Technology and Investment, 2013, 4, 27-30
Published Online February 2013 (http://www.SciRP.org/journal/ti)

Based on ODE and ARIMA Modeling for the population of China

Xiaohua Hu, Min Yu
School of Mathematics and Statistics, Hainan Normal University, Haikou, China
Email: 1241957415@qq.com

Received 2012

ABSTRACT

The economic data usually can be also composed into a deterministic part and a random part. We establish ordinary differential equations (ODE) model for the deterministic part from a purely mathematical point of view, via the principle of integral and difference, establishing $ARIMA(p,d,q)$ model for the random part, to combine the two established models to predict and control the original series, then we apply the method to study the population of China from 1978 to 2007, establishing the corresponding mathematical model, to obtain the forecast data of the population of China in 2008(1.3879503 billion), finally we make further stability analysis.

Keywords: Natural Asset; Financial Value; Neural Network

1. Introduction

A lot of time series (as economic data) can be regarded as a discretization of continuous-time process. if a time series Y possesses biology background, we can establish a differential equations model for Y in accordance with the growth rate, for example, suppose that constant L is the growth limit of the variable Y, the rate of growth dY/dt is proportional to Y or $L-Y$; suppose that the relative rate of growth $(dY/dt)/Y$ is proportional to $(L-Y)/L$ or $\ln L-\ln Y$; some simple growth models can be established respectively, such as the Logistic model, Gompertz model[1] and so on, they are called the growth mechanism models.

In general, a time series $Q_t, t=0,1,...,n,$ can be decomposed into two parts[2]: a deterministic part F_t and a random part

$h_t, t=0,1,...,n$, or $Q_t = F_t + h_t, \quad t=0,1,...,n$ (1)

where F_t is usually described by the trend, seasonal and cyclic term, and h_t is described by a more complex stochastic approach. In this paper, we ignore specific economic backgrounds of variable, from a purely mathematical point of view, to establish differential equations model for the discrete-time series F_t on some conditions, while h_t can be described by $ARIMA(p,d,q)$ model[3].

2. Principles or technology of differential equations modeling

Suppose that $y(t), 0 \le t < +\infty$, is a continuous function, F_t is regarded as the discretised value of $y(t)$. Set

$$X(t) = \int_0^t y(t)d \approx \sum y(t)\Delta t$$

$X'(t) = y(t) \approx (X(t+\Delta t) - X(t))/\Delta t$, as Δt is small.

Now let's consider the discrete case (time series) and let $\Delta t = 1$ (unit time). It follows that

$$X(t) \approx \sum y(t), X'(t) = y(t) \approx X(t+1) - X(t)$$

Denoting that $y_t = y(t), X_t = X(t)$, if given a original time series $y_t = y(t), t = 0,1,...,n$,its cumulative sum series X_t can be generated as $X_t = \sum_{k=0}^{t} y_k, t = 0,1,...,n$,so that, if we can find the relationship between the original series y_t and its cumulative sum series X_t as below,

$$F(y_t, X_t, \varepsilon_t) = 0, t = 0,1,...,n$$

where ε_t is a residual term, a random variable, satisfies $E(\varepsilon_t) = 0, D(\varepsilon_t) = \sigma^2, \sigma > 0$ .We now view $y(t), X(t)$ as continuous, because of $X'(t) = y(t)$, ignoring ε_t , corresponding to the one-order differential equation model can be established as follows.

$$F(X(t), X'(t)) = 0 \quad or \quad X'(t) = f(X(t)) \qquad (2)$$

We just need to solve out $X(t)$ from (2), so, $y_t = y(t) = X'(t) \approx X(t+1) - X(t), y_0 = y(0) = X(0)$, $t = 1,...,m.m \ge n$ (where m is a positive integer) .

In the same way, $X_t = X(t), t = 0,1,...,n.$ is viewed as a new time series, we can generate its cumulative sum series $Z(t) = \sum X(t)$, or, the second cumulative sum of y_t ,denote $Z_t = Z(t) \cdot Z'(x) = X(t) \approx Z(t+1) - Z(t)$ .we consider to establish the relationship of three series y_t, X_t, Z_t, if there exists the relationship as follows

$$F(y_t, X_t, Z_t, \varepsilon_t) = 0, t = 0,1,...,n$$

where ε_t is the same with the previously mentioned. We now view $y(t), X(t)$ as a continuous case, because of $Z''(t) = X'(t) = y(t)$, ignoring ε_t ,the two-order differential equation model can be established.

$$F(Z''(t), Z'(t), Z(t)) = 0 \quad or \quad Z''(t) = f(Z(t), Z'(t)) \tag{3}$$

we just need to solve out $Z(t)$ from (3), so,

$$X_t = X(t) = Z'(t) \approx Z(t+1) - Z(t),$$
$$y_t = y(t) = X'(t) \approx X(t+1) - X(t),$$
$$y(0) = X(0) = Z(0), t = 1,..., m.m \geq n$$

In a general way, suppose that the original time series is $X_0(t)$. If $X_0(t)$ is not pure random data, its value's change on unit time is not random, or data has a trend. its the first cumulative sum series is $X_1(t)$, ..., the p-th cumulative sum series is $X_p(t), t = 1,...,n$,where p is a positive integer, if we can find the relationship of $X_0(t), X_1(t),..., X_p(t)$ as below

$$F(X_0(t), X_1(t),..., X_p(t), \varepsilon_t) = 0, t = 0,1,...,n$$

where ε_t is the same as previous, ignoring ε_t ,the p-order differential equation model can be established.

$$F(\frac{d^p X_p(t)}{dt^p}, \frac{d^{p-1} X_p(t)}{dt^{p-1}}, ..., \frac{d_p X(t)}{dt}, X_p(t)) = 0$$
$$or \quad \frac{d^p X_p(t)}{dt^p} = f(\frac{d^{p-1} X_p(t)}{dt^{p-1}}, ..., \frac{d_p X(t)}{dt}, X_p(t))$$

We just need to solve out $X_p(t)$, so

$$X_i(t) = X_{i+1}(t+1) - X_{i+1}(t), i = 1,..., p-1, t = 1,..., m.m \geq n$$
$$X_0(0) = X_1(0) = ... = X_p(0)$$

It is usually difficult to find F, f ,but, we can consider to establish the multiple linear (or nonlinear) regression model. Given a significance level α (for example, $\alpha = 0.05$),if the significance test for the regression equation can be established, we can find the corresponding differential equations at 1-α confidence level, which can explain the reasonable degree for the established differential equations. The adjusted R-squared (R^2) or goodness-of-fit can describe the fitting degree of good or bad. We call $F = 0$ the main model. for the residual series ε_t ,we test that it is or isn't random by some test methods, if it isn't a pure random series(such as white-noise series[4]), it shows that there exists some valuable information hid in residual series, the information should be extracted out from the residual series by making use of the B-J method[5], at this time, we shall establish the model for residual series, It is called as the auxiliary model. Finally, we combine the main model and the auxiliary model to forecast.

3. Empirical Analysis Application

We study the Chinese population data[6] from 1978 to 2007, the total number of sample observations is 21, see

Table 1 Data of the population of China(unit:ten thousand)

1978	1980	1985	1990	1991	1992	1993	1994	1995	1996	1997
96259	98705	105851	114333	115823	117171	118517	119850	121121	122389	123626
1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	
124761	125786	126743	127627	128453	129227	129988	130756	131448	132129	

We use the sample data from 1978 to 2005 to model, leaving two sample data of 2006 and 2007 as a reference to assess the short-term forecast to see the accuracy of the established model.

3.1. Analyse the Relationship Between the Original Series and its Cumulative Series

Let the original sample time series be $y(t) = y_t, t = 0,1,...,20$, the time range to model is from 1978 to 2005.that is $t = 0,1,...,18$, the first cumulative sum series of y_t is $X_t = X(t) = X$, see fig.1,fig.2. We generate new series Z_t, R_t via y_t, X_t as follows

$$Z_t = Z(t) = \frac{y_t}{X_t}, R_t = R(t) = \ln X(t)$$

The scatter plot of Z_t and R_t is such as fig.3.We establish regression model below

$$Z_t = c(1) + c(2) R_t + c(3) R_t^2. \tag{4}$$

By making use of least-squares method and EViews6.0, it is easy to obtain the estimation value of parameters $c(1), c(2), c(3)$, see fig.4 and Tab.2.

$$Z_t = 22.19772 - 3.0877147 R_t + 0.107731 R_t^2 \tag{5}$$

Table 2. Results of estimate and test (significance level α=0.05)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
c(1)	22.19772	1.334405	16.63491	0.0000
c(2)	-3.0877147	0.202384	-15.25674	0.0000
c(3)	0.107731	0.007644	14.09369	0.0000

The adjusted R^2 = 0.986264.The t-statistics of three regression coefficients $c(1), c(2), c(3)$ in (4) are respec-

tively 16.64, -15.26,14.09, their Prob=0.0000<0.05,

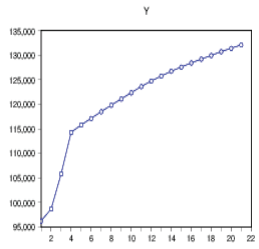

fig.1: the original time series y

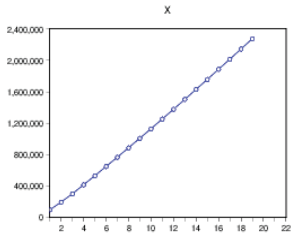

fig.2: the cumulative sum series X

F-statistics=647.2. Prob = 0.0000<0.05. the significant test for coefficients and the whole regression equation were resulted. We think that the Chinese population data from 1978 to 2005 can be described by (5) at 95% confidence level, Goodness-of-fit reaches 98.6%. However, DW= 1.317 shows existence of autocorrelation in the residual series, It shows that there are still some valuable information not to be extracted out from the residual series, so we will establish the model for the residual series.

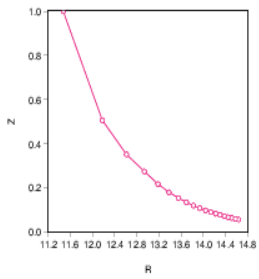

fig.3: the scatter plot of R and Z

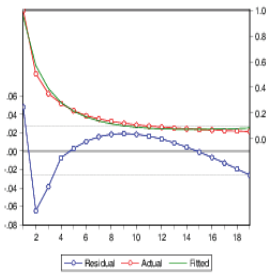

fig.4: the fitting and residuals plot of R and Z

3.2. Establish the Main Model–Differential Equations Model

It follows from (2),(4)and(5)that

$$\frac{dX}{dt} = X'(t) = [c(1) + c(2)\ln X(t) + c(3)\ln^2 X(t)]X(t) \tag{6}$$

$$\frac{dX}{dt} = X'(t) = [22.198 - 3.088\ln X(t) + 0.108\ln^2 X(t)]X(t) \tag{7}$$

by formula[7],when $\sqrt{4ac-b^2} > 0$,

$$\int \frac{du}{a+bu+cu^2} = \frac{2}{\sqrt{4ac-b^2}}\arctan\frac{2cu+b}{\sqrt{4ac-b^2}} + C,$$

It is easy to obtain from (6)

$$\frac{2}{\sqrt{4c(1)c(3)-c^2(2)}}\arctan\frac{2c(3)\ln X + c(2)}{\sqrt{4c(1)c(3)-c^2(2)}} + C = t$$

so

$$X = \exp\{\frac{1}{2c(3)}[-c(2) + \sqrt{4c(1)c(3)-c^2(2)}]\times \tan[\frac{1}{2}\sqrt{4c(1)c(3)-c^2(2)}(t-C)]\}$$

or

$$X = \exp\{14.296 + 1.074\tan(\frac{t}{8.624} + C1)\}$$

$t = 0, X(0) = 96259, C1 = -69.519$, so

$$X = \exp(14.296 + 1.074\tan(\frac{t}{8.624} - 69.519)) \tag{8}$$

$$X' = 0.125\sec^2(\frac{t}{8.624} - 69.519))\exp(14.296 + 1.074\tan(\frac{t}{8.624} - 69.519)) \tag{9}$$

3.3. Establish the Auxiliary Model for Residual Series

Let the residual series be h_t ,we establish $ARIMA(p,d,q)$ model for $h_t = h(t)$,make 1-order, 2-order difference for h_t below

$$\nabla h_t = h(t) - h(t-1) = h_t - h_{t-1}, \nabla^2 h_t = \nabla h_t - \nabla h_{t-1}$$

based on the analysis of the autocorrelation coefficients and partial autocorrelation coefficients of h_t and 2-order difference of h_t .we take $p = q = d = 2$,establish $ARMA(2,2)$ model for $"^2 h_t$.

$$\nabla^2 h_t = \varphi_1\nabla^2 h_{t-1} + \varphi_2\nabla^2 h_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} \tag{10}$$

$$h_t = (2+\varphi_1)h_{t-1} + (\varphi_2 - 2\varphi_1 - 1)h_{t-2} + \varphi_1 h_{t-3} + \varphi_2 h_{t-4} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$$

this is a $ARIMA(4,2,2)$ for h_t ,where ε_t is a white-noise series. the parameters in (10) are estimated by EViews6.0, $\hat{\varphi_1} = 0.8436, \varphi_2 = 0.0328, \hat{\theta_1} = 0, \theta_2 = 0.9992$.

$$h_t = 2.8436h_{t-1} - 2.6544h_{t-2} + 0.8436h_{t-3} + 0.0328h_{t-4} + \varepsilon_t - 0.9992\varepsilon_{t-2} \tag{11}$$

3.4. Forecast Based on the Main and Auxiliary Model

By (8) and (11), as t=19, 20, 21, y=131190.5504, 136191.0595, 138795.0589; h_t =-0.007129, -0.0133, -0.01979, so, by (1),the predictive data of the population of China in 2006,2007 are 1.3119054,1.3619104 billion, respectively, however, the actual data are 1.31448, 1.32129 billion, respectively; the absolute errors are 2.58, 40.62,respectively;the relative errors are 0.2%, 3.1%,respectively;It shows that the combination model (1),(8) and (11) have a higher prediction accuracy, and the predictive data of the population of China in 2008 is

1.3879503 billion. We also find that predictive value of the auxiliary model (11) is little impact on the total predictive value. the total predictive value mainly depends on the main model (8),or, mainly depends on the predictive value of the differential equations model (7).So, we can see that the short-term forecast accuracy is very high based on differential equations modeling for the time series on some condition.

We further consider the stability of equilibrium point of (6) or (7). Let

$$\frac{dX}{dt} = [c(1) + c(2)\ln X(t) + c(3)\ln^2 X(t)]X(t) = f(X) = 0$$

$X \neq 0, [c(1) + c(2)\ln X(t) + c(3)\ln^2 X(t)] = 0$, when $c^2(2) - 4c(3)c(1) > 0$, there are two real roots, denoted by u_1, u_2 .there are two equilibrium points X_1, X_2 , $X_1 = \exp(u_1), X_2 = \exp(u_2)$. on the other hand,

$$f'(X) = c(1) + c(2) + (c(1) + 2c(3))\ln X + c(3)\ln^2 X,$$

When $c^2(2) - 4c(3)c(1) > 0,$

$$f'(X_1) = \sqrt{c^2(2) - 4c(3)c(1)},$$

$$f'(X_2) = -\sqrt{c^2(2 \quad 4c(3)c(1)}. f'(X_1) > 0, f'(X_2) < 0$$

The equilibrium point X_2 of (6) is stable, or, $t \to +\infty, X \to X_2$.the equilibrium point X_1 of (6) is unstable, or, $t \to +\infty, X \to +\infty$. so, we must control those factors that impact $c(1), c(2), c(3)$ in (6),such that $c^2(2) - 4c(3)c(1) > 0$. otherwise, $t \to +\infty, X \to +\infty$

However, in fact, for model (7), $c^2(2) - 4c(3)c(1) < 0$,it show that there is no equilibrium point in (7), it is obviously from (9),as $t / 8.624 - 69.519 \to \frac{\pi}{2},$ or $t \to 613, y \to +\infty$. it show that China's population will tend to infinity after 613 years, so, the model (7) is only suitable for short-term prediction.

4. Acknowledgements

Acknowledgements: the author is grateful to the anonymous referees for his helpful comments and suggestions.

REFERENCES

[1] Zhu Minhui. Fitting Gompertz Model and Logistic Model.J. Mathematics in Practice and Theory,2003;2:705-709.

[2] Peter J.Brockwell et al. Time series: theory and methods (2nd edn).China Higher Education Press Beijing and Springer-Verlag Berlin Heidelberg:Beijing,2001;75.

[3] Yi Danhui.Data analysis and Eviews application. China Statistics Press:Beijing,2002;66-70.

[4] Zhang Shiying.The financial time series analysis. Tsinghua University Press:Beijing,2008;90-93.

[5] Philip Hans Franses.Time Series Models for Business and Economic Forecasting[M].Beijing: Chinese People's University Press.2002.

[6] http//www.cpachn.org.cn/chinese/Teaching/Information.asp?,2009-05-02.

[7] Tongji University Department of Applied Mathematics. Advanced Mathematics (5th edn), Beijing: Higher Education Press.2004.

