Study on population forecast of Shaanxi Province Based on ARMA and BP neural network model

Ruixia Suo

Xi'an University of Technology, Xi'an 710054, China

454069279@qq.com

Abstract

Based on the analysis of the population historical data of Shaanxi Province, this paper established the ARMA model and the BP neural network model of two single forecasting models, and application of the method of the reciprocal of the variance to build a combined forecasting model. The fitting results show that the combined forecasting model is better than the single model. At last, the population of Shaanxi province is predicted, which can be a reference for the development strategy and plan.

Keywords

ARIMA model; BP neural network; population; prediction.

1. Introduction

The number of people not only affects the development of the national economy, but also affects the social stability, production arrangement, labor employment and the sustainable use of resources and so on. Therefore, it is important to study the number of people in different time. It can provide a theoretical basis for the government to formulate population policies, and provide the reference for the relevant departments to formulate development strategies and plans, and provide basic data for the calculation of social and economic development [1, 2]. Based on the analysis of the population historical data of Shaanxi Province, this paper established the ARMA model and the BP neural network model of two single forecasting models [3,4], and application of the method of the reciprocal of the variance to build a combined forecasting model.

2. Establishment of ARIMA model

ARMA (p, d, q) model referred to as autoregressive moving average model[3], which is composed of autoregressive and moving average of two parts, P, Q respectively denote the order of autoregressive and moving average part, d stands for the differential order, the general expression for

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

2.1 The time series data's stationary analysis

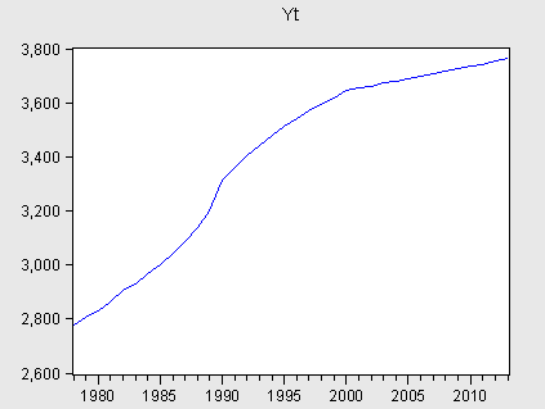


Fig.1 The population data of Shaanxi province in 1978-2013

According to the population sequence data between the 1978-2013(such as Fig. 1), the y_t is a non stationary series, and the Eviews6.0 is used to carry on the first order difference, from the autocorrelation and partial correlation Fig.2 of the first order difference, The autocorrelation coefficient of the sequence is not fastly near to zero, and the time series is still not stable, so the two order difference is carried on. Two order differential sequence unit root test results are shown in Table 1.

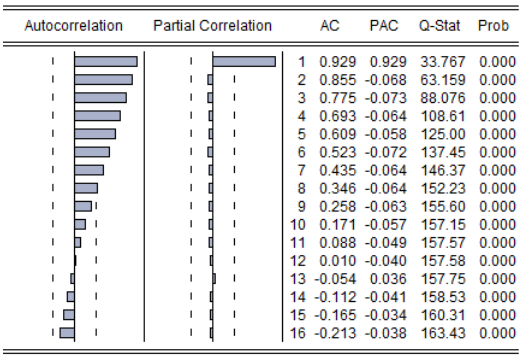


Fig.2 Dy's related and partial correlation graph

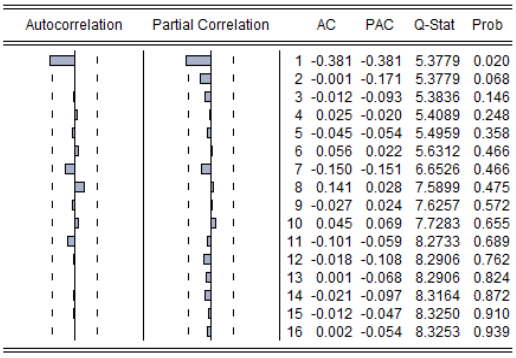


Fig.3 D(Dy)'s related and partial correlation graph

Table1 Two order differential sequence unit root test results

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.320193	0.0000
Test critical values:	1% level	-3.646342	
	5% level	-2.954021	
	10% level	-2.615817	

From the test results, we can know that there is no unit root of the two order difference, which is stationary sequence. The autocorrelation and partial autocorrelation analysis of the two order differential sequences are plotted, as shown in Fig.3, we can show that the partial and autocorrelation coefficient of the two order difference sequence is neared to the random interval. The trend has been basically eliminated, so d=2.

2.2 Identification and establishment of model

According to the AIC, the SC criterion, after reasonable selection and selection, P takes 1, Q takes 1, takes the ARIMA (1, 2, 1) model as the final model. Model estimation results are shown in Fig.4.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.517113	0.502208	-3.020886	0.0051
AR(1)	0.465814	0.156109	2.983900	0.0056
MA(1)	-0.978845	0.031303	-31.27002	0.0000

R-squared	0.273579	Mean dependent var	-0.393939
Adjusted R-squared	0.225151	S.D. dependent var	17.24445
S.E. of regression	15.17952	Akaike info criterion	8.364279
Sum squared resid	6912.534	Schwarz criterion	8.500325
Log likelihood	-135.0106	Hannan-Quinn criter.	8.410055
F-statistic	5.649184	Durbin-Watson stat	2.189470
Prob(F-statistic)	0.008276		

Inverted AR Roots	.47
Inverted MA Roots	.98

Fig. 4 ARIMA (1, 2, 1) model parameters

From Fig.4, the model's estimation result is:

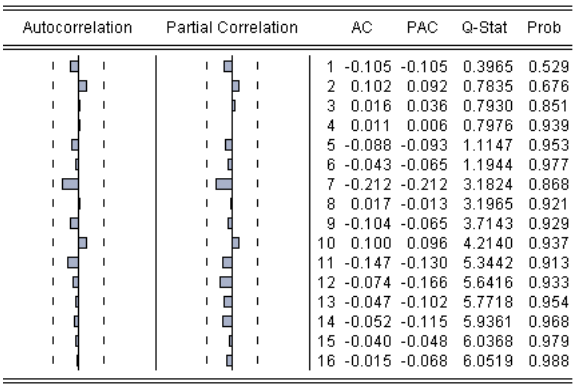
$$x_t = -1.5171 + 0.4658x_{t-1} - 0.9788u_t$$

Which,

$$x_t = \Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$$

2.3 Residual test of sequence model

The residual of the model is tested and the test results are shown in Fig.5.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.105	-0.105	0.3965	0.529
		2	0.102	0.092	0.7835	0.676
		3	0.016	0.036	0.7930	0.851
		4	0.011	0.006	0.7976	0.939
		5	-0.088	-0.093	1.1147	0.953
		6	-0.043	-0.065	1.1944	0.977
		7	-0.212	-0.212	3.1824	0.868
		8	0.017	-0.013	3.1965	0.921
		9	-0.104	-0.065	3.7143	0.929
		10	0.100	0.096	4.2140	0.937
		11	-0.147	-0.130	5.3442	0.913
		12	-0.074	-0.166	5.6416	0.933
		13	-0.047	-0.102	5.7718	0.954
		14	-0.052	-0.115	5.9361	0.968
		15	-0.040	-0.048	6.0368	0.979
		16	-0.015	-0.068	6.0519	0.988

Fig.5 The figure of relevant and partial correlation of the residual sequence

Because the probability values of the right side of Figure 4 are more than 0.05, that is all Q values are less than the threshold value of the chi-square distribution of 0.05.The random error term of the established model is a white noise sequence. So the model is appropriate.

2.4 Model prediction

According to the ARIMA (1, 2, and 1) model, the population of Shaanxi province was predicted, the average relative error was 1.36% (shown in Tab.2 and Fig.6), the forecast effect was good, so the model can be used to predict the population of Shaanxi province. The population between 2014 and 2020 in Shaanxi province is predicted, and the forecast results are shown in Table 2.

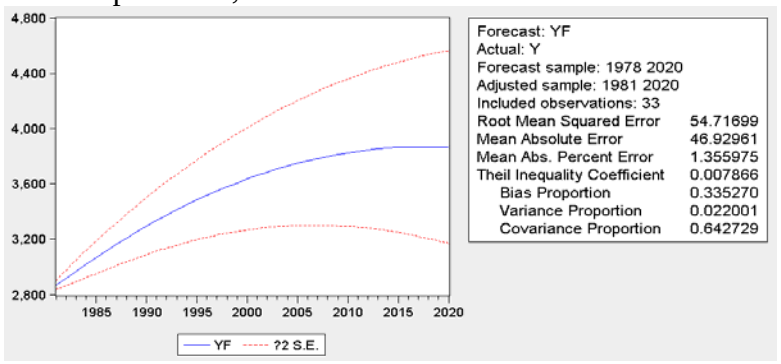


Fig. 6 ARIMA (1, 2, 1) model prediction results

3. Establishment of BP neural network model

Based on the population data of Shaanxi Province in 1978-2013, this paper chooses the data of three years ago to predict the data of fourth years, when the BP neural network model is trained by the following parameters: (1) the number of input layer nodes is 3; (2) the number of hidden nodes is 7 (3), the number of output layer node is 1 (4), (5) the network training error is 0.0001,the initial weights, learning rate and the threshold of the network are automatically selected by the network, and then BP neural network fitting precision is higher. By the Matlab programming, the weight between the input layer and the hidden layer is

$$w=\begin{bmatrix} 0.9193 & -1.0442 & 0.2503 & -0.6862 & -0.4324 & -0.2911 & -0.3041 \\ 0.0794 & -0.2340 & 0.2915 & -0.7463 & -0.3188 & -0.3134 & 0.4064 \\ 1.3057 & -0.3692 & -0.9527 & -0.3421 & -1.2777 & -1.3002 & 1.3027 \end{bmatrix}$$

The threshold between the input layer and the hidden layer is

$$st1=\begin{bmatrix} 0.4082 & 0.5866 & 0.2499 & 0.7379 & 0.2469 & 0.4938 & 0.7261 \end{bmatrix}$$

The weight between the hidden layer and the output layer is

$$v=\begin{bmatrix} 2.5737 & -1.4434 & -0.3834 & -2.1846 & -2.4126 & -2.2436 & 1.1095 \end{bmatrix}$$

The threshold between the hidden layer and the output layer is

$$st2 = 0.3046$$

The average relative error is 0.4684% (see Table 2), the BP model can be used for the prediction of future moments.

4. Establishment of combined forecasting model

According to the variance reciprocal method, the weight w of the combined forecasting model is (0.3982, 0.6018), and the model can be obtained as follows

$$Y_t = 0.3982 y_{1t} + 0.6018 y_{2t}$$

Where, y_{1t} Y for the forecast value of the ARMA model, y_{2t} for the forecast value of the BP model.

The fitting value of the model is obtained by using the model, as shown in Table 2.

Table 2 The fitting value and error of population of Shaanxi province in 1981-2013

Year	Actual value	Fitting value			Relative error（%）		
		ARMA model	BP model	Combined forecasting	ARMA model	BP model	Combined forecasting
1981	2865	2871.44	2883.1	2878.46	0.22	0.63	0.47
1982	2904	2918.74	2911.5	2914.38	0.51	0.26	0.36
1983	2931	2968.41	2942.6	2952.88	1.28	0.40	0.75
1984	2966	3018.38	2977.1	2993.54	1.77	0.37	0.93
1985	3002	3067.69	3014.7	3035.80	2.19	0.42	1.13
1986	3042	3115.86	3049.0	3075.63	2.43	0.23	1.11
1987	3088	3162.71	3089.5	3118.65	2.42	0.05	0.99
1988	3140	3208.12	3133.9	3163.45	2.17	0.19	0.75
1989	3198	3252.06	3184.0	3211.10	1.69	0.44	0.41
1990	3316	3294.49	3240.4	3261.94	0.65	2.28	1.63
1991	3363	3335.42	3319.2	3325.66	0.82	1.30	1.11
1992	3405	3374.84	3388.8	3383.24	0.89	0.48	0.64
1993	3443	3412.74	3464.5	3443.89	0.88	0.62	0.03
1994	3481	3449.12	3502.8	3481.42	0.92	0.63	0.01
1995	3513	3483.99	3536.3	3515.47	0.83	0.66	0.07
1996	3543	3517.34	3565.1	3546.08	0.72	0.62	0.09
1997	3570	3549.17	3590.9	3574.28	0.58	0.59	0.12
1998	3596	3579.48	3612.5	3599.35	0.46	0.46	0.09
1999	3618	3608.28	3631.8	3622.44	0.27	0.38	0.12
2000	3644	3635.56	3648.4	3643.29	0.23	0.12	0.02
2001	3653	3661.33	3664.3	3663.12	0.23	0.31	0.28
2002	3662	3685.58	3675.4	3679.45	0.64	0.37	0.48
2003	3672	3708.31	3685.4	3694.52	0.99	0.36	0.61
2004	3681	3729.52	3690.8	3706.22	1.32	0.27	0.69
2005	3690	3749.22	3696	3717.19	1.60	0.16	0.74
2006	3699	3767.40	3701.3	3727.62	1.85	0.06	0.77
2007	3708	3784.06	3706.2	3737.20	2.05	0.05	0.79
2008	3718	3799.20	3711.1	3746.18	2.18	0.19	0.76
2009	3727	3812.83	3716	3754.56	2.30	0.30	0.74
2010	3735	3824.94	3720.8	3762.27	2.41	0.38	0.73
2011	3743	3835.53	3725.5	3769.32	2.47	0.47	0.70
2012	3753	3844.61	3729.8	3775.52	2.44	0.62	0.60
2013	3764	3852.17	3734.1	3781.12	2.34	0.79	0.45

From the calculation results, the maximum relative error of the combination forecast is 1.63%, the maximum relative error of the ARMA and BP neural network is respectively 2.47% and 2.28%, , which can be used for the prediction of population in Shaanxi Province in the future. In 2014-2020 the population prediction results in Shaanxi Province are shown in table 3.

Table 3 The population forecast value of Shaanxi province in 2014-2020

Year	ARIMA（1,2,1）forecast value	BP forecast value	Combined forecasting value
2014	3858.21	3733.6	3826.21
2015	3862.74	3734.1	3829.70
2016	3865.74	3729.7	3830.81
2017	3867.23	3728.9	3831.71
2018	3867.21	3728.5	3831.59
2019	3865.67	3727.3	3830.13
2020	3862.60	3726.8	3827.73

From the prediction results show that, under the existing fertility pattern and fertility policy, the population for 38.2773 million people of Shaanxi Province in 2020, with an average annual growth rate of 0.24%, and the population will be in negative growth from 2018, so the release two-child policy is necessary, to promote and realize the long-term stable development of the population greatly.

5. Conclusions

In this paper, the population of Shaanxi province is predicted by ARMA model and BP neural network model, and then by using the variance reciprocal method to determine the weight of each single forecasting method. The results show that the combined forecasting model has smaller prediction error, and the population of Shaanxi Province in 2020 is 38.2773 million, and the average annual growth rate is 0.24%.

Acknowledgements

Shaanxi Provincial Social Science Fund (2014D38)

The philosophy and Social Sciences prosperity development plan of XUST (2014SY10)

References

[1] Jiang Yuanying,Wang Xiang. Application of population development equation model in population forecast in China[J]. Statistics and Decision,(2011)NO.15,p52-54.

[2] Yu Li; Yang Lintao.China Population Development Forecast Based on bilinear model, Statistics and Decision, (2014) NO.22, p: 90-92

[3] Ren Qiang, Hou Dadao.Stochastio Model for population forecast based on Leslie matrix and ARMA model. Vol.35 (2011) NO.14, p28-41

[4] YIN Chunhua, CHEN Lei. Study and Application of Population Forecast Model Based on BP Neural [J].Population Journal, (2005) NO.2, p: 44-48