Monitoring the SARS Epidemic in China:
A Time Series Analysis

Dejian Lai

The University of Texas and Jiangxi University of Finance and Economics

Abstract: In this article, we studied three types of time series analysis methods in modeling and forecasting the severe acute respiratory syndrome (SARS) epidemic in mainland China. The first model was a Box-Jenkins model, autoregressive model with order 1 (AR(1)). The second model was a random walk (ARIMA(0,1,0)) model on the log transformed daily reported SARS cases and the third one was a combination of growth curve fitting and autoregressive moving average model, ARMA(1,1). We applied all these three methods to monitor the dynamic of SARS in China based on the daily probable new cases reported by the Ministry of Health of China.

Key words: ARIMA, forecasting, growth curve, random walk, SARS.

1. Introduction

Since the sudden epidemic of severe acute respiratory syndrome (SARS) in Guangdong, China in November 2002, this infectious disease has infected more than 8000 people worldwide by the end of May 2003 and about 2/3 of these cases were in mainland China (WHO 2003). SARS was called atypical pneumonia in China due to its unknown etiology initially. The name, a typical pneumonia, has still been used in China in most publications in Chinese because of the convenience and consistency although it was officially named as SARS by the World Health Organization (WHO) in March 2003.

SARS is a highly infectious disease with relatively high mortality (about 10%) and possible severe damage to organs in the survivors of the disease. After a few months in Guangdong, the disease found its way to the capital, Beijing, and other places in China. Later, it crossed the boarder of mainland China to Guangdong's neighbor, Hong Kong, in March 2003. Hong Kong, a special administration region of China, has a very high population density and it is also a highly internationalized city. Following the outbreak of the disease in Hong Kong, SARS has infected more than 30 countries/regions worldwide according to WHO's report (WHO 2003) since March 2003. The virus that causes SARS was

identified as a novel coronavirus in March 2003 and confirmed by WHO in April by the intensive and rapid work of many scientists around the world (Drosten *et al.* 2003; Ksiazek *et al.* 2003; Peiris *et al.* 2003). The papers on the genome sequence of the SARS virus were published in *Science* (Rota *et al.* 2003; Marra *et al.* 2003). The discovery of the main proteinase structure of the novel coronavirus may be useful for developing anti-SARS drugs (Anand *et al.* 2003).

At present, there is no effective drug to cure the disease and the vaccine against the disease has yet to be developed. Hence, SARS poses a great threat to the global public health of human beings and may lead to a possible devastating impact to world's economic and development. Although the seriousness of the epidemic of SARS was not initially recognized by some Chinese officials, the World Health Organization pressured Chinese government for more information and issued travel warning to many places in China since March 2003 (WHO 2003). In the middle of April 2003, the Chinese government removed a few high ranking officials who failed to actively lead the efforts for curbing the spread of SARS. Since then, various agencies in China have engaged in extraordinary efforts to stop the spread of the disease. In so doing, large number of people were quarantined due to possible contacts with probable or suspected SARS patients. Quarantining a large number of people is an enormous task for any government. It requires a large amount of resource and results a great loss of productivity in many business sectors in addition to the abnormal life of the people under quarantine. One important issue for the policy makers in planning the efforts in controlling SARS would be the possible future number of SARS cases.

Using data from Hong Kong, Singapore and Canada, the epidemic and the transmission dynamics of SARS were studied (Riley *et al.* 2003; Lipsitch *et al.* 2003; Chau and Yip 2003; Choi and Pak 2003). Mathematical dynamical models were also used for simulating the SARS outbreak in Beijing (Wang and Ruan 2004). However, there seems a lack of statistical analysis on the data from China in the literature. In our study, we used the data from China and compared three popular statistical models for forecasting. The first model was a Box-Jenkins model, autoregressive model with order 1 (AR(1)). The second model was a random walk (ARIMA(0,1,0)) model on the log transformed daily new reported SARS cases and the third one was a combination of growth curve fitting and autoregressive moving average model, ARMA(1,1). Due to limited number and nonstationarity of the data, a direct use of ARMA seems not appropriated. These validity issues of the models are discussed in the last section of this article.

We tried all three models on the officially reported SARS cases in mainland China since day 22 (May 12) after April 21. For an illustration, we presented the results up to day 30 (May 20) in this article. For a comparison, we also tabulated the one day ahead forecasts, their 95% confidence intervals and the

actual observations in Tables 1, 2 and 3 by these three models, respectively. Figures 1, 3 and 5 showed the forecasting values.

2. Data Source

It seemed that the first SARS patient was able to be traced back in November 2002 in Guangdong, China. Due to a lack of an effective public health surveillance system in China for new diseases and a lack of understanding of the potential threat to the general public at that time, the disease was not acknowledged as serious as it should be by Chinese authorities. Nevertheless, the disease spread rapidly. According to various news reports, the number of SARS cases in Guangdong peaked around the Spring Festival in February 2003. The government tried to calm down the general public by not recognizing the seriousness of SARS. Later, the disease spread to the capital, Beijing, and the neighboring region of Hong Kong, then to other countries by air travelers. In March 2003, SARS infected and killed a WHO doctor who treated a SARS patient in Vietnam. On March 12, WHO issued an emergency travel advisory to several affected areas. The travel advisory guidelines have been issued and modified accordingly since then. By May 27, many places were removed from the list of considering postponing all but essential travel. These places included Guangdong and Hong Kong. A great progress of controlling SARS have been achieved. There have been very few new SARS cases in China since June 1, 2003. By August 16, 2003, the last two SARS patients in the 2003 epidemic were released from the hospital in Beijing. Since the end of the 2003 epidemic of SARS, there were a few SARS cases occurred in Guangzhou in December 2003. In late March and mid April of 2004, two researchers at the National Institute of Virology developed SARS. Because of the effective control, the isolated SARS cases in December 2003 in Guangzhou and in April 2004 in Beijing did not start another SARS epidemic in China.

Prior to April 21, 2003, there was no systematical publicly available daily number of SARS cases in China. Starting from April 21, 2003, the daily reporting of SARS was published on the official website of the Ministry of Health of China[1]. Around that time, the number of newly infected cases increased dramatically in the capital, Beijing, while new cases in Guangdong province decreased. We followed the daily reports closely. The time series of the SARS cases modeled in this article were extracted from the official reports of the Ministry of Health of China up to May 31 , 2003 (MOHC 2003). The number of daily reported new probable SARS cases in China is presented in Figure 1. From Figure 1, one can clearly see that in the first 19 days following April 21, 2003 the daily reported new SARS cases were all greater than 100. Most of these cases must

---

[1]See, http://www.moh.gov.cn/zhgl/yqfb/index.htm.

have been infected before April 21 because the strict quarantine policy, the close of schools and the ban of large public gatherings after April 21 could effectively stop the rapid spread of SARS. Due to these strong interventions in Beijing and across China starting in the middle of April, the daily new SARS cases decreased dramatically from May 10 to May 15. From May 16 to May 25, the daily new probable SARS cases remained relatively low. Since May 26, the number of new daily SARS cases has been less than 10. In June, the number of daily SARS case has been mostly below 5. To statistically quantify the dynamic of the infections of SARS and forecast the new cases of infections, we used and compared three time series models since May 12, 2003 (day 22) almost daily up to the end of May. In this article we report our findings of our modeling on the SARS time series.
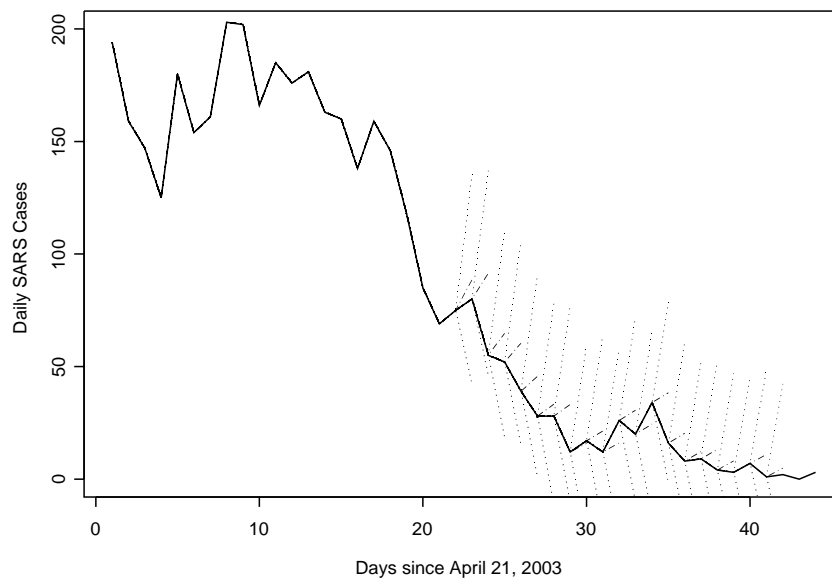

Figure 1: The daily new SARS cases in mainland China from April 21 to June 3, 2003 and the one-day ahead forecast and its 95% confidence interval based on AR(1) model for the daily new SARS cases in mainland China from May 12 to May 31, 2003.

3. Box-Jenkins Modeling

Box-Jenkins methods (Box and Jenkins 1976) are well established techniques for analyzing stationary linear time series. From Figure 1, the dynamic of the

Table 1: The Estimates of the parameters in the Box-Jenkins model and the forecasting of the daily new SARS cases: $\hat{\alpha}_1$ the estimate of the coefficient in AR(1), $\hat{\mu}$ the estimate of the mean, $\hat{\sigma}$ the estimate of the standard deviation of the innovation in AR(1), (a) nne-day ahead forecast, (b) the 95% confidence interval of the one-day ahead forecast, (c) actual one-day ahead observation

Day	$\hat{\alpha}_1$	$\hat{\mu}$	$\hat{\sigma}$	(a)	(b)	(c)
22	0.8238	152.1	23.88	89	(42,135)	80
23	0.8351	149.0	23.28	91	(46,137)	55
24	0.8921	145.0	23.74	65	(18,111)	52
25	0.9069	141.3	23.27	60	(15,106)	39
26	0.9323	137.4	23.08	46	(1,91)	28
27	0.9499	133.3	22.80	33	(0,78)	28
28	0.9492	129.6	22.35	33	(0,77)	12
29	0.9662	125.5	22.23	16	(0,59)	17
30	0.9584	121.9	21.80	21	(0,64)	12
31	0.9616	118.4	21.46	16	(0,58)	26
32	0.9490	115.5	21.14	31	(0,72)	20
33	0.9525	112.6	20.86	24	(0,65)	34
34	0.9434	110.3	20.57	38	(0,79)	16
35	0.9520	107.6	20.58	20	(0,61)	8
36	0.9562	104.8	20.35	12	(0,52)	9
37	0.9552	102.2	20.05	13	(0,52)	4
38	0.9574	99.6	19.80	8	(0,47)	3
39	0.9575	97.2	19.53	7	(0,45)	7
40	0.9554	94.9	19.25	11	(0,49)	1
41	0.9576	92.6	19.05	5	(0,42)	2

daily new SARS cases in China seemed not stationary. To recognize the non-stationary features of the daily observations of the SARS cases in China, one could create an indicator variable (deterministic trend) and fit it to the observations. In Section 5, we discuss such approach via growth curve fitting. Alternatively, one may assume that the trend was due to stochasticity and use difference to remove the trend. We discuss this approach in Section 4 on log scale. In current section, we report the results from the plain ARMA(p,q) model:

$$Y_t - \mu = \sum_{i=1}^{p} \alpha_i(Y_{t-i} - \mu) + Z_t + \sum_{j=1}^{q} \beta_j Z_{t-j}, \qquad (3.1)$$

where Y_t is the number of daily new SARS cases at day t, μ is the mean and Z_t is white noise with mean zero and variance σ^2.

Using the first 30 observations, we computed the sample autocorrelation coefficients (ACF) and the sample partial autocorrelation coefficients (PACF). The

ACF and PACF showed the process had a strong autocorrelation of lag one and a significant partial autocorrelation of lag one. These features in ACF and PACF indicated an ARMA(1,0) (AR(1)) may be applicable. Similar results were also true for the number of daily new cases at different lengths. These plots are not shown in the paper and they are available up on request from the author.

After fitting equation (3.1) by maximal likelihood method via S-plus based on the first 30 observations, we got $\hat{\alpha}_1$=0.9584 with standard deviation 0.0028, after removing the mean from the time series. The estimate of coefficient α_1 was significantly different from zero with p-value < 0.0001. A more meaningful test would be Dickey-Fuller test on unit root (Fuller 1976), which led to models in next section since the test showed that the estimated coefficients were not significant different from 1 (Fuller 1976). Similar autocorrelation structure was found for the log transformed SARS time series. Using the AR(1) model on the SARS time series, we can forecast the number of SARS cases at k days ahead. For example, for a one-day ahead forecasting, we obtained the mean value 21 with standard deviation 22. Hence, the 95% confidence interval would be (0, 64). The lower bound would be negative from the computation, which was replaced by 0. Results from models on observations with different lengths are tabulated in Table 1. The one day ahead forecasting since day 22 are shown in Figure 1.

4. Random Walk

In Section 3, we assumed that the observed daily new SARS cases were stationary and we fitted AR(1) model to the demeaned time series. The mean value of the one day ahead forecasting seemed reasonable. However, the standard deviation estimated from the process was relatively large and resulted nonpositive lower bound for the 95% confidence intervals for many one day ahead projections (Table 1). The large values of the estimates of the standard deviations were mostly due to the great variation at the beginning of the daily time series of the SARS cases. One popular way to stabilize the variation is to take log transformation of the time series. The log tranformed time series of the first 30 observations are shown in Figure 2(a).

From Figure 2(a), one can see that the trend persists. In this section, we assumed the trend was resulted from nonstationary process. Differencing the time series is an effective way to remove a stochastic trend. That is, we let X_t=log Y_t-log Y_{t-1}. This first order difference could remove the linear trend. Figure 2(b) is the plot of X_t. The sample autocorrelation and the sample partial autocorrelation coefficients are plotted in Figure 2(c) and Figure 2(d), respectively. These two plots of ACF and PACF showed that X_t was not serially correlated. Hence, we

may use a simple ARIMA(0,1,0) (random walk) model for $\log Y_t$:

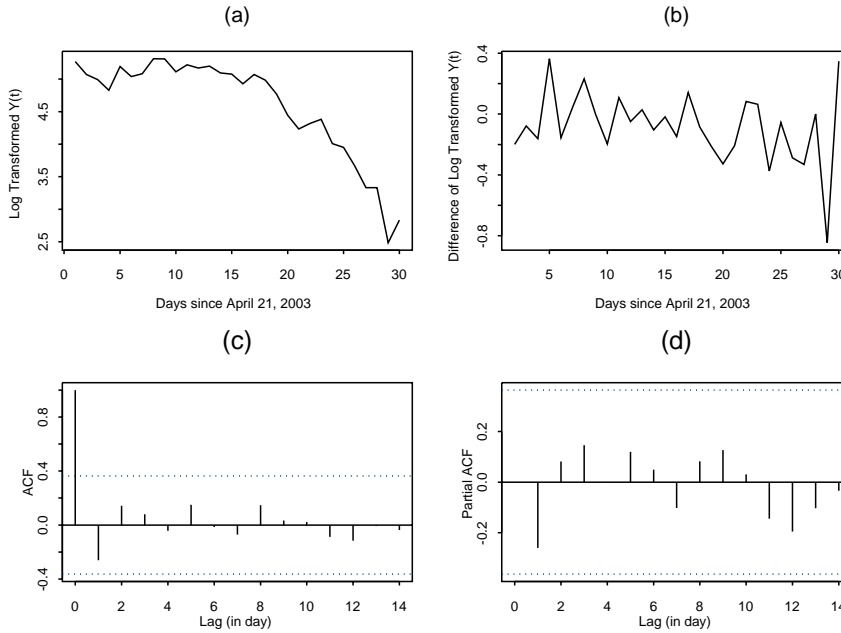$$\log Y_t = \log Y_{t-1} + \mu + Z_t, \tag{4.1}$$

Figure 2: (a) The log-transformed daily new SARS cases in mainland China from April 21 to May 20, 2003, (b) the first order differences of the log-transformed time series, (c) the sample autocorrelation coefficients, (d) the Sample partial autocorrelation coefficients.

where Z_t is a white noise with mean zero and variance σ^2 and $X_t = \mu + Z_t$. This random walk model has been widely used in many diversified fields such as in modeling financial markets (Tsay 2002) and monitoring clinical trials (Lan and Wittes 1988). Using model (4.1), we can construct a k day ahead prediction of $\log Y_t$ and its 95% confidence intervals. Equivalently, the prediction of Y_t and its 95% confidence intervals. For example, for a one day ahead forecasting for $\log Y_{t+1}$ at day $t+1$, the 95% confidences interval would be

$$(\log Y_t + \mu_t - 1.96\sigma_t \quad \log Y_t + \mu_t + 1.96\sigma_t),$$

where σ_t is the sample standard deviation of Z_t up to day t. From our calculation, at day 30, $\hat{\mu}_{30}$=-0.0840, $\hat{\sigma}_{30}$=0.2374 and the one day ahead prediction was 16 with the 95% confidence interval (10,25). The 95% confidence intervals of the one day ahead forecasting are shown in Figure 3 for days after May 12 (day 22). The numerical results are tabulated in Table 2.

Table 2: The estimates of the parameters in the log-transformed random walk model and the forecasting of the daily new SARS cases: $\hat{\mu}$ the estimate of the mean in the log scale, $\hat{\sigma}$ the estimate of the standard deviation of the innovation in log scale, (a) one-day ahead forecast, (b) the 95% confidence interval of the one-day ahead forecast, (c) actual one-day ahead observation.

Day	$\hat{\mu}$	$\hat{\sigma}$	(a)	(b)	(c)
22	-0.0453	0.1673	72	(52,99)	80
23	-0.0403	0.1649	77	(56,106)	55
24	-0.0548	0.1756	52	(37,73)	52
25	-0.0548	0.1717	49	(35,69)	39
26	-0.0642	0.1744	37	(26,51)	28
27	-0.0744	0.1787	26	(18,37)	28
28	-0.0717	0.1759	26	(18,37)	12
29	-0.0993	0.2264	11	(7,17)	17
30	-0.0840	0.2374	16	(10,25)	12
31	-0.0928	0.2382	11	(7,17)	26
32	-0.0648	0.2811	24	(14,42)	20
33	-0.0710	0.2787	19	(11,32)	34
34	-0.0528	0.2937	32	(18,57)	16
35	-0.0734	0.3132	15	(8,27)	8
36	-0.0910	0.3258	7	(4,14)	9
37	-0.0852	0.3230	8	(4,16)	4
38	-0.1049	0.3401	4	(2,7)	3
39	-0.1097	0.3368	3	(1,5)	7
40	-0.0852	0.3660	6	(3,13)	1
41	-0.1317	0.4659	1	(0,2)	2

Figure 3: The One-day Ahead Forecast and Its 95% Confidence Interval Based on the Random Walk Model for the Daily New SARS Cases in Mainland China from May 12 to May 31, 2003.

5. Growth Curve

In sections 3 and 4, the dynamic of the daily SARS cases in China was assumed to be stochastic. As it was mentioned in section 3, one would fit a deterministic trend to the time series such as $Y_t = f(t) + Z_t$, where $f(t)$ might be a linear or nonlinear function. Taking into account the changing rate (daily new cases) of infection, a growth model on the cumulative case would be a good choice. Growth models have been used widely in many fields. (Wiorkowski 1981; Sandland and McGilchrist 1979).
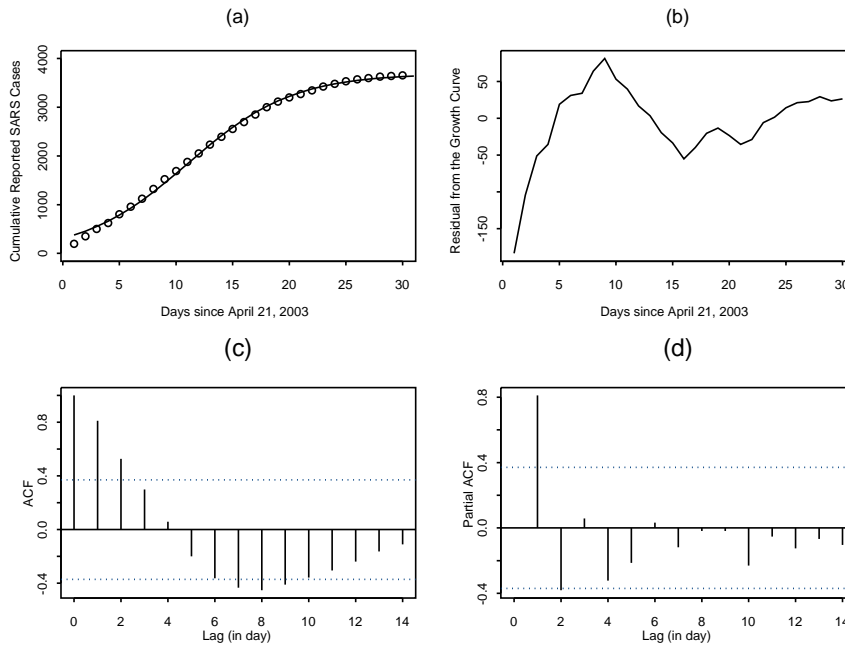

Figure 4: (a)The Fitted Growth Curve to the Cumulative SARS Cases in Mainland China from April 21 to May 20, 2003, (b) the Residuals of the Observed Values to the Fitted Model, (c) the Sample Autocorrelation Coefficients of the Residuals, (d) the Sample Partial Autocorrelation Coefficients of the Residuals.

When the trend was assumed to follow a parametric function, we can use regression techniques to estimate the parameters in the function and establish the model. For the number of SARS cases in China since April 21, we cumulated the daily new cases up to June 3. The cumulative cases are plotted in Figure 5. The data prior to April 21 and after May 31 were not included in our analysis.

Table 3: The estimates of the parameters in the growth curve and the forecasting for the cumulative SARS cases: $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\kappa}$ the estimates of the parameters in the growth curve, $\hat{\alpha}_1$ and the $\hat{\beta}_1$ the estimates of the coefficients in ARMA(1,1) of the residuals, $\hat{\sigma}$ the estimate of the standard deviation of the innovation in ARMA(1,1) of the residuals, (a) one-day ahead forecast, (b) the 95% confidence interval of the one-day ahead forecast, (c) actual one-day ahead observation.

Day	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\kappa}$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\sigma}$	(a)	(b)	(c)
22	3543	11.35	0.2300	0.7664	0.4983	18.01	3393	(3358,3428)	3426
23	3578	11.27	0.2270	0.7797	0.5110	18.23	3468	(3432,3504)	3481
24	3606	11.20	0.2245	0.7716	0.5155	17.49	3508	(3475,3542)	3533
25	3631	11.13	0.2222	0.7888	0.4947	17.39	3561	(3527,3595)	3572
26	3651	11.06	0.2204	0.7817	0.5017	16.81	3590	(3557,3623)	3600
27	3665	11.01	0.2189	0.7809	0.4923	16.35	3615	(3583,3647)	3628
28	3677	10.96	0.2177	0.7821	0.4939	16.02	3642	(3610,3673)	3640
29	3684	10.93	0.2169	0.7689	0.4926	15.61	3646	(3616,3677)	3657
30	3691	10.89	0.2162	0.7716	0.4832	15.37	3667	(3637,3697)	3669
31	3696	10.87	0.2157	0.7664	0.4839	15.04	3674	(3644,3703)	3695
32	3702	10.83	0.2150	0.7752	0.4675	15.19	3705	(3675,3734)	3715
33	3708	10.79	0.2142	0.7752	0.4814	14.93	3717	(3688,3746)	3749
34	3717	10.74	0.2131	0.8057	0.4599	15.58	3756	(3726,3787)	3765
35	3725	10.68	0.2120	0.8033	0.4721	15.25	3760	(3731,3790)	3773
36	3733	10.63	0.2111	0.8141	0.4571	15.05	3771	(3741,3800)	3782
37	3740	10.58	0.2102	0.8193	0.4620	14.84	3779	(3750,3808)	3786
38	3746	10.54	0.2094	0.8199	0.4598	14.58	3782	(3753,3810)	3789
39	3751	10.50	0.2088	0.8216	0.4585	14.37	3785	(3757,3814)	3796
40	3755	10.47	0.2082	0.8270	0.4576	14.22	3794	(3766,3822)	3797
41	3759	10.44	0.2077	0.8246	0.4579	14.00	3792	(3764,3819)	3799

The daily rate of increasing of new SARS cases was high from April 21 to May 10 before it decreased, which suggested an S-shaped parametric function. In our analysis, we selected the autocatalytic (logistic) model (Seber and Wild 1976). The growth function f satisfied the following condition:

$$\frac{df}{dt} = \frac{\kappa}{\alpha} f(\alpha - f), \tag{5.1}$$

where $k > o$ and $0 < f < \alpha$. The growth function f can be solved as

$$f(t) = \frac{\alpha}{1 + e^{-\kappa(t-\gamma)}}. \tag{5.2}$$

Or,

$$f(t) = \frac{\alpha}{1 + \beta e^{-\kappa t}}, \tag{5.3}$$

where $\beta = e^{\kappa\gamma}$.

Using the data of the first 30 days after April 21 and non-linear least squares method (nls function in S-plus), we got $\hat{\alpha}$=3691, $\hat{\beta}$=10.89 and $\hat{\kappa}$=0.2162 with the standard deviations 26.62, 0.51 and 0.005, respectively. All the estimates indicated that the parameters in the model were statistically different from zero with p-values all less than 0.0001. The fitted curve and the actual observations are shown in Figure 4(a). The residuals from the model are shown in Figure 4(b). As it is expected, the residuals from the model would be correlated. From the residual plot, we can see that there were two outliers from the initial observations. After removing these two initial observations, we computed the sample autocorrelation coefficients and the sample partial autocorrelation coefficients, which are presented in Figure 4(c) and Figure 4(d). Based these two ACF and PACF plots, we fitted an ARMA(1,1) model to the residuals e_t:
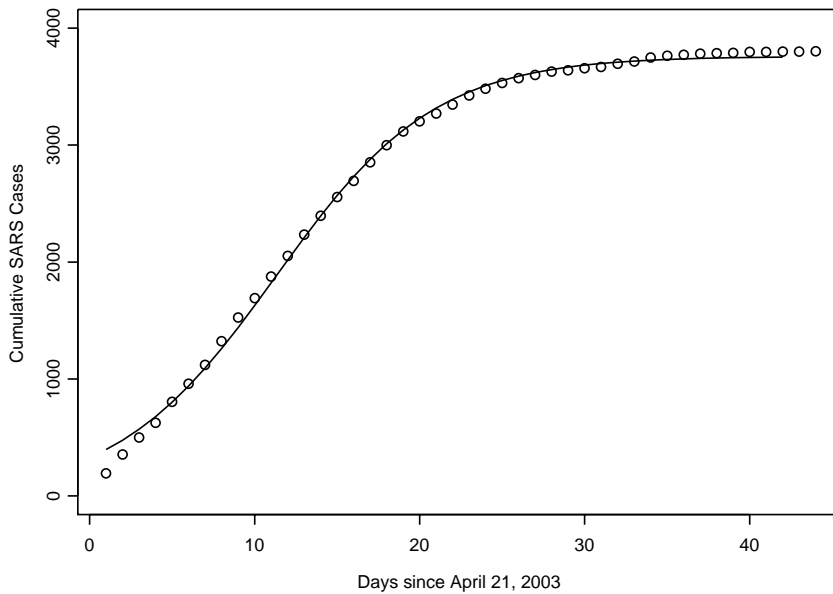

Figure 5: The fitted growth curve to the observed values of the cumulative SARS cases in Minland China from April 21 to May 31, 2003.

$$e_t = \alpha_1 e_{t-1} + Z_t + \beta_1 Z_{t-1}, \qquad (5.4)$$

where Z_t is white noise with mean zero and variance σ^2. The estimate of parameters in equation (5.4) were $\hat{\alpha}_1$=0.7716, $\hat{\beta}_1$=0.4832 and the $\hat{\sigma}_{30}$=15.37. The diagnostic checking of equation (5.4) (not shown here) indicated a reasonable fit

of the model to the data. From the fitted model, we can forecast the residuals of the cumulative model and then construct the 95% confidence intervals. Combining equation (5.3) and equation (5.4) (ignoring the variability of the estimates of the parameters in the growth curve model), we got the one day ahead forecast of the cumulative cases at day 30 being 3667 with the 95% confidence interval (3637,3697). The actual observation was 3669. We repeated this forecast starting from day 22, the results are tabulated in Table 3. Because the large scale in Figure 5, the 95% confidence intervals could not be shown properly in the plot. Hence, we did not include the confidence intervals in Figure 5. The fitted growth curve was based on the data up to day 41 (May 31).

6. Concluding Remarks

Monitoring the health of population is an important task for all countries. In many developed countries, the use of quantitative methods have been a routine practices for governmental agencies and academic institute. However, in many developing countries, there is much more to be done in this field. In this article, we applied three different methods for modeling the dynamic of the SARS cases in mainland China. All these methods involved components of time series analysis. The well developed ARIMA techniques were named after Box-Jenkins' seminar work in the late 1960 (Box and Jenkins 1976). These techniques have been used on modeling influenza in the US (Scuffham 2003) and other infectious diseases (Schnell, Zaidi and Reynolds 1989; Zaidi, Schnell and Reynolds 1989). From Tables 1 to 3, we can compare the forecasting results derived from these three methods with the actual observations. The Box-Jenkins' AR(1) model assumes stationarity of the daily new SARS cases. This was probably not true because of the extraordinary interventions implemented by the Chinese government after April 21. Hence the forecasts from the AR(1) model were usually greater than the actual observed numbers. The random walk model on the log transformed daily SARS cases followed the dynamic of the daily new cases reasonably well and it used mostly the present information to predict the future. The log transformation reduced the variability significantly. For the growth model, it fitted a parametric function globally to the cumulative SARS cases. Although it appeared that the model gave an extremely good fit from the plot, it was mostly due to scale effect. From the results in Table 3, one can see that the variability remained.

The three models studied in this article may be used for different purposes. From the AR(1) model, the differences between the observed and the forecasted could be used as a measure of the effectiveness of the interventions. For a mainly short term forecasting task, the random walk model seemed providing satisfactory results. However, if one wants to assess the relative long-term effect of SARS from a cumulative point of view, the growth curve may be more useful.

Mathematical and statistical models can not only provide description and understanding of the dynamics of diseases, they can also be used in forecasting the future values. To quantify the possible number of cases in the future is fundamental to decision making in public health, especially for highly infectious diseases with a relatively high mortality rate such as SARS.

Prior to systematic time series analysis techniques, there were other mathematical methods used in forecasting the trends of various diseases. Extrapolation of polynomial trends and exponential smoothing were two popular methods (21). Extrapolation a polynomial trend is very unstable because of the nature of the polynomials. Exponential smoothing can be viewed as a special case of ARIMA modeling (Diggle 1990). Recently, nonlinear time series models including chaos have been applied to study the measle dynamics (Bjornstad, Finkenstadt and Grenfell 2003; Finkenstadt and Grenfell 2000). The applicability of chaos theory in monitoring the SARS dynamics is under investigation.

Although the rapid increasing of SARS cases in China in April and early part of May was replaced with a much lower rate and the urgency of forecasting the new cases seemed diminishing, it is generally agreeable that the SARS virus may be able to survive the coming winter and may come back next spring. Hence constant vigilance by governmental officials and academic researchers as well as the general public is essential to keep the virus under control.

SARS is a new disease and there is much more to be learned about its cause and its transmission among the population. The time series models used in this article were mostly based on the history of the observations. When more risk factors are identified from epidemiological study, one can combine these factors in constructing better models to describe the dynamic of the disease and forecast the future. These efforts would contribute to the ultimate goal of eradicating SARS.

Acknowledgments

References

Anand, K. *et el.* (2003). Coronavirus main proteinase (3CL$^{\text{pro}}$) structure: Basis for design of anti-SARS drugs. *Science* **200**, 1763-1767.

Bjornstad, O. N., Finkenstadt, B. F. and Grenfell, B. T. (2002). Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*, **72**, 169-184.

Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Chau, P. H. and Yip, P. S. F. (2003). Monitoring the severe acute respiratory syndrome epidemic and assessing effectiveness of interventions in Hong Kong Special Administrative Region. *Journal of Epidemiology and Community Health*, **57**, 766-769.

Choi, B. C. K. and Pak, A. W. P. (2003). A simple approximate mathematical model to predict the number of severe acute syndrome cases and deaths. *Journal of Epidemiology and Community Health*, **57**, 831-835..

Diggle, P. J. (1990). *Time Series Analysis: A Biostatistical Introduction*, Oxford Science Publication, New York.

Drosten, C. *et al.* (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New England Journal of Medicine*, **348**, 1967-1976.

Finkenstadt, B. F. and Grenfell, B. T. (2000). Time series modelling of childhood diseases: A dynamical systems approach. *Journal of the Royal Statistical Society, Series C*, **49**, 187-205.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, Wiley, New York.

Ksiazek, T. G. *et al.* (2003). Coronavirus associated with severe acute respiratory syndrome. *New England Journal of Medicine*, **348**, 1953-1966 (2003).

Lan, K. K. G. and Wittes, J. The B-value - A tool for monitoring data, *Biometrics*, **44**, 579-585, 1988.

Lipsitch, M., *et al.* (2003). Transmission dynamics and control of severe acute respiratory syndrome, *Science*, **300**, 1966-1970. Published online May 23, 2003; 10.1126/science.1086616.

Marra, M. A., *et al.* (2003). The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399-1404.

Peiris, J. S. M. *et al.* (2003) Coronavirus as a possible cause of severe acute respiratory syndrome. *The Lancet*, **361**, 1319-1325.

Riley, S. *et el.* (2003) Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, **300**, 1961-1966. Published online May 23, 2003; 10.1126/science.1086478.

Rota, P. A. *et al.* (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394-1399. Published online May 1, 2003; 10.1126/science.1085952.

Sandland, R. L. and McGilchrist, C. A. (1979). Stochastic growth curve analysis. *Biometrics*, **35**, 255-271.

Schnell, D., Zaidi, A. and Reynolds, G. (1989). A time-series analysis of gonorrhea surveillance data. *Statistics in Medicine*, **8**, 343-352.

Scuffham, P. A. (2003). Estimating influenza-related hospital admissions in children and adults - A time series analysis. *Disease Management and Health Outcomes*, **11**, 259-269.

Seber, G. A. F. and Wild, C. J. (1989). Nonlinear Regression. Wiley, New York.

Tsay, R. S. (2002). *Analysis of Financial Time Series*, Wiley, New York.

Wang, W. D. and Ruan, S. G. (2004). Simulating the SARS outbreak in Beijing with limited data. *Journal of Theoretical Biology.* **227**, 369-379.

Wiorkowski, J. J. (1981). Estimating volumes of remaining fossil-fuel resources - A critical review. *Journal of the American Statistical Association*, **76**, 534-548.

World Health Organization (WHO 2003) on severe acute respiratory syndrome (SARS): http://www.who.int/csr/sars/en/.

Zaidi, A., Schnell, D. and Reynolds, G. (1989). Time-series analysis of syphilis surveillance data. *Statistics in Medicine*, **8**, 353-362.

Dejian Lai
School of Public Health
The University of Texas Health Science Center
Houston, TX 77030
dlai@sph.uth.tmc.edu

Also: Faculty of Statistics
Jiangxi University of Finance and Economics
Nanchang, China 330013