Modeling the Cumulative Cases from SARS

Renbao Chen, Ph.D.
Department of Finance and Accounting
Graduate School of Business
National University of Singapore
10 Kent Ridge Crescent
Singapore 119260
Email: bizchenr@nus.edu.sg

Ping Wang, Ph.D.
School of Risk Management
St John's University
101 Murray St., New York, NY 10007,
USA
Email: wangp1@stjohns.edu

Abstract

In this paper, we fit the model of $y = \frac{a}{[1+\exp(b-ct)]}$ to the actual data of cumulative cases from the 2003 SARS outbreak. The model proves a good fit for the four areas we examine, China (mainland), Canada, Hong Kong and Singapore, as evident from the high R^2 values and the randomness of the residuals.

In addition, we obtain second order derivatives of our curve fits and propose the use of T^*, the time in days taken to slow the rate of the increase in the number of cumulative SARS cases, as a measure of the effectiveness of control measures in each country.

We then incorporate explanatory variables to link the model parameters to economic and demographic indices of areas hit heavily by SARS. The work enables prediction of the future behavior of SARS or other epidemics should they ever strike again.

1. Background

The Severe Acute Respiratory Syndrome (SARS) epidemic in 2003 affected as many as 29 countries across the world and was the first severe contagious disease to emerge in the 21st century (World Health Organization, 2004a). This new disease sparked numerous researches into the epidemiology of SARS, unveiling transmission dynamics of the disease and the impact of control measures[1]. Even though all human chains of SARS transmission had been declared successfully contained in July 2003, four new cases in Beijing and Anhui Province, China were reported in April 2004 (World Health Organization, 2004b). These new cases, which were subsequently contained, raised serious concerns about the possible future re-emergence of the outbreak.

2. Motivation

Among numerous researches conducted on SARS, a few have attempted to model the cumulative incidence of cases attributed to SARS during the global outbreak in 2003, but very few have carried out the study on a global scale. If the SARS outbreak were to re-emerge, such a model will be useful in predicting the cumulative cases and deaths that may arise. The emergence of the avian influenza in early 2004 has sparked warnings of a possible pandemic. The World Health Organization (WHO) asserted that it could take years to eliminate the avian influenza from the environment while recent findings reported its widespread occurrence across poultry and wild birds in Asia (2004c). The model derived from the SARS outbreak will be helpful to examine other diseases such as the avian influenza.

In this study, we examine SARS data from Singapore, Canada, Hong Kong, Taiwan and China and propose a model for the SARS outbreak. Following which, we examine the validity of our model using these data and analyze the resulting curve fits using R^2 and residual analysis. Economic and demographic explanatory variables are then introduced to enable predictive function of the model. Implications of the findings will be discussed, before we end off with the limitations of this study.

[1] See Lipsitch et al. (2003) and Ghani et al. (2005)

3. Data and Methodology

We gather data from the World Health Organization (2005), the Ministry of Health Singapore (2004) and the Public Health Agency of Canada (2003) on the cumulative SARS cases in five regions: Singapore, Canada, Hong Kong, China (mainland) and Taiwan. These regions were chosen based on their high number of cumulative cases ranging from 238 (in Singapore) to 5327 (in China) (World Health Organization, 2004a), as well as the availability of the data required.

After examining the data, we find the figures on the number of cumulative cases in Taiwan to be unreliable for modeling due to the many errors in the initial estimation of probable SARS cases. Out of over 600 cumulative cases, 325 cases (about 50%) were discarded (World Health Organization, 2004a). Because these discarded cases are a substantial percentage of the total, and we do not have data on the dates these discarded cases were first labeled as probable SARS cases, we are not able to make a reliable model based on the available data. Hence, Taiwan is omitted from this project.

Thanks to the availability of epidemic curves of the SARS outbreaks in Singapore and Canada from their public health authorities, the graphs of cumulative cases that we generate for these two countries represent the number of cumulative cases by symptom onset date. In contrast, as epidemic curves of the SARS outbreaks in Hong Kong and China are unavailable in local reports, we turn to the cumulative SARS cases reported by WHO, though one must be reminded of the possible time lag between symptom onset and reporting of a new probable SARS case.

3.1 Fit each region separately with no explanatory variables

With varying periods of outbreaks, the start dates of t=0 are fixed as the onset date of the first probable SARS case in each region.

The ending date for the regression is set at 20 days after the onset of the last probable SARS case as defined by WHO as the date marking the containment of the disease and a break to the

human-to-human transmission chain of SARS. Table 1 summarizes the various dates of the first and last cases, as well as the end dates of our regression:

Table 1: Summary of First and Last Cases and Ending Dates for Regression

Regions	Cumulative cases		
	First case, t=0	**Last case**	**End date for regression**
Singapore	Feb 25, 2003	May 5, 2003	May 25, 2003
Hong Kong	Feb 15, 2003	May 31, 2003	Jun 20, 2003
Canada	Feb 23, 2003	Apr 19, 2003	May 9, 2003
	Apr 20, 2003	Jun 12, 2003	Jul 2, 2003
China	Nov 16, 2002	Jun 3, 2003	Jun 23, 2003

Researchers have contributed in the area of epidemic modeling following the 2003 SARS outbreak. Brauer (2005) reviews and updates the so-called Kermack-McKendrick approach and its application to general epidemic models. Hsu and Roeger (2007) introduce "a basic reproductive number" to establish the relation between the final susceptible population and the initial susceptible population. Massad et al. (2005) design a deterministic mathematical model of the susceptible-infected-recovered type to analyze the impact of control measures against SARS and validate their model using the SARS data of Hong Kong, China and Toronto, Canada. A similar research is conducted by Zhang (2007) for the analysis of SARS data of the mainland of China.

Of previous works, several papers (for example, Brauer (2005), Zhang (2007) and Rozema (2007)) identify that the development pattern of SARS in terms of cumulative cases can be modeled by the logistic function $y = \frac{a}{1+\exp(b-ct)}$. These works are all focused on examination of SARS data of one single region. The current paper makes use of the information from multiple regions.

With the accumulation of data, graphs of the cumulative cases are plotted against time. The plotting is done to determine the approximate shape of the SARS trends. Plot points corresponding to data which are not reported on certain days are classified as missing data and these points are eliminated. Also, upon examination of the epidemic curve of the SARS outbreak in Canada, we spot two waves of the outbreaks, one starting on February 23, 2003 and the other starting around April 20, 2003. The first wave was started by a woman returning to Toronto from Asia while the second phase was reportedly evolved among nurses and visitors who had been in close proximity to an elder patient first admitted to hospital for a fractured pelvis (Centers for Disease Control and Prevention, 2003). As a result, we generate two curve fits for the cumulative SARS cases to correspond to the two separate outbreaks in Canada.

Following previous research and the hint revealed by the curve of raw data, we adopt logistic curve to model the cumulative number of cases. Specifically, let y be the number of cases by time t, M the ultimate cumulative number of cases, then M-y denotes the number of cases yet to be developed. The product of y and M-y hence is the total number of possible contacts between the two groups. Assuming the growth rate of case number is proportional to the number of contacts between the two groups we have

$$\frac{dy}{dt} = \frac{r}{M} y(M - y)$$

where r is a constant. Solving the equation for y we obtain the logistic curve.

For each region the model is

$$y_{i,t} = \frac{a_i}{1 + \exp(b_i - c_i t)} \qquad (1)$$

where i denotes the region and t the number of days after the first onset.

We perform preliminary analysis to establish the best fitting curves to the observed data through non-linear regression. Residuals and goodness-of-fit statistics are obtained to further analyze the validity of our model.

3.2 Fit all regions altogether incorporating explanatory variables

As mentioned earlier, previous research utilizing actual SARS data share a common disadvantage by concentrating on one region and isolating the SARS behavior of one region from another, thus neglecting the fact that the 2003 SARS outbreak took place in the 21st century, when an infected person can travel thousands of miles in hours from one country to another and spread the disease before the onset of any symptoms.

To tackle this issue Hufnagel, Brockmann and Geisel (2004) introduce a probabilistic susceptible-infected-recovered model to describe the worldwide spread of infectious diseases in today's closely-connected world.

In this paper making use of the SARS data from four regions we incorporate economic and demographic indices of each region in the analysis of the behavior of the specific disease. Specifically with the region index i and time variable t, the parameters a_i, b_i and c_i in the equation $y_{i,t} = \frac{a_i}{1+\exp(b_i - c_i t)}$ for each region are rendered as functions of demographic and economic indices, namely, population, human development index (HDI), health manpower, and a dummy variable indicating a second wave hitting an area shortly after the first wave. In detail,

$$\begin{aligned} a_i &= \boldsymbol{\alpha}^{\mathbf{T}} * \mathbf{X}_i = \alpha_0 + \alpha_1 X_{1i} + \ldots + \alpha_k X_{ki} \\ b_i &= \boldsymbol{\beta}^{\mathbf{T}} * \mathbf{X}_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} \\ c_i &= \boldsymbol{\gamma}^{\mathbf{T}} * \mathbf{X}_i = \gamma_0 + \gamma_1 X_{1i} + \ldots + \gamma_k X_{ki} \end{aligned} \qquad (2)$$

where $\boldsymbol{X_i}$ is the vector of all explanatory variables for region i, including the constant, and coefficient vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are common to all regions. The estimate of the coefficient vectors are achieved by minimizing the overall sum of squared errors of the estimated case number from the actual ones. That is,

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \underset{\alpha,\beta,\gamma}{\arg\min} \left(\sum_i \sum_t \left(y_{i,t} - \hat{y}_{i,t} \right)^2 \right).$$

The advantage of fitting all regions together and incorporating explanatory variables over separate fitting include 1) the behavior of epidemic in a region is related to its overall economic, demographic and public health status, which are key in the battle against the disease; 2) the common coefficients apply to all regions observed, and potential to other countries. This makes it possible to predict the pattern of next SARS outbreak should it ever strike again. More generally the approach can be employed in the analysis of other epidemics. Table 2 lists the explanatory variables employed in the analysis and data sources.

Table 2: Summary of Explanatory Variables and Data Sources

Variable	**Description and Source**
Population	Population in millions, logarithmic value is used. (WHO, 2006 a, b)
Population density	Population per square kilometer, logarithmic value is used. (WHO, 2006 a, b)
HDI	Human development index reported by United Nations. A relative measure used to determine whether a country is developed, developing, or underdeveloped, it is an equally weighted function of the achievements in three basic dimensions of human development: general health care, education, and standard of living. Original value varies from 0 to 1. HDI*100 is used in the analysis. (United Nations Development Programme, 2005)
Manpower	Number of medical professionals (doctors, nurses, pharmacists) per thousand of population. ((WHO, 2006 a, b)
2^{nd} Wave	Indicator of the second wave outbreak. It equals 1 for the 2^{nd} wave in Canada and 0 otherwise.

4. Results and discussion of separate fitting with no explanatory variables

From Figure 1, we observe a logistic distribution in the cumulative cases, which is consistent to previous research. This observation is reasonable as the tapering-off of a logistic curve towards the end of the period would correspond to the depletion of those susceptible to SARS and the implementation of control measures to contain the spread of the disease.

We fit the model below to the data of each region by conducting non-linear regression analysis:

$$y_{i,t} = \frac{a_i}{1 + \exp(b_i - c_i t)} \ ,$$

Where y_t = cumulative SARS cases at time t, t = time in days from the corresponding start dates as mentioned in the Methodology and *a, b, c* = model parameters.

Figure 1. Collected Data

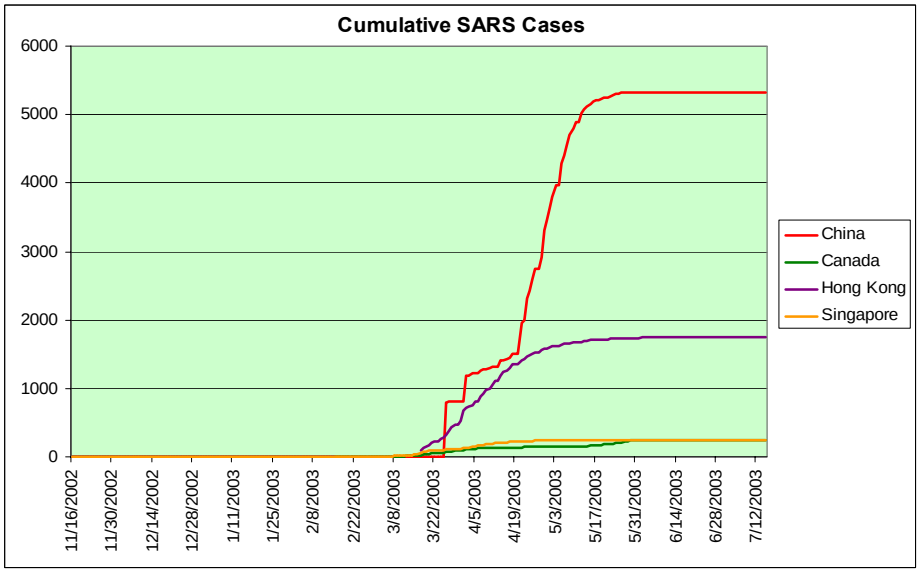


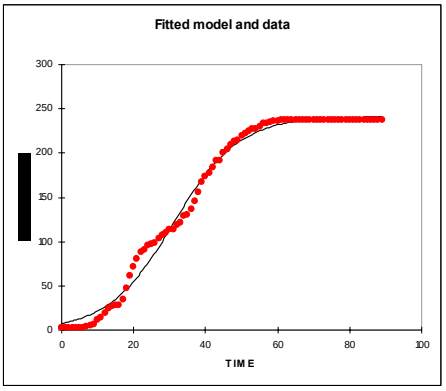
Table 3 summarizes curve fits we have derived. For a graphical representation of the plotted points and the curve fits, See Figure 2.

Table 3. Equations of Curve Fits from Non-linear Regression, by Region

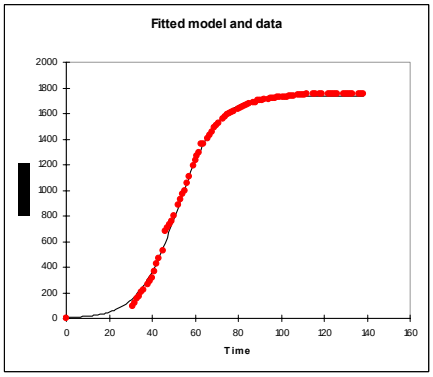
Region	Cumulative cases
Singapore	$y = \frac{241.3513}{1+\exp(3.4293-0.1104t)}$
Hong Kong	$y = \frac{1738.3818}{1+\exp(5.7134-0.1097t)}$
Canada (1st wave)	$y = \frac{142.8214}{1+\exp(5.1354-0.1644t)}$
Canada (2nd wave)	$y = \frac{107.8214}{1+\exp(5.3650-0.2294t)}$
China	$y = \frac{5436.6893}{1+\exp(15.8454-0.0992t)}$

Figure 2. Fitted Model and Data

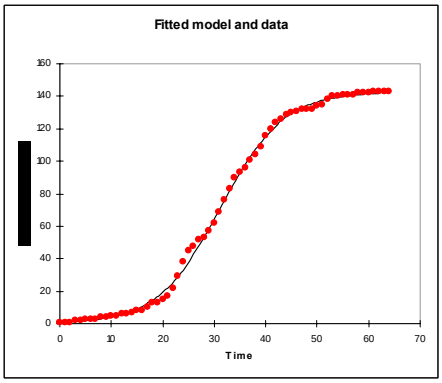
2a. Singapore



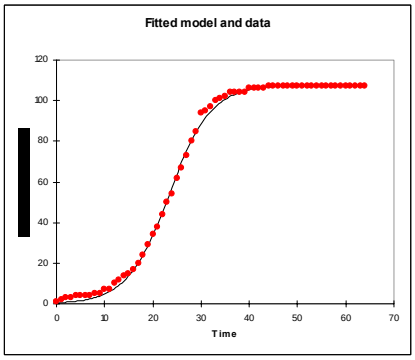
2b. Hong Kong



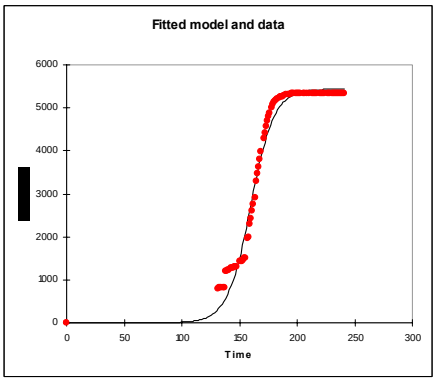
2c. Canada (first wave)



2d. Canada (second wave)



2e. China



4.1 Pseudo-R^2 and Analysis of Residuals

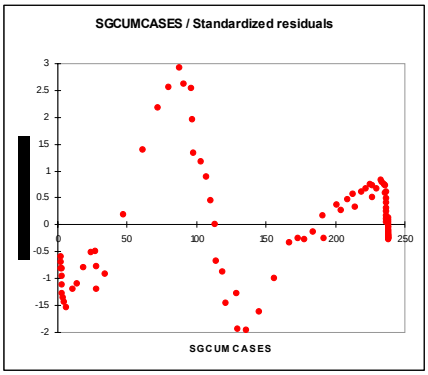
Before we enter into a discussion of our results, it is necessary to analyze the goodness-of-fit of the nonlinear regression equations in Table 3 to determine if they sufficiently reflect the observed data in each country. Firstly, we shall observe the plotted residuals of our curve-fits (See Figure 3). Residuals reflect the difference between the observed data and the predicted data (from the nonlinear regression equation) at each data point. When the curve is a good fit, it will reflect random errors and appear to behave randomly in a graphical plot. There should be equal probability that the predicted data occurs above and below the actually *y* variable for the errors to be random. As we need to make a fair comparison between five different sets of residual data with different spreads in their standard deviations, we need to standardize the spreads of the raw data onto the same scale. The standardized deviations are given by the equation:

Standardized Residual = (Raw Residual – Mean Raw Residual)/ Standard Deviation.

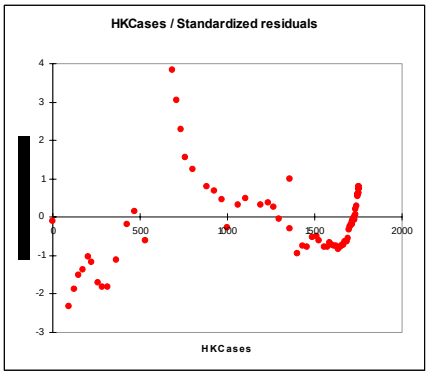
The standardized residuals of the dependent *y* variables are graphed in 3a to 3e of Figure 3. The residuals appear to be randomly distributed about zero, indicating that the model describes the data well.

Figure 3: Residuals

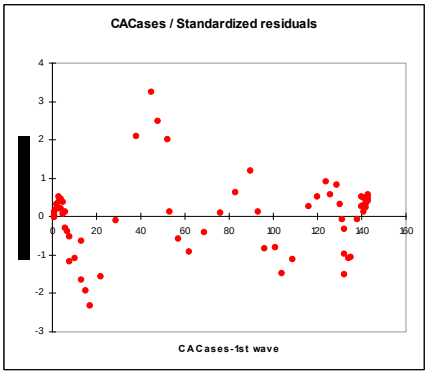
3a. Singapore



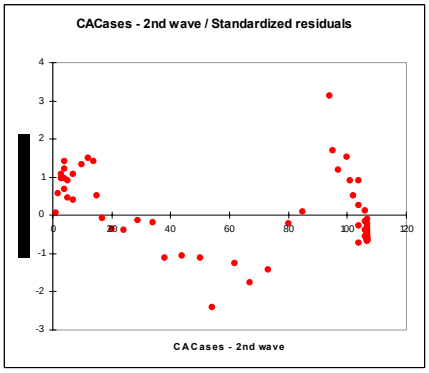
3b .Hong Kong



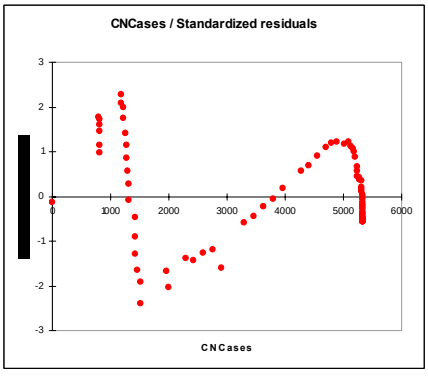
3c. Canada (first wave)



3d. Canada (second wave)



3e. China



To further analyze our fits, the following goodness-of-fit statistics are obtained and summarized in Table 4.

Table 4 Goodness of fit statistics

Region	SSE	Pseudo-R^2
Singapore	5823.8068	0.9922
Hong Kong	55,791.4780	0.9980
Canada Wave 1	320.9885	0.9985
Wave 2	175.7540	0.9988
China	5,975,758.772	0.9822

From our data, the spread of SSEs are large, with SSE of 5,975,759 for China, indicating a possible less-than-desired fit. However, a disadvantage of SSE is that there is no basis of comparison for the obtained values. The relative size of the SSE is inevitably affected by the number of data points in each non-linear regression, and the relative size of the data values of one data set compared to another data set. As such, we calculate the pseudo-R^2 values to derive a better comparative analysis of the curve fits.

R^2 of linear regression provides a standardized measure of goodness-of-fit. Application of this measure to nonlinear models, however, may lead to a measure that may lie outside the [0,1] interval. A measure, relatively closely corresponding to R^2 in the nonlinear case is:

$$Pseudo\text{-}R^2 = 1 - SS(Residual)/SS(Total_{Corrected}).$$

For more details of non-linear regression see, for example, http://www.ats.ucla.edu/STAT/sas/library/SASNlin_os.htm. As seen from Table 4, our curve fits show high values of pseudo-R^2, supporting the suitability of the derived model.

4.2 Analysis of results

Following the reaffirmation of our curve fits in Section 4.1, we proceed to analyze the implications of our results.

4.2.1 Model and parameters

The logistic model $y = \frac{a}{1 + \exp(b - ct)}$ proposes a trend in the cumulative number of cases which initially increases at an increasing rate and then slows down after a particular point in time. The initial exponential increase can be attributed to the unfamiliarity of this new disease, leading to the subsequent rapid spread of it due to either a lack of or inappropriate knowledge and measures to curb the disease. The rate of increase in the initial stage of the outbreak is affected by, among other factors, population, population density of the region (i.e. a measure of the number of people in contact with probable SARS cases) and the initial attack size (i.e. the number of index cases), since this disease is contagious.

Subsequently, the increase in the cumulative cases/deaths slows down due to new-acquired knowledge of SARS and the implementation of adequate control measures. As countries learn more about the killer respiratory disease, they are able to react more efficiently and thus identify and quarantine patients much faster. Loss control reduces the frequency and severity of SARS cases and can be carried out through loss prevention (e.g. avoidance of travel to SARS infected areas, donning of protective masks and adopting hygienic living habits, home-quarantine of contacts with probable SARS cases and travelers returning from SARS-affected countries) and loss reduction (e.g. daily monitoring of SARS symptoms and immediate report when SARS-like symptoms appear).

On closer examination, we realize parameter a (i.e. the numerator) of our model reflects the total number of cumulative cases. This observation is intuitive as our model is a logistic equation which approaches an asymptotic value as time tends to infinity. Consequently, the numerator of the equation (which reflects that asymptote) will have to point toward the total cumulative cases within each region.

Holding *t*, *a* and *c* constant, changing parameter *b* results in a roughly parallel shift of the curve along the horizontal axis of graphs (i.e. it determines the position of the curve in the x-direction).

Parameter *c* varies directly with the number of cumulative cases. When *c* increases, the number of cumulative cases increases for each value of *t* (i.e. the curve fits become steeper with increasing *c* values). The slope represents the rate of increase of the cumulative values. When the slope increases and the first and last cumulative values are fixed, the cumulative values take a shorter time to reach the final value. In addition, parameter *c* is a coefficient of time *t*. Hence, we postulate that parameter *c* represents the time period over which SARS cases appear. The following table compares the time periods and *c* values for the four regions.

Table 5. Comparison of Parameter c and Time Periods

Region	Date onset first probable case	Date onset last probable case	Time period of cases (days)	*c* value
Singapore	February 25	May 5	70	0.1104
Hong Kong	February 15	May 31	106	0.1097
Canada (wave 1)	February 23	April 19	56	0.1644
Canada (wave 2)	April 20	June 12	54	0.2294
China	November 16	June 3	200	0.0992

As shown in Table 5, parameter *c* corresponds inversely with the length of time over which the SARS cases appeared. The time period in increasing order for cumulative cases is 54 (Canada wave 1), 56 (Canada wave 2), 70 (Singapore), 106 (Hong Kong), 200 (China). The corresponding *c* values exhibit an inverse order with the time period.

4.2.2 Indicator of the effectiveness of control measures

By obtaining the second order differential equation of our curve fits and equating it to zero, we are able to determine the time ($t = T^*$) at which the growth rate of cumulative cases stops increasing and starts to decrease. This marks the point in time where control measures slow the spreading rate effectively and hence signals the start of the containment of the epidemic. As

such, the number of days from *t=0* to the time *T** has a direct relationship with the responsiveness in implementation and the effectiveness of control measures in the country. Table 6 shows the value of *T**:

Table 6. Time Taken to Slow the Rate of Increase of Cumulative Cases

Region	***T****	**Total days of outbreak**
Singapore	31.0	70
Hong Kong	52.2	106
Canada		
First wave	31.2	56
Second wave	23.4	54
China	159.7	200

Currently there are no indicators to the efficacy of SARS control measures. *T** is a suitable gauge to the speed at which control measures are adopted and the effectiveness of these measures, with a smaller value indicating higher responsiveness and effectiveness. Based on this indicator, Singapore was the most efficient in containing the disease since it only took 31 days for the increase in cumulative cases to slow down. Singapore has been praised by WHO officials for their serious attitudes in controlling SARS (Lian He Zao Bao, May 1 2003). For example, Singapore is the first country among these SARS affected regions in forming a national team to fight for SARS led by the Prime Minister, first country decided to close all primary and high schools, and many more similar measures in managing SARS risk.

5 Results and discussion of fitting with explanatory variables incorporated

5.1 Results

Non-linear regression is conducted utilizing data from all four regions and incorporating explanatory variables, as described in Section 3.2. To compare the results with that of separate fitting, sum of squared errors is calculated. As pointed out earlier, a single value of SSE is affected by the number of data points and the relative size of the data values of the data set, thus is not an ideal measure of goodness of fit. However SSE does serve the purpose to compare two or more approaches applied to the same data set.

Values of sum of squared errors are reported below to compare the goodness of fit of cumulative case number for each region when fitted with and without explanatory variables.

Table 7. Comparison of Sum of Squared Errors

Region	Each region fitted separately, without explanatory variables	All regions fitted altogether, incorporating explanatory variables
Singapore	5823.8	5853.9
Hong Kong	55791.5	56260.0
Canada (1st wave)	321.0	432.7
Canada (2nd wave)	175.8	277.9
China (mainland)	5975758.8	5976849.6
Total	6037870.9	6039674.1

Note: The second column reports the SSE for each region when fitted separately, while the third column the SSE when all regions are modeled altogether by a single regression

.

We see that by incorporating explanatory variables the fitting is almost as well as separate modeling, in the sense that the total sum of squared errors for all five series of cumulative case number stays virtually unchanged. A close examination reveals that SSE for Singapore is about the same as that of separate fitting, the value of SSE jumps by about 50% for Canada (both waves) compared to their original relative small values, while the SSE for Hong Kong and China experience less than one percent increase.

Table 8 reveals the value of parameters *a*, *b* and *c* for each region under two fitting approaches. We can see all parameter values are very close under different fitting strategies. By modeling the SARS case numbers of all regions together we suffer a little loss of goodness of fit but achieve a lot more. The major advantage of incorporating explanatory variables is that we are able to identify factors that may have affected the spread-out pattern of the disease. This has more significant value than just a good fitting in terms of the ability to predict the pattern of another strike in the same or different country.

Table 8. Parameter Values from Different Fitting Methods

	Each region fitted separately without explanatory variables			All regions fitted together, with explanatory variables		
Region	*a*	*b*	*c*	*a*	*b*	*c*
Singapore	241.3513	3.4293	0.1104	242.2040	3.4478	0.1105
Hong Kong	1738.3818	5.7134	0.1097	1739.3491	5.7607	0.1106
Canada (1st wave)	142.8214	5.1354	0.1644	144.3380	5.2371	0.1648
Canada (2nd wave)	107.8214	5.3650	0.2294	109.1387	5.4703	0.2293
China (mainland)	5436.6893	15.8454	0.0992	5441.1477	15.8627	0.0993

Table 9. Effects of Explanatory Variables ← Not mentioned in the paper

	Coefficient vector α in the *a*-equation		Coefficient vector β in the *b*-equation		Coefficient vector γ in the *c*-equation	
Variable	Estimate	Std Error	Estimate	Std Error	Estimate	Std Error
Constant	7328.6163	174.3252	25.0101	0.0105	-0.5907	0.0004
Population	1164.3580	4.1787	2.1413	0.0018	0.0124	0.0001
Population Density	1579.4615	0.7744	1.9134	0.0032	-0.0137	0.0001
HDI	-193.1080	1.9217	-0.3966	0.0001	0.0084	0.0001
Manpower	497.7438	1.8014	0.7484	0.0014	-0.0065	0.0001
2nd Wave	-35.1992	0.2253	0.2332	0.0389	0.0644	0.0016

The results suggest that all the explanatory variables play a role explaining the behavior of the dread disease. However caution must be exercised when interpreting and applying the results to other countries. There are only five time series of four regions in the sample for analysis although the total number of data points is fairly large.

5.2 Analysis of the results

In linear regression models, the effect of explanatory variable on dependent variable is straightforward to tell as revealed by the sign of the estimated coefficient. In non-linear regression, the relationship between dependent and explanatory variable, however, is not always ready to be revealed by the sign of coefficient only due to the complexity of the regression equation. In our analysis, as can be seen in equations (1) and (2) the explanatory variables appear in both denominator and numerator of the cumulative case number *y,* which suggests more complex association between dependent and explanatory variables. It is worthwhile to further analyze the relationship.

Let's drop the subscript *i* and *t* from the equations (1) and (2) when taking partial derivative of *y* with respect to explanatory variable X_k. Easy calculus shows that

$$\frac{\partial y}{\partial X_k} = \frac{\alpha_k (1 + EXP) + a \cdot EXP \cdot (-\beta_k + \gamma_k t)}{(1 + EXP)^2}$$

where $EXP = \exp(\beta^T X - \gamma^T X \cdot t) = \exp(b - ct)$, and α_k, β_k, and γ_k are the coefficients of variable X_k in the three equations of (2).

When the value of X_k increases, how the dependent variable *y* will respond is jointly determined by the coefficients α_k, β_k, and γ_k, the value of parameters *a*, *b* and *c* for each region, and the time variable *t*. Complexity is expected. Table 10 displays the sign of the partial derivative of *y* with respect to each of the explanatory variables evaluated for each region.

Table 10. Sign of Partial Derivative of Cumulative Case Numbers with Respect to Explanatory Variables, by Region

Explanatory variable X_k	Singapore	Hong Kong	Canada (1st wave)	Canada (2nd wave)	China
Population	positive	negative for smaller t, then stay positive	positive	positive	negative for smaller t, then stay positive
Population density	positive	negative for smaller t, then stay positive	positive	positive	negative for smaller t, then stay positive
HDI	negative	positive for smaller t, then stay negative	negative	negative	positive for smaller t, then stay negative
Manpower	positive	negative for smaller t, then stay positive	positive	positive	negative for smaller t, then stay positive

Common sense tends to predict that cumulative case number would increase with population, population density, and decrease with human development index (HDI) and health manpower. However it is not always the case with our data set as reflected in Table 10. For Singapore and Canada the relationship between the cumulative case number and the first three explanatory variables are consistent with the common expectation, although the effect of Manpower is surprising. For Hong Kong and the mainland of China none of the partial derivatives keeps its sign throughout the time period in analysis ($0 \leq t \leq 241$), rather the sign varies with the value of time t. A revisit to the data set reveals that the characteristic common to Hong Kong and China which distinguishes them from Singapore and Canada is their rather large ultimate case number. However it is too impulsive to make any conclusion yet due to the small number of region in the data set. Further study is in call when more data is available.

6. Limitations

Although our model proves to be a good fit to the observed data, and it is useful to predict the pattern of future SARS outbreak, there are certain limitations regarding its applicability.

Firstly, we have identified certain explanatory variables in Section 5, however, further research will still be needed to validate them and expose other underlying factors that may drive the parameter values.

Secondly, the data we have gathered from various sources was inconsistent, especially when comparing those from the World Health Organization with the local authorities and other publications. Attempts have been made to reconcile the data which sometimes meant the selection of one source of data over another. In addition, data from WHO may not reflect the actual cases on the respective days due to the time lag in the receipt and publishing of reports, as well as the lapse of time from the onset of symptoms to the report of the case.

Moreover, SARS is a diagnosis of exclusion (WHO, 2003b), meaning its presence cannot be established with complete confidence from examination or testing. Diagnosis is therefore by elimination of other reasonable possibilities (Wikipedia, 2005). This definition has led to the presence of discarded cases in Hong Kong and China. As a result, there are inherent inaccuracies with the curve fits as knowledge of when these discarded cases were first reported was not available. Nevertheless, this limitation is not significant enough to affect the overall model.

7. Conclusions

In this paper, we fit the model of $y = \frac{a}{[1+\exp(b-ct)]}$ for the cumulative cases of SARS. The model proves a good fit for the five case number series from four regions, namely China, Canada, Hong Kong and Singapore. In addition, we obtain the second order derivatives of our curve fits and propose the use of T^*, the time in days taken to slow the rate of the increase in the number of cumulative SARS cases, as a measure of the efficacy of control measures in each country.

We then further incorporate explanatory variables in modeling all five case number series at the same time. Fitting with demographic variables enables the model to be useful in predicting the model parameters of future outbreak of the disease. In the midst of escalating fears of worldwide bird flu pandemic, research can be conducted to determine the applicability of our model to other disease.

Despite limited modeling and epidemiological knowledge, we have put forth a model well-fitted to the actual epidemiological data of SARS. We believe that using the model and methods introduced in this paper and with more data available, researchers can carry out more rigorous tests, refine the model to bring the world an improved understanding of the SARS epidemiology.

References

Brauer, Fred (2005), *The Kermack-McKendrick Epidemic Model Revisited*, Mathematical Biosciences 198, 119-131.

Centers for Disease Control and Prevention (2003), *Update: Severe Acute Respiratory Syndrome ----Toronto, Canada,* Morbidity and Mortality Weekly Report, 52(23), 547-550. (available at http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5223a4.htm)

Ghani, A. C., Donnelly, C. A., Cox, D. R., Griffin J. T., Fraser, C., Lam, T. H. et al., (2005), *Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease*, American Journal of Epidemiology, 162 (5), pp.479-486.

Hsu, Sze-Bi, Roeger, L-I W. (2007), *The Final Size of a SARS Epidemic Model Without Quarantine,* J. Math Anal. Appl., 333, 557-566.

Hufnagel, L, Brockmann, D., Geisel, T, May, R. (2004), *Forecast and Control of Epidemics in a Globalized World,* Proceedings of the National Academy of Sciences of USA, Vol. 101, No. 42, 15124-15129.

Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L. et al. (2003), *Transmission Dynamics and Control of Severe Acute Respiratory Syndrome*, Science Express, May 23 2003, pp. 1-5.

Massad, Eduardo, Burattini, M. N., Lopez, L. F., and Coutinho, F. A. B.(2005), *Forecasting versus Projection Models in Epidemiology: The Case of the SARS Epidemics,* Medical Hypotheses, Vol. 65, 17-22.

Ministry of Health, Singapore (2004), *Severe Acute Respiratory Syndrome: Related statistics,* June 10 2003, [cited 2005 Nov] Available from: http://www.moh.gov.sg/corp/sars/index.html.

Public Health Agency of Canada (2003), Summary of Severe Acute Respiratory Syndrome (SARS) cases: Canada and International, July 4, [cited 2005 Nov], Available from: http://www.phac-aspc.gc.ca/sars-sras/eu-ae/index.html.

Rozema, Edward (2007), *Epidemic Models for SARS and Measles*, The College Mathematics Journal, Vol 38, No. 4, 246-259.

South China Morning Post (2003a), *WHO acts on HK outbreak of pneumonia*, South China Morning Post, 14 March 2003, South China Morning Post Publishers Limited.

South China Morning Post (2003b) Sars *toll reaches 299 as woman loses her fight for life*, South China Morning Post, 18 July 2003, South China Morning Post Publishers Limited.

South China Morning Post (2003c), *Crisis linked to Hong Kong spreads globally. Pneumonia prompts WHO health alert. WHO issues warning*, South China Morning Post, 16 March 2003, South China Morning Post Publishers Limited.

South China Morning Post (2003d), *Sars toll hits 349 as patient dies in Beijing*, South China Morning Post, 29 July 2003, South China Morning Post Publishers Limited.

The Straits Times (2003a), *Hospitals to lift restrictions – last patient out of SARS ward, The Straits Times*, July 17 2003, Singapore Press Holdings Limited.
The Straits Times (2003b), *One more patient dies of SARS*, The Straits Times, 18 June 2003, Singapore Press Holdings Limited.
United Nations Development Programme (2005), *Human Development Report 2005,* New York.
Wikipedia (2005), Diagnosis of exclusion, [cited 2005 Nov], Available from: http://en.wikipedia.org/wiki/Diagnosis_of_exclusion
World Health Organization (2003a), China updates data on cases and deaths, March 26. [cited 2005 Nov 10] Available from: http://www.who.int/csr/don/2003_03_26/en/index.html.
World Health Organization (2003b), Case definitions for surveillance of Severe Acute Respiratory Syndrome, May 1. [cited 2005 Nov] Available from: http://www.who.int/csr/sars/casedefinition/en
World Health Organization (2004a), Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003, April 21. [cited 2005 Nov], Available from: http://www.who.int/csr/sars/country/table2004_04_21/en/index.html.
World Health Organization (2004b), Investigation into China's recent SARS outbreak yields important lessons for global public health, July 2. [cited 2005 Nov], Available from: http://www.wpro.who.int/sars/docs/update/update_07022004.asp.
World Health Organization (2004c), WHO warns governments to prepare for SARS and bird flu, September 14. [cited 2005 Nov], Available from: http://www.wpro.who.int/sars/docs/update/update_09142004.asp.
World Health Organization (2005), Cumulative number of reported probable cases of Severe Acute Respiratory Syndrome, [cited 2005 Nov], Available from: http://www.who.int/csr/sars/country/en/index.html.
World Health Organization (2006a), *World Health Statistics 2006*.
World Health Organization (2006b), *The World Health Report 2006: Working Together for Health*.
Yellis: Curriculum, Evaluation and Management Centre (n.d.), Understanding residuals. [cited 2005 November]. Available from: http://www.yellisproject.org/Handbook/residuals.asp.
United Nations Development Programme (2005), annual Human Development Report, http://hdr.undp.org/reports/global/2005/pdf/HDR05_complete.pdf
University of Aberdeen (1998), Normality of residuals. Department of Physics. [cited 2005 Nov]. Available from:http://www.abdn.ac.uk/physics/physcomp/minitab/tsld018.htm .
Zhang, Zhibin (2007), *The Outbreak Pattern of SARS in China as Revealed by a Mathematical Model,* Ecological Modelling, Vol. 204, 420-426.