

文章编号:1005-3085(2003)07-0029-06

# SARS 传播的数学原理及预测与控制

邹宇庭, 郑晓练, 缪旭晖

指导教师: 谭 忠

(厦门大学, 福建厦门 361005)

**编者按:**本文建立了 SARS 传播的具有负反馈的差分方程模型。用两个参数分别刻画疾病传播能力和对疾病采取有效控制措施的能力,并用标准差趋于稳定来判定数据拟合的合理性,是本文的突出优点。文章将这一方法用于北京,广州,山西和香港,均获得较好效果。

**摘 要:**众所周知,SARS 对中国社会带来了重大的影响。我们以北京地区 4 月到 6 月有关 SARS 的数据为参考资料,就病毒的实际传播特征引入了电子线路中的负反馈的概念,建立了 SARS 传播的负反馈系统,并在分析该系统参数实际意义的情况下,建立时间序列的模型。该模型将传染率定义为时间的函数,以拟合数据和实际数据之间的总残差最小为目标,利用 matlab 中的 fminsearch 函数模拟得到最优的模型参数。该模型可以较好的预测 SARS 的发展趋势,且可以就此趋势提出如何控制 SARS 传播的措施。继而,本文通过模拟出在不同日期提前或滞后 5 天实施隔离政策所引起 SARS 发展趋势变化的曲线,分析了卫生部门实施隔离政策的日期对 SARS 发展趋势的影响。

在 SARS 对经济影响的这个问题上,本文适当选取医疗业具有代表性的 17 支股票,构造了医疗板块指数,以此测度医疗业的经济表现。在传统的 CAPM 模型中,我们引入了虚拟变量,利用 OLS 技术进行估计分析,检验出 SARS 这一事件对医药业的经济影响是正影响。该影响反映在医疗版指数的日收益上,但这个影响是由 SARS 引起的,会随着 SARS 的结束而结束。

**关键词:** SARS; 负反馈系统; 时间序列模型; 资本资产定价模型

**分类号:** AMS(2000) 62M10

**中图分类号:** O212.3

**文献标识码:** A

## 1 数学模型的分析与建立

### 1.1 假设与符号说明

1) 统计数据是可靠的; 2) 病人处于潜伏期时不传染他人; 3) 采取的所有控制措施对于阻止 SARS 病毒的传播都是有效的。 $I_n$ : 到第  $n$  天为止累计确诊的病人数;  $D_n$ : 到第  $n$  天为止累计的死亡人数;  $S_n$ : 第  $n$  天的疑似病人数;  $C_n$ : 到第  $n$  天为止治愈病人数;  $d$ : 死亡率;  $g$ : 治愈率;  $S_1$ : 新增病人与新增疑似病人的比值;  $S_2$ : 疑似病人转化为正常人的比率;  $K_0$ : 区域内的自反馈参量;  $F_n$ : 反馈变量;  $K_f$ : 反馈系数。

### 1.2 现在我们分析问题并建立相应的数学模型:

社会的反应往往是一个渐变的过程,会随疫情的变化而变化,是一个负反馈过程,比如,

当疫情严重时人们会自觉地减少与他人接触,相反的当疫情不显著时人们会放松警惕而增加与他人接触的机会。为此我们建立如下负反馈系统(图1)。将实际情况与控制力度的关系通过反馈系数实现自动调节,这样就能及时采取控制措施,使得病情传播速度迅速减小,以求最后达到消灭病情的目的。

在以上系统中:

①  $A$  是开环增益(Open Loop Gain),表示基本放大器的放大倍率,即为输出信号与输入信号的比值。在研究 SARS 传播的模型中,我们将此定义为 SARS 病毒在未受控制情况下的自然传播速度,即病毒的基本传播率。②  $K_f$  是反馈网络中的反馈系数,即反馈信号与输出信号的比值。在该模型中我们将它定义为各种反应因素的总和,通过负反馈可以减小病毒传播速度。③ 输入信号  $X_i$  定义为当前情况下 SARS 的传播情况。④ 输出信号  $X_0$  定义为  $t$  时间后社会上 SARS 的传播情况。

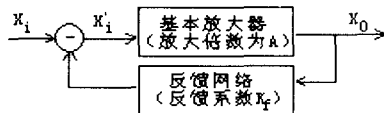


图1 系统示意

◆通过以上系统,我们得到参量  $A, K_f$  与输入信号( $X_i, X_i'$ ),输出信号( $X_0$ )以及反馈信号  $X_f$  之间的关系:

$$X_0 = A \cdot X_i' \quad (1), \quad X_f = K_f \cdot X_0 \quad (2), \quad X_i = X_i' + X_f \quad (3)$$

定义闭环增益(Closed Loop Gain)  $A_f$  为病毒的实际传播率,定义反馈深度  $F$ ,表示对调节元件灵敏度的折中,并将式(1)(2)(3)代入则

$$A_f = \frac{X_0}{X_i} = \frac{A}{1 + \frac{X_f}{X_i}}, \quad F = 1 + \frac{X_f}{X_i} = 1 + \frac{X_f}{X_0} \cdot \frac{X_0}{X_i} = 1 + K_f \cdot A, \quad A_f = \frac{A}{F}$$

由上式分析得到:

①当反馈系数  $K_f < 0$  时,系统是负反馈的,这正是我们需要的系统。这样社会上病毒传播速度即系统中的输出量变化时,系统参数自我调节为:  $X_i \uparrow \rightarrow X_i' \uparrow \rightarrow X_0 \uparrow \rightarrow K_f \uparrow \rightarrow A_f \downarrow \rightarrow X_i' \downarrow$

②当反馈系数  $K_f > 0$  时,系统是正反馈的,由于控制措施的目的是阻止病毒传播,因此  $K_f$  不可能是正值,除非处理 SARS 的方法是错误的。对于这个系统来说,可以不考虑这种情况。

◆基于对以上系统的分析,我们考虑建立现实的 SARS 传播模型:

①因为要控制的实际量为 SARS 确诊病人数,因此  $X_i$  就为  $I_n, X_0$  为  $I_{n+1}, I_n$  表示到第  $n$  天为止累计确诊的病人数。

②量化时间,以天为单位。

③选定某一区域为研究区域,在模型中建立各反馈参量:

a)设该区域内的自反馈参量为  $K_0$ ,表示该地区在未采取控制措施时 SARS 的传播能力。

b)设该地区反馈量  $F_n$  的变化率为  $f_k$ ,即每增加一个病人引起反馈量  $F_n$  的变化量。 $f_k$  表示该地区的病情控制情况。

由此,建立以下时间序列模型:

$$F_n = K_0 + f_k \cdot (I_n + S_n) \quad (1)$$

$$I_{n+1} = I_n + F_n \cdot I_n - C_n - (D_n - D_{n-1}) \quad (2)$$

$$D_{n+1} = D_n + d(I_n - D_n - C_n) \quad (3)$$

$$S_{n+1} = S_n + (I_{n+1} - I_n) \cdot S_1 - S_n \cdot S_2 \quad (4)$$

$$C_{n+1} = C_n + g \cdot (I_n - I_{n-1}) \quad (5)$$

其中式(1)表示反馈量是自反馈参量和人为控制后的反馈参量的和;式(2)表示确诊病人总数的变化情况;式(3)表示死亡人数的变化情况;式(4)表示疑似病例数量的变化情况;式(5)表示治愈病人数的变化情况。如果病情得到控制,既病人不再增加,那么反馈量应该是0,易得  $f_k$  必须为负数。所以

$$\lim_{x \rightarrow \infty} F_n \rightarrow 0 \quad \lim_{x \rightarrow \infty} I_{n+1} \rightarrow I_n \quad \lim_{x \rightarrow \infty} D_{n+1} \rightarrow D_n \quad \lim_{x \rightarrow \infty} S_{n+1} \rightarrow 0$$

### 1.3 问题的求解

设实际数据为  $I_{n0}$ , 拟合数据为  $I_n$ , 则我们确定参数的目标是使总残差最小, 即:  $\min E = \sum_{n=0}^n (I_n - I_{n0})^2$ 。我们用 matlab 中的 fminsearch 函数来求解, 得到总残差最小时的参数  $k_0, f_k$  并由模型中式(1)、(2)迭代求出  $I_n$  发展趋势的曲线。取用实际数据量不同, 各参数以及最小总残差便不同。模拟时, 只要达到一定的数据量就可以很好地拟合出曲线, 我们以标准差来判断数据量是否已经足够。数据量从 10 天变化到 64 天, 标准差变化的曲线如图 2 所示: 将标准差接近平稳状态的数据量取值区间抽出:

数据量(天)	23	24	25	26	27	28	29	30	31
标准差	14.047	13.894	13.833	13.801	13.78	13.762	13.782	13.82	13.816

可见, 数据量取到 25 天以上后标准差已经稳定。因此, 可以认为拟合北京市 SARS 发展趋势线只要取其从 4 月 1 日开始的 25 天数据即可。因篇幅有限删除具体计算程序。取前 25 天数据拟合出的各发展曲线如图 3 所示

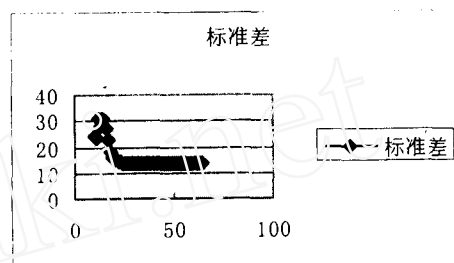


图2 标准差曲线

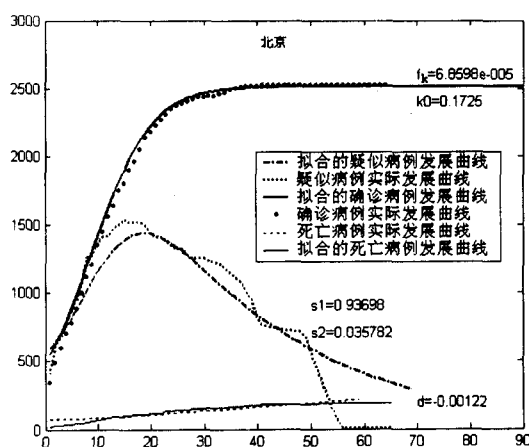


图3 发展曲线

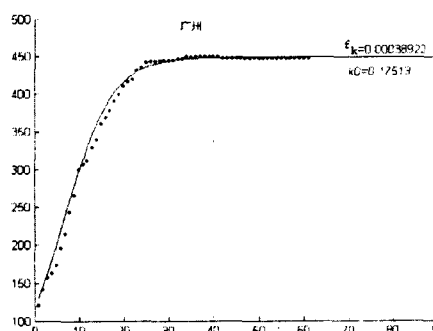


图4 广州

该曲线图不仅描绘出 25 天的发展情况, 而且通过预测描绘出 25 天以后的发展情况。对比 25 天后预测数据和实际数据: ①对于累计确诊病人, 它们之间的标准差稳定在 13.

586附近,因此两者是十分吻合的。②对于疑似病人数,由于疑似病例的判断受主观因素的影响,拟合情况不如其他两条曲线。③对于死亡病人数,吻合程度依然比较高,误差产生的原因可能是个人的免疫能力不同这个原因无法在定义的反馈系数中体现出来。由此可以看出以上的预测十分成功。

◆为验证该方法的有效性,我们分别计算了广州,山西和香港的累积确诊病例发展趋势,得到的预测数据和实际数据吻合程度很高,可见该方法是十分有效的。具体见图4-6:

#### 1.4 对卫生部门措施的评论

##### A)对发展趋势曲线的影响:

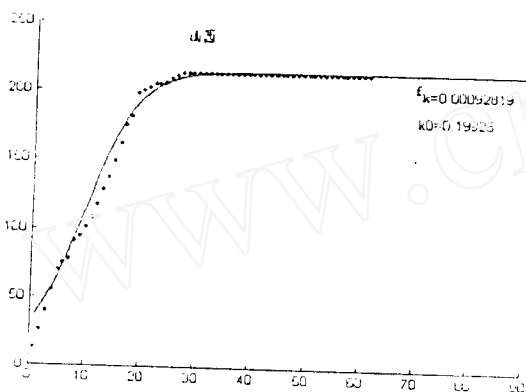


图5 山西

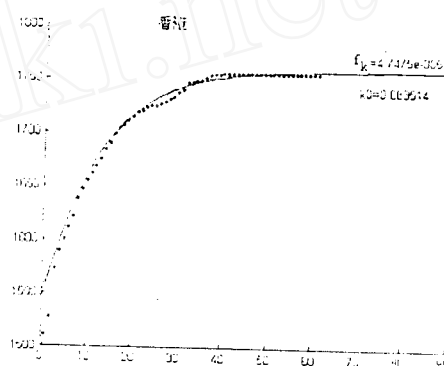


图6 香港

图7体现了从不同时间点开始提前或滞后5天实施隔离政策,得到的总患病人数随日期的发展变化,如图中取了①、②、③、④四个时间点。原始数据反映的总患病人数的曲线为(-),若提早5天采取隔离政策,则对应四个时间点的总患病人数的曲线分别为(二)、(四)、(六)、(八),若推迟5天,则对应四个时间点的总患病人数的曲线分别为(三)、(五)、(七)、(九)。从图中可以很明显的看到从任一给定时刻起提前5天采取政策比正常情况下可减少一定的患病者,而滞后5天则会增加患病者,这一时刻取得越早,就可使越多的人免于得病,因而政策的效用就越大。

##### B)对总确诊人数的影响:(图8中的每一对点对应统一的横坐标)

影响值	第7天	第8天	第9天	第10天	第11天
滞后5天	1.899	1.837	1.767	1.692	1.613
提前5天	0.62	0.653	0.689	0.727	0.766
影响值	第12天	第13天	第14天	第15天	第16天
滞后5天	1.532	1.452	1.376	1.306	1.245
提前5天	0.803	0.839	0.87	0.898	0.921
影响值	第17天	第18天	第19天	第20天	第21天
滞后5天	1.192	1.149	1.114	1.086	1.065
提前5天	0.939	0.954	0.966	0.974	0.981
影响值	第22天	第23天	第24天	第25天	第26天
滞后5天	1.048	1.036	1.026	1.019	1.014
提前5天	0.986	0.99	0.993	0.995	0.996
影响值	第27天	第28天	第29天	第30天	第31天
滞后5天	1.011	1.008	1.006	1.004	1.003
提前5天	0.997	0.998	0.999	0.999	0.999

上表为从第5个时间点即第5天开始考虑提前和滞后5天采取隔离措施带来的总确诊人数的影响,影响值以影响后的总确诊人数与无影响下的总确诊人数的比值表示。由上表

和图 7-8 可见,越早实施隔离政策,最后总确诊病例数就越少,而且新增病人就越快接近 0,这意味着 SARS 病情越早得以控制。

## 2 SARS 对经济影响的数学模型分析与建立

SARS 对国民经济的各方面均带来冲击,并给国民经济带来重大损失,但同时 SARS 也给某些行业,如医疗业带来正面的影响。下面构建带虚拟变量的 CAPM 模型实证检验

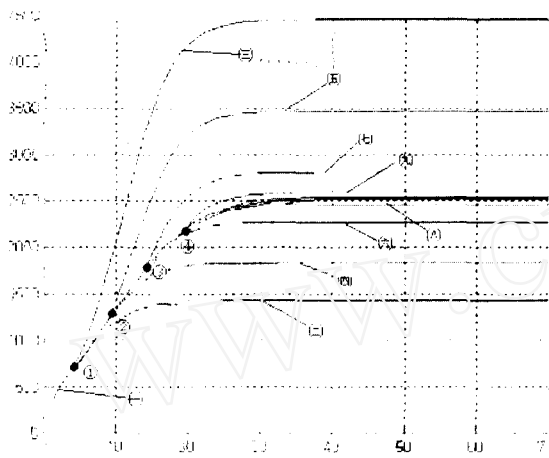


图 7 对发展趋势曲线的影响

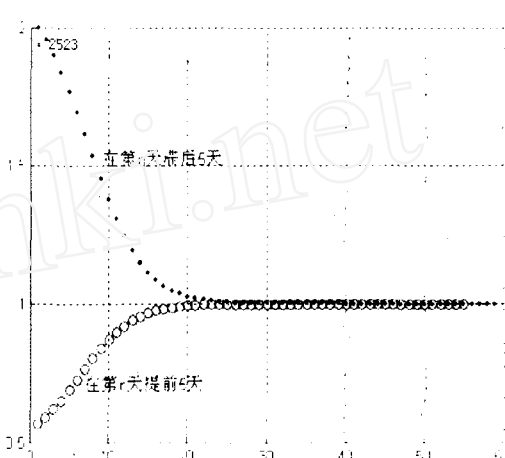


图 8 提前或滞后 5 天的总确诊病人数

SARS 是否给医疗业带来影响和多大程度的影响。行业的股票指数在一定程度上反映了整个行业在大经济环境下的业绩表现。模型的数据的获取从 <http://www.moneywise.com.cn/downspj.htm> 取得,选取的时间段从 2002 年 9 月 2 号到 2003 年 6 月 30 号。选取的股票为:万东医疗,浙江医药,恒瑞医药,南京医药,鲁抗医药,上海医药,华东医药,三九医药,广济药业,金陵药业,新华制药,百科药业,四环药业,东北药,ST 海药,恒和制药,云南白药。定义:  $P_{it}$  为第  $i$  只股票第  $t$  日的收盘价格;  $P_{Mt}$  为深圳成指的第  $t$  日的收盘价;  $r_{Et}$  为医疗股指的日收益;  $r_{Mt}$  为市场指数的日收益率。构建医疗业板块的股票指数:以等权重法计算医疗业股指,计算过程中考虑到了新增股和个股退市的权重调整,具体表达式为

$$P_{Et} = \frac{\sum_i P_{it}}{\sum_i 1} \quad r_{Et} = \ln \frac{P_{Et}}{P_{Et-1}} \quad r_{Mt} = \ln \frac{P_{Mt}}{P_{Mt-1}}$$

建立模型如下

$$(r_E - r_f)_t = \alpha + \beta(r_M - r_f)_t + \gamma D_t + \epsilon_t$$

$$D_t = \begin{cases} 0 & t \text{ 处于 } 2002 \text{ 年 } 9 \text{ 月 } 2 \text{ 号到 } 2003 \text{ 年 } 2 \text{ 月 } 11 \text{ 号} \\ 1 & t \text{ 处于 } 2003 \text{ 年 } 2 \text{ 月 } 11 \text{ 号到 SARS 结束以后} \end{cases}$$

其中  $D_t$  为虚拟变量,当有 SARS 影响时  $D_t$  为 1,无影响则为 0。由于 SARS 开始的日期为 2003 年 2 月 11 日,因此采集的数据时往前至 2002 年 9 月 2 号往后到 2003 年 6 月 30 号。 $\alpha, \beta, \gamma$  为待估计系数,其中:  $\alpha$  为截距项;  $\beta$  即为 CAPM 中的“ $\beta$ ”系数,用以衡量个股(板块)的系统风险;  $\gamma$  为分析的重点,如果  $\gamma$  统计上显著异于零,则说明 SARS 对医疗业发生了显著影响,并且影响程度为  $\gamma$ ;  $\epsilon_t$  为扰动项;  $r_f$  为无风险利率,在本模型中用银行 3 月期存款利率测度;  $r_M$  为市场指数收益,本模型中用深圳成份指数测度。

## 2.2.3 模型求解

Variable	Coefficient	Std. Error	t - Statistic	Prob.
C	-0.001428	0.001497	-0.954424	0.3413
RM - RF	0.292576	0.088641	3.300665	0.0012
D	0.000213	0.002555	3.860554	0.0126
R - squared	0.766419	Mean dependent var		-0.00137
Adjusted R - squared	0.7763	S.D. dependent var		0.01566
S.E. of regression	0.015227	Akaike info criterion		-5.51277
Sum squared resid	0.036172	Schwarz criterion		-5.45487
Log likelihood	441.2653	F - statistic		5.549341
Durbin - Watson stat	1.991851	Prob(F - statistic)		0.004697

从表中可以看出:SARS的爆发对医疗设备,医疗药物的需求带来了冲击性的影响,反映在医疗版指数的日收益上,带来了0.0213%的额外日收益。当然,这些收益是由于SARS带来的,随着SARS的平息,这个正的效益也会随之平息。这便是SARS对医疗行业的经济影响。

## 参考文献

- [1] 姜启源. 数学模型. 北京:高等教育出版社,2002
- [2] 邓聚龙,郭洪. 灰色理论. 北京:全华出版,2002
- [3] 谢嘉奎. 电子线路—线形部分. 北京:高等教育出版社,1999
- [4] James D,Hamilton. 时间序列分析. 北京:中国社会科学出版社,1999

## The Mathematical Principle Of The Spread Of SARS and Its Application On Forecasting and Controlling SARS Epidemic

ZOU Yu-ting, ZHENG Xiao-lian, MIAO Xu-hui

Advisor: TAN Zhong

(Xiamen University, Xiamen, Fujian 361005)

**Abstract:** It is well known that SARS has a tremendous effect on Chinese society. Based on the statistics related to SARS in Beijing from April to June in 2003, we modeled the Negative Feedback system of the spreading of SARS by introducing the concept of Negative Feedback System in electro - circuit and the nature of the SARS virus. Moreover, by analyzing the practical meanings of the parameters in this system, a complete Time Sequences derived model is presented. It defines the spreading rate as a function of time, and adopts the fminsearch function in matlab to give the best parameters, with the object of minimum total difference between the actual statistics and computed results. This model well forecasts the spreading trend of SARS, according to which proper advices can be put forward on controlling policies. Further, curves are drawn that simulate the different development trends as the result of the quarantine carried out five days ahead of or five days later than a series of different dates. Comparison between these curves makes it possible to analyze the impact exerted by the date when the sanitary department carried out a quarantine.

Concerning the influence of SARS on economy, 17 representative stocks of medicine industry are chosen in order to measure specifically and indirectly the overall trend of this industry in spreading of SARS. Dummy variations are introduced to the traditional CAMP model in order to verify the positive effect imposed on medicine industry by SARS epidemic, employing OLS technique. This positive effect was reflected by the daily interest calculated from these stocks. Yet, this impact would be gone when its cause, the SARS epidemic, comes to an end.

**Keywords:** SARS epidemic; negative feedback system; time sequences; CAMP model