

基于时间维度的影响力量化模型

摘 要

本文针对上海世博会影响力评价问题,在多方面收集数据、初步分析和处理数据的基础上,给出了量化评估方案,提出了基于时间维度的量化模型,分三个阶段讨论影响力在每个阶段的不同表现形式。

首先,考虑到上海世博会规模大,产生影响的时间长,且在不同时期影响力侧重点不同的特点,我们分析了世博会从筹备期至正式会期的各类相关数据,发现世博会的影响力在展会的不同阶段具有不同的表现形式。在筹备阶段,世博会的影响力主要体现在对承办城市经济发展的拉动作用;在正式会期,世博会的影响力作用综合体现在多个方面;在会议结束后,世博会的影响力主要通过各种媒体舆论来体现。

针对筹备阶段,我们基于对1978年-2009年上海市经济状况的数据分析,建立了上海世博会在筹备阶段对上海市经济产生影响的量化模型。模型应用时间序列分析理论,利用世博会申办成功前的经济指标数据预测了申办成功后上海市去除世博因素的经济指标,并将预测值与受世博因素影响的实际值进行差值分析,实现了世博会筹备期间影响力的量化。本模型还运用柯布-道格拉斯生产函数,评估了筹备世博会对生产总值、从业人员和固定资产投资三项指标的不同影响,并探讨了各影响因素间的内在关系。

针对世博会正式会期,鉴于上述量化模型无法评价世博会会展期间产生的影响力,我们建立了基于主成分分析法(PCA)的综合影响力模型。本模型通过分析8届世博会的会议指标样本,采用相关系数矩阵的主成分分析得出样本中世博会影响力的综合评价,并将此8届世博会的综合评价进行横向量化比较。

针对世博会结束之后,在上述模型分析的基础上,我们有必要对世博会后续影响力的变化趋势进行量化分析。考虑到世博会一类大型活动结束后,我们以Google-Timeline网络关键词搜索量为衡量标准,基于Logistic微分方程模型建立了后续影响力的发展趋势模型,对世博会后影响力的强度和持续时间进行定量预测。

最后,我们对研究结果进行了总结,根据上海世博会前期、即时、后续三个阶段的影响力,给出了影响力的综合评价。

关键词: 时间维度 时间序列分析 柯布-道格拉斯生产函数 主成分分析 影响力指数 Logistic模型

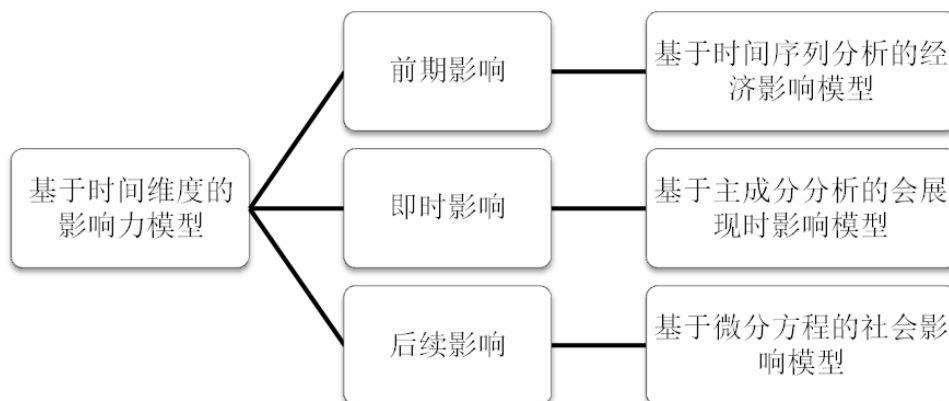
1 问题重述

世博会属于“特大事件”范畴，即规模庞大以致于影响整个经济，并在全球媒体中引起反响的活动^[1]。它包括奥林匹克运动会和世界博览会。此类特大事件由于其规模巨大，影响范围广，因而对其影响力的量化评估较为困难。现有的此类事件影响力评估基本为定性的理论分析，定量分析多基于依赖于主观因素的问卷调查结果。综合以上背景，本文要解决以下几个问题：

1. 上海世博会的影响有哪些主要表现形式？
2. 这些影响的程度如何，怎样测度？
3. 与类似的重大事件相比，上海世博会的特点是什么？
4. 能否得出类似重大事件影响力的理论规律？

2 问题分析

上海世博会规模庞大，影响因素众多。如何在诸多因素中剥离并测度事件的真实影响是评估的难点。另外，上海世博会对该地区造成影响的时间和空间广度较大，数据分散，如何获得和整合数据也是本研究的难点。考虑到上海世博会的持续时间长，影响力评价机制复杂，对上海世博会影响力的分析需在时间维度上分为三个子模型，采用定性和定量相结合的方法，提高研究的有效度。模型的有效性需要用实际现象来验证。



3 模型一 基于时间序列分析的经济影响模型

本模型通过定量分析上海世博会筹备期对上海经济的影响来研究上海世博会筹备期的影响力。研究时间跨度为1978-2009年，研究指标为上海地区生产总值，从业人员和固定资产投资，模型分别对这3个因素建立了时间序列模型。根据上海世博会在2003年申办成功这一事件，将数据分为申办前（1978-2003）和申办后（2004-2009）两个时间段，前者用于预测在没有世博会影响下的各指标值，而后者用来和前者比对，从而计算得到上海世博会对上海经济的影响力指数。通过比较各指标在世博会筹备期的总影响力指数，可知世博会的申办成功对三个经济指标都产生了正面影响，其中对固定资产投资的影响最大，对生产总值的影响相对较小。本模型最终由生长函数统一了三大经济指标的影响力指数，并反映了各影响力指数的内在联系。虽然本模型探讨的是上海世博会在筹备期对上海经济的影响，该影响力指数也可以反映上海世博会在筹备期的总影响，因为世博会在筹备期的影响主要体现在经济方面，尤以在建筑业，制造业等产业的影响为主。

3.1 模型假设

- 1. 上海世博会在申办成功前的影响可忽略不计。
- 2. 2004-2009年间未发生大事件，即2004-2009年间各种小事件对经济指数的影响属于随机误差，不会对时间序列预测造成本质影响。
- 3. 所有数据（《上海统计年鉴》和上海统计网）真实地反映了实际情况，没有错误。

3.2 样本选取

本模型选用的样本区间为1978-2009年，共32个样本，所有原始数据均来自《上海统计年鉴》和上海统计网^[1]。样本选取的参数有各年上海市生产总值(GDP)，固定资产投资，从业人员，商品零售价格总价格指数以及固定资产投资价格指数（见表2）。数据分析处理中所采用的软件有Matlab，SAS和Excel。考虑到各年物价指数的差异，特选用商品零售价格总价格指数和固定资产投资价格指数分别对生产总值和固定资产投资数据进行处理，获得上海市实际生产总值和实际固定资产投资。实际值的计算公式为实际值=原值/指数*100。然而，由于早期的数据资料有限，固定资产投资价格指数存在数据缺失（缺1978-1994年）。根据经济学基本原理可知，固定资产投资价格指数和商品零售价格总价格指数高度相关，因此通过研究1995-2009年上海固定资产投资价格指数和商品零售价格指数的关系，建立回归模型以补全1978-1994年的缺失数据。

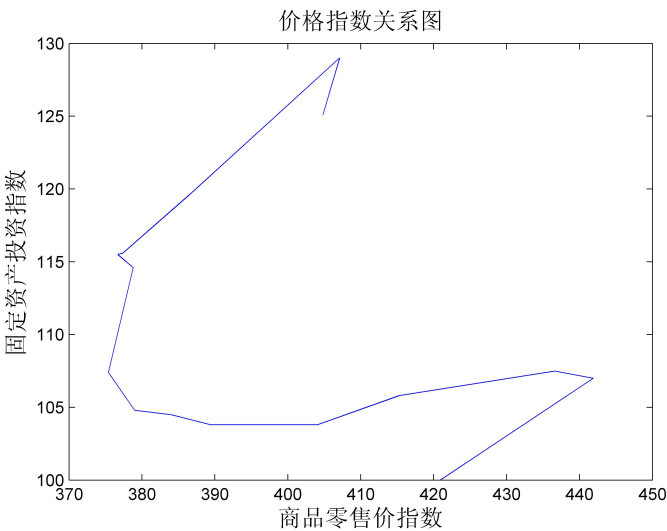


图 1 价格指数关系

由两指数的关系图可看到两指数间并不存在简单线性关系，因此选用二次多项式对数据进行回归分析。设固定资产投资价格指数为y，商品零售价格总价格指数为x，回归方程为

$$y = -0.0043x^2 - 3.4524x - 574.8439 \quad (0)$$

此回归模型较准确地反映了两指数间的关系且误差较小（见附件）。因此用该模型的预测值对1978-1994年固定资产投资指数数据进行了补全（表2）。

表 1 变量说明

x_1	年份	x_4	上海从业人员（万人）	x_7	固定资产投资指数（拟合）
x_2	上海市生产总值（亿元）	x_5	商品零售价格总价格指数	x_8	上海市生产实际总值（亿元）
x_3	上海固定资产投资（亿元）	x_6	固定资产投资价格指数	x_9	固定资产投资实际值

表 2 原始数据和处理

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1978	272.81	27.91	698.32	100.0		100.00	272.8	27.91
1979	286.43	35.58	712.59	101.0		102.58	283.6	34.68
1980	311.89	45.43	730.77	107.6		119.40	289.9	38.05
1981	324.76	54.60	750.22	109.2		123.42	297.4	44.24
1982	337.07	71.34	764.03	109.5		124.17	307.8	57.45
1983	351.81	75.94	768.90	109.6		124.42	321.0	61.03
1984	390.85	92.30	769.79	112.0		130.41	349.0	70.78
1985	466.75	118.56	775.53	130.4		174.61	357.9	67.90
1986	490.83	146.93	782.99	139.1		194.48	352.9	75.55
1987	545.46	186.30	788.12	151.4		221.47	360.3	84.12
1988	648.3	245.27	792.13	183.6		285.90	353.1	85.79
1989	696.54	214.76	784.96	214.3		338.97	325.0	63.36
1990	781.66	227.08	787.72	224.6		354.94	348.0	63.98
1991	893.77	258.30	798.13	245.9		385.06	363.5	67.08
1992	1 114.32	357.38	806.91	269.8		414.17	413.0	86.29
1993	1 519.23	653.91	787.25	317.0		457.13	479.3	143.05
1994	1 990.86	1 123.29	786.04	372.4		482.92	534.6	232.60
1995	2 499.43	1 601.79	794.19	420.9	100.0	472.93	593.8	338.69
1996	2 957.55	1 952.05	851.21	441.9	107.0	479.93	669.3	406.74
1997	3 438.79	1 977.59	847.25	436.6	107.5	480.47	787.6	411.60
1998	3 801.09	1 964.83	836.21	415.2	105.8	478.74	915.5	410.41
1999	4 188.73	1 856.72	812.09	404.0	103.8	476.73	1036.8	389.47
2000	4 771.17	1 869.67	828.35	389.5	103.8	476.73	1225.0	392.18
2001	5 210.12	1994.73	745.24	384.0	104.5	477.46	1356.8	417.78
2002	5 741.03	2 187.06	752.26	379.0	104.8	477.77	1514.8	457.76
2003	6 694.23	2 452.11	792.04	375.4	107.4	480.29	1783.2	510.55
2004	8 072.83	3 084.66	813.05	378.8	114.6	487.48	2131.2	632.77
2005	9 247.66	3 542.55	863.32	376.7	115.5	488.40	2455.1	725.34
2006	10 572.24	3 925.09	885.51	377.4	115.6	488.52	2801.7	803.47
2007	12 494.01	4 458.61	909.08	386.5	119.6	492.53	3232.4	905.25
2008	14 069.87	4 829.45	1 053.24	407.1	129.0	501.98	3456.4	962.08
2009	15 046.45	5 273.33	1 064.42	404.8	125.1	498.03	3716.8	1058.84

3.3 模型建立

根据2003年上海世博会申办成功这一事件，将数据分为1978-2003和2004-2009这两个阶段。其中，1978-2003年的数据用于建立各经济指标的时间序列模型以获得2004-2009年的预测值，即模拟若不举办上海世博会这些指标的理论数值。2004-2009年的数据用于和预测值进行比对，以研究上海世博会举办所造成的影响。为了便于生长函数模型的计算以及指数的标准化问题，参考前人处理方法^[3]，取实际值/预测值*100的对数为影响力指数。当实际值=预测值，即没有影响时，ln(实际值/预测值*100)=4.60517，此值作为均衡影响力指数。此外，每个指标的预测值和实际值的趋势图直观地体现了世博会申办成功后的筹备期对上海经济的影响，两趋势线所围区域为上海世博会对该经济指标产生的总影响。

3.3.1 上海市生产总值时序模型

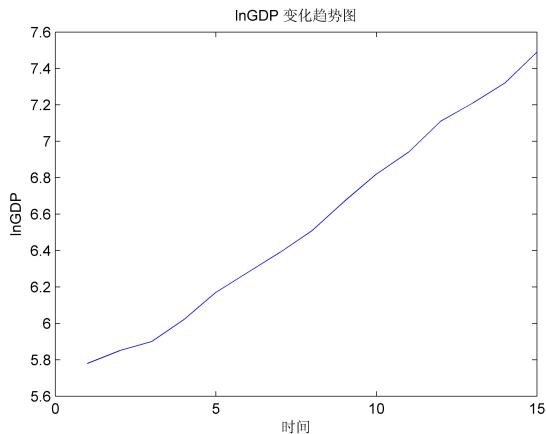


图 2 lnGDP线性趋势

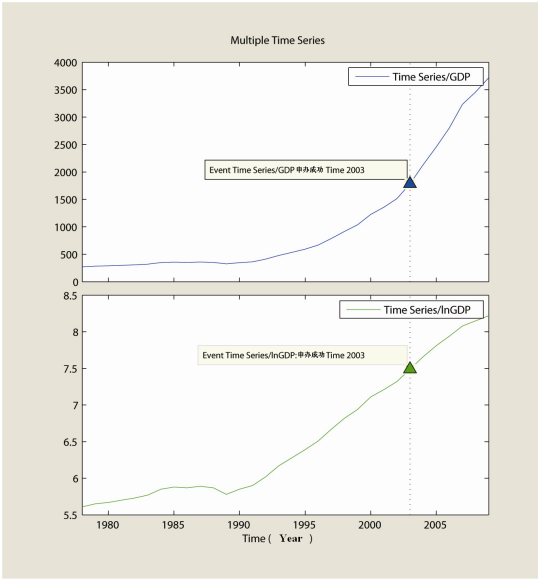


图 3 GDP&lnGDP趋势

由图3可看到上海市生产总值呈逐年上升趋势。对数据取对数后，上升趋势呈明显线性，并消除了数据可能存在的异方差。其中上升趋势明显分为两大线性阶段，且数值与时间相关。为更准确预测2004-2009年数据并排除历史原因对数据的干扰，选用1990-2003数据进行时间的一元线性回归，得到回归方程如下

$$y = 5.5534 + 0.1263x \tag{1}$$

$$R^2 = 1.00 \quad p : 0.0000$$

其中 $R^2 = 1.00$ 说明回归方程对样本的拟合优度非常高，置信度非常高，可用来预测上海2004-2009年的生产总值（表3）。

将GDP各年实际值和预测值加和并处理得到上海世博会筹备期对上海地区生产总值的总影响，其影响力指数为 $4.686892 > 4.60517$ （均衡影响力指数）。通过比较上海生产总值的预测值和实际值可发现实际值各年均大于预测值，由于预测值是排除举办世博会影响得到的，因此两者之差可大致归结于上海世博会的申办成功以及举办准备对全市生产总值的推动作用。这种影响可以更直观地在图中体现，图中两条趋势线所夹区域即为上海世博会的影响。

此外，由生产总值的预测值和实际值还可得出上海世博会在筹备期间对生产总值的影响占生产总值的比例为7.8%，即上海世博会的筹备工作对上海生产总值的贡献约7.8%，这说明了世博会对上海经济的巨大影响力。

表 3 生产总值影响

年份	lnGDP预测值	GDP预测值	GDP实际值	实际值/预测值*100	影响力指数
2004	7.5742	1947.302	2131.159	109.44163	4.695391
2005	7.7005	2209.452	2455.055	111.11601	4.710575
2006	7.8268	2506.894	2801.654	111.75798	4.716336
2007	7.9531	2844.379	3232.350	113.63992	4.733035
2008	8.0794	3227.296	3456.385	107.09848	4.673749
2009	8.2057	3661.763	3716.751	101.50168	4.620075
前期总和		16397.086	17793.354	108.51534	4.686892

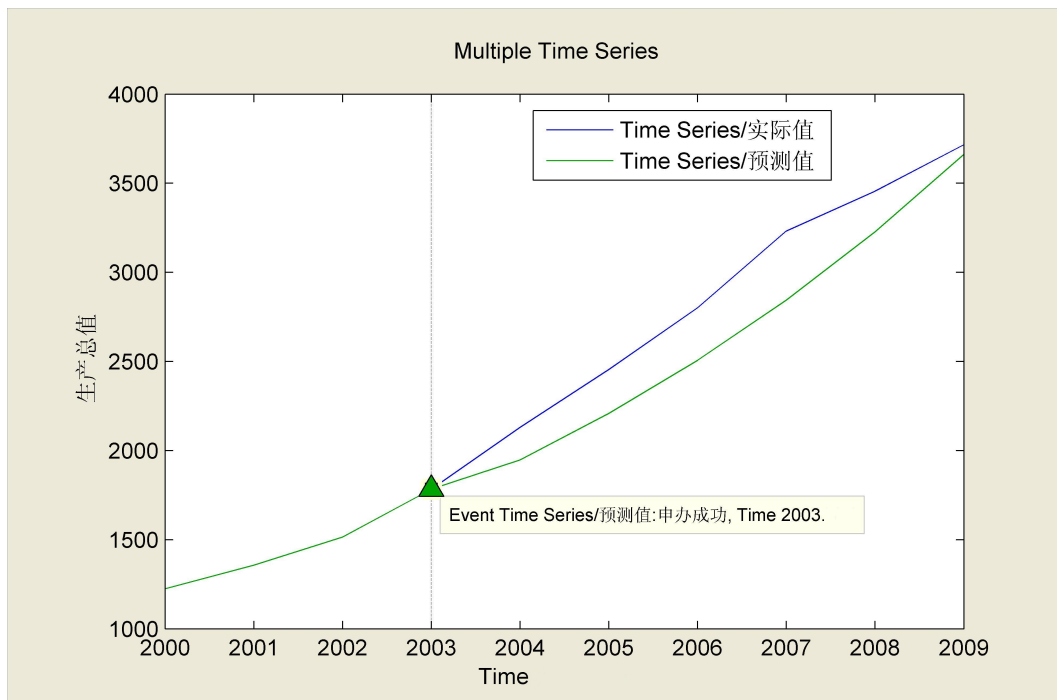


图 4 GDP比较图

3.3.2 上海市从业人员时序模型

根据上海统计年鉴查得1978-2003年间上海市从业人员数据，其变化趋势见图5。从业散点。根据该数据计算得到1978-2003年上海市从业人员增长率数据，其趋势图较图5平稳，为时间序列分析作铺垫。

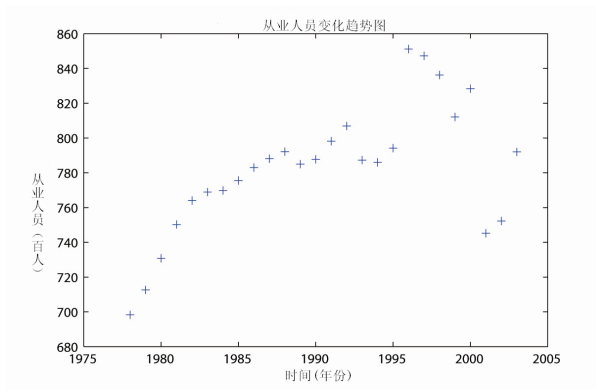


图 5 从业人员变化趋势

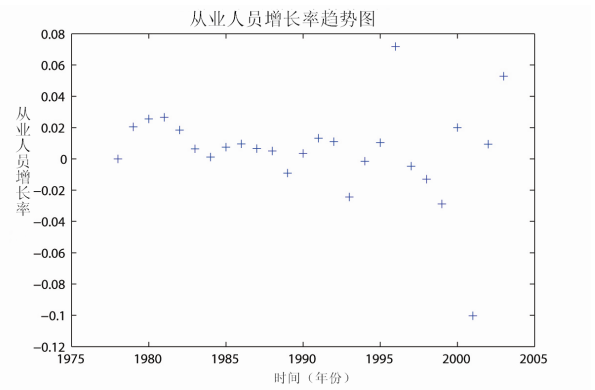


图 6 从业人员增长率趋势

在建立时序模型前，先对该时间序列的平稳性进行检验。自相关图和偏相关图见图7和图8。

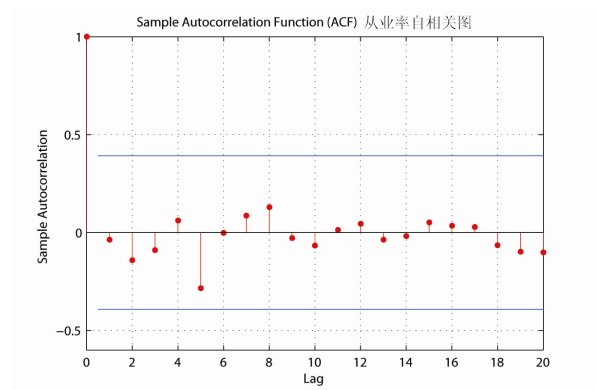


图 7 从业率自相关图

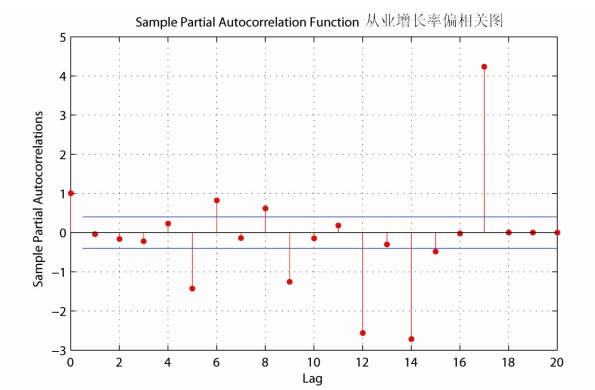


图 8 从业率偏相关图

从业人员的增长率时间序列在自相关图中一阶拖尾，在偏相关图中高阶截尾。单位根检验显著 ($h=1;p=0.001$)，即平稳性较好可以用AR模型。根据经验和相关图中拖尾情况，选用AR(12)模型预测时间序列，结果如下

$$A(q) = 1 + \frac{0.08482}{q^{12}} - \frac{0.4604}{q^{11}} + \frac{0.5539}{q^{10}} - \frac{0.5874}{q^9} - \frac{0.07874}{q^8} + \frac{0.3417}{q^7} - \frac{1.249}{q^6} + \frac{1.416}{q^5} - \frac{0.8482}{q^4} + \frac{0.2536}{q^3} + \frac{0.5843}{q^2} - \frac{0.5872}{q} \tag{2}$$
$$\tag{3}$$

其AR模型的Loss function=0.000485271，FPE=0.000933213，说明该模型可以较准确地拟合样本数据，其预测2004-2009年的数据见表4。

虽然2004年从业人员的预测值大于实际值，但在后面5年间实际值都大大超过了预测值，该期间世博会对上海从业人员的总影响力指数为4.76676 > 4.60517（均衡影响力指数），说明上海世博会自申办成功以来制造了更多的就业机会。在图9中，两条趋势线所包含的面积即为扩大的就业岗位/人员。

表 4 从业人员影响预测

年份	从业增长率预测	从业人员预测值	从业人员实际值	实际值/预测值*100	影响力指数
2004	0.07132	848.53	813.05	95.8188	4.562459
2005	-0.06685	791.81	863.32	109.0317	4.691639
2006	0.02712	813.28	885.51	108.8810	4.690256
2007	-0.03791	782.45	909.08	116.1840	4.755175
2008	0.02317	800.57	1053.24	131.5606	4.879467
2009	-0.10422	717.14	1064.42	148.4255	5.000083
前期和		4753.78	5588.62	117.5616	4.766962

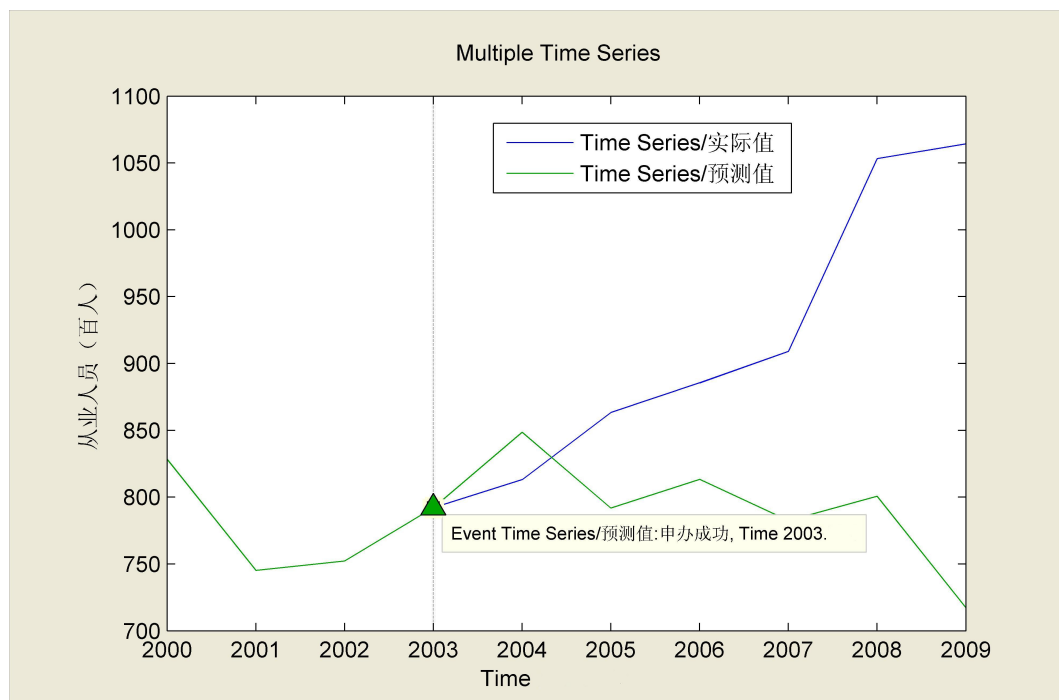


图 9 从业人员比较图

3.3.3 上海市固定资产投资时序模型

为将时间序列转化为易于处理的平稳序列，先对实际固定资产投资数据取对数以消除样本可能存在的异方差，再对数据处理得到较为平稳的各年增长率数据。其单位根检验显著(h=1; pValue=0.035)，自相关和偏相关图见图10和图11。

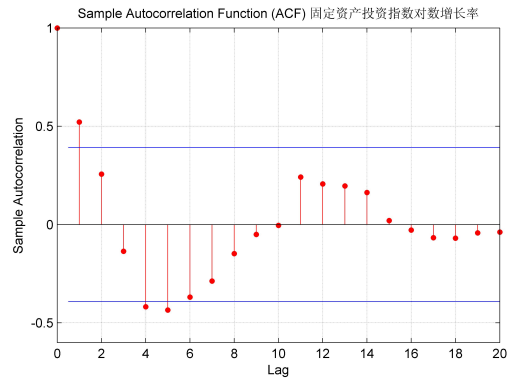


图 10 固定资产投资对数增长率自相关图

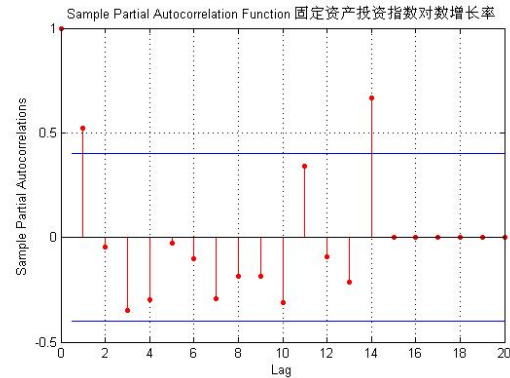


图 11 固定资产投资对数增长率偏相关图

根据自相关图的拖尾性和偏相关图的截尾性用一阶的AR模型来预测04-09上海固定资产投资的短期行为，预测值见表5。AR(1)模型如下

$$A(q)y(t) = e(t) \tag{4}$$

$$A(q) = 1 - 0.6787\frac{1}{q} \tag{5}$$

Loss function 和FPE的数值都较小，可知AR（1）能较好地拟合和预测该时间序列。

表 5 固定资产投资影响

年份	固定资产投资预测值	固定资产投资实际值	实际值/预测值*100	影响力指数
2004	510.7	632.772	123.9029	4.8194982
2005	508.1	725.338	142.7549	4.9611294
2006	508.4	803.473	158.0395	5.0628452
2007	511.6	905.246	176.9442	5.1758343
2008	516.3	962.083	186.3419	5.2275831
2009	521.8	1058.838	202.9203	5.3128132
前期总和	3076.9	5088.000	165.4000	5.1080833

固定资产投资的预测值和实际值形成了巨大的反差。预测值趋势非常平缓，是1978-2003年投资的体现，该预测值可能也受到了历史因素的影响。而实际值在2004-2009期间上升很迅速，幅度也很大，不符合历史数据的发展趋势，这也间接说明了在2003年左右有大事件（上海世博会）发生而对固定资产投资产生了较大影响。此外，这一阶段世博会对上海固定资产投资的总影响力指数为5.10808，大大高于均衡影响力指数，也体现了上海世博会的申办成功对固定资产投资的较大影响。预测值和实际值的趋势对比可见图12。

3.3.4 生产函数模型

由经济学基本原理可知有生产函数：

$$Y = AK^{\alpha}L^{\beta}e^{\gamma X} \tag{6}$$

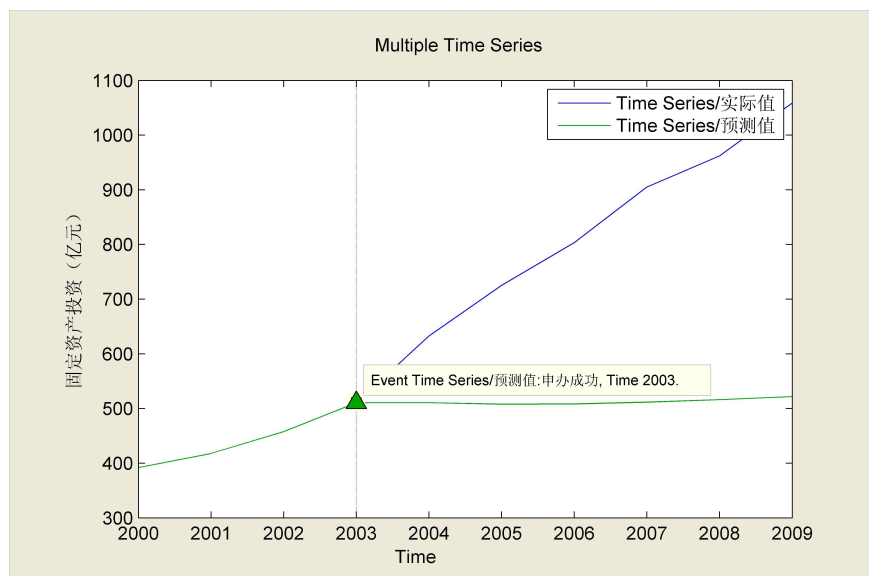


图 12 固定资产投资比较图

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \gamma X \quad (7)$$

其中Y为上海实际生产总值, L为上海从业人员, K为上海固定资产投资, A代表技术进步对生产总值的影响, X表示世博申办成功后的年数, 如X=1表示申办成功后第1年即2004年。其中lnY, lnK和lnL已由前面各时序模型求得, X根据预测年份也可以确定, 具体数据见表6。

表 6 影响力指数汇总

年份	lnY	lnL	lnK	X
2004	4.695	4.5625	4.8195	1
2005	4.711	4.6916	4.9611	2
2006	4.716	4.6903	5.0629	3
2007	4.733	4.7552	5.1758	4
2008	4.674	4.8795	5.2276	5
2009	4.620	5.0001	5.3128	6
前期总和	4.687	4.7669	5.1081	

根据上表数据对生产函数进行回归分析可拟合出函数中其他参数, 获得生产函数如下

$$\ln Y = 1.79585 + 0.91027 \ln K - 0.30927 \ln L - 0.07643 X \quad (8)$$

其中部分重要统计参数如表7和表8所示 (完整的数据请见附件)。

表 7 显著性检验

Source	DF	Sum of Squares	MeanSquare	F Value	Pr > F
Model	3	0.0079	0.00263	22	0.0438
Error	2	0.000239	0.0001197		
Corrected Total	5	0.00814			

由上述统计结果可知模型较显著 ($Pr > F = 0.0438$), 拟合程度也较好 ($R^2 = 0.9706$)。该生产函数模型可用于研究上海世博会通过对固定资产投资, 从业人员以及时间的影响而产生的对生产总值

表 8 回归统计参数

Root MSE	0.01094	R-Square	0.9706
Dependent Mean	4.69153	Adj R-Sq	0.9265
Coeff Var	0.23316		

的影响。比如, $\ln K$ 的系数 0.91927 说明由世博会申办成功而导致的固定资产投资对数每 1% 的增加, 将带动生产总值 0.91927% 的增长。也就是说 $\ln K_{t+1} - \ln K_t = 1$, $\ln Y_{t+1} - \ln Y_t = 0.91927$, 即 $L_{t+1}/L_t = e$, $Y_{t+1}/Y_t = e^{0.91927}$ 。而 X 的系数 -0.07643 表明了时间增长和世博会影响的负相关关系。世博会申办成功后每过 1 年, 世博会对上海经济的影响就会减少 $e^{0.07643}$ 。而这种影响的递减规律也存在于世博会后期 (在模型 3 中将具体讨论)。该生长函数模型的建立不仅统一了三大经济指标的影响力指数并建立了三大经济指标影响的内在联系, 还可以用于今后的影响力预测。

3.4 结果与分析

通过比较世博会对这三大指标的影响力指数, 可以发现上海世博会的申办成功在前期对固定资产投资的影响最大 (5.108), 对从业人员的影响次之 (4.767), 对生产总值的影响相比而言最小 (4.687), 而这与实际情况及理论分析亦是相符的。根据经济学基本理论, 固定资产投资是建造和购置固定资产的经济活动, 包括固定资产更新、改建、扩建、新建等活动。世博会的筹备阶段主要就是围绕新场馆的建设以及城市基础设施的更新、改建和扩建展开的, 而上海是一个发展趋于饱和的城市, 不同于正处于快速发展中的城市那样需要大力地投资于固定资产。因此世博会的筹备工作会瞬时带来较多的固定资产投资, 和城市的基础水平相比较后更突显世博的影响力。生产总值是反映一个地区经济状况的指标, 受到诸多因素的影响, 包括固定资产投资, 从业人数等, 因此其受世博的影响和固定资产投资这种单因素受世博影响相比较必然偏小。即便如此, 由生产总值时序模型也分析得到了世博会对生产总值 7.8% 的贡献量, 足以说明世博筹备对上海经济带来的巨大影响力。

4 模型二 基于主成分分析的会展即时影响模型

模型一衡量了上海世博会在筹备期间对国民经济指标产生的影响。而模型二衡量的是世博会期间会议本身的影响力。因此, 我们选取持续时间、参展国个数、参加人数、日均人数、占地面积、投资等几个指标进行分析, 用主成分分析法获得世博会影响力函数。与奥运会等大型活动不同, 世博会活动期间在参加人数和持续时间等方面有独特性, 因此我们需要选取若干届世博会的数据进行分析建立模型, 用该模型衡量此若干世博会的影响力大小, 作为上海世博会影响力大小的比较标准。

4.1 模型假设

1. 设定投资额为狭义投资, 即世博会的直接资金投入和运营成本, 配套设施产生的成本不予考虑。
2. 认为预测得到的上海世博会参与人数数据真实可靠。
3. 设定世博会的即时影响只受模型中列出的展会自身客观因素影响, 且可由这些因素充分影响, 不考虑其他因素。

4.2 数据处理

互联网上每届世博会的数据比较散乱, 这些数据主要来源于参考文献 [4]-[8]。考虑到不同货币间的汇率^[9]和不同年份间的通货膨胀^[10], 投资额一项数据需要预标准化处理成 2010 年同等购买力的美元值。

由于 2010 年上海世博会仍在举办中, 本届世博会的参与人数还无法得到。我们运用时间序列分析方法, 利用已有的 19 周周末入园人数数据预测出闭园前每每周的入园人数, 从而计算出总参与人数。对

样本为19周的参观人数数据，作一次差分得到增长率，为时间序列分析铺垫，两者趋势如图13和图14。

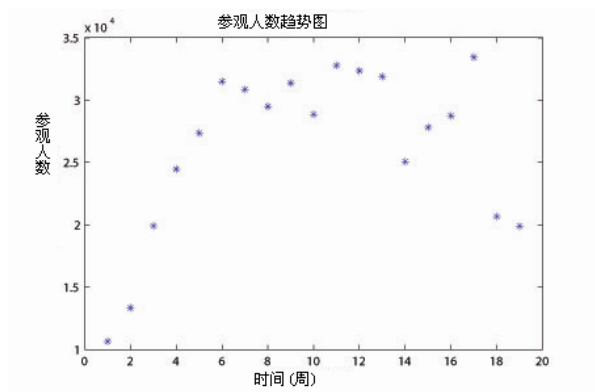


图 13 参观人数趋势图

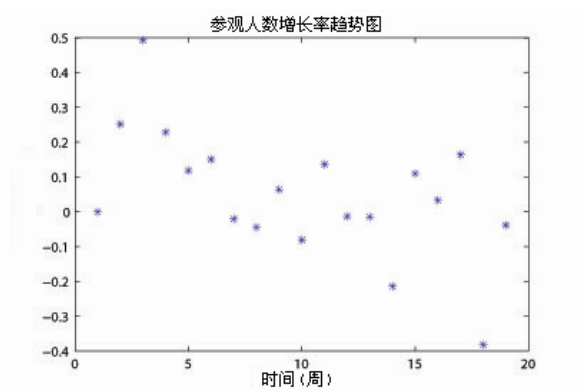


图 14 增长率趋势图

参观人数增长率时间序列较平稳，其中单位根检验显著（ $h=1$ ； $pValue=0.0049$ ），自相关图拖尾，偏相关图截尾（见附件）。根据经验和相关图的滞后性和拖尾性，选用ARMA(2,4)模型进行时间序列预测，预测模型如下

$$A(q)y(t) = C(q)e(t) \tag{9}$$

$$A(q) = 1 + \frac{0.6116}{q^2} - \frac{0.1305}{q} \tag{10}$$

$$C(q) = 1 + \frac{0.8949}{q^4} - \frac{0.6689}{q^3} + \frac{0.817}{q^2} - \frac{0.8125}{q} \tag{11}$$

Loss function 0.00529295 and FPE 0.00902916

该模型较准确地拟合了样本数据，并计算得到未来7周的参观人数增长率数据和参观人数值如表9。

表 9 参观人数预测

时间（周）	20	21	22	23	24	25	26
参观人数增长率	0.020	0.073	-0.061	-0.040	0.022	0.071	-0.031
参观人数预测值（百人）	20278.86	21763.22	20442.70	19619.33	20060.59	21483.77	20817.73

进而与前19周总参观人数加和得上海世博会预计总参观人数为6448万人。数据整合和初处理结果见表11。

表 10 变量说明

X_1	X_2	X_3	X_4	X_5	X_6
持续时间(天)	参展国个数	参与人数(万)	日均人数（万）	占地面积(英亩)	投资（亿美元）

4.3 模型建立与求解

对上表中的6项评价指标进行定性考察，不难发现指标之间存在较大的关联性。例如用来衡量展会成本规模的展区面积和投资量均为衡量世博会规模的直接变量，在表中有较强的相关性。如果直接用这些指标进行综合评估，会造成信息重叠，影响评估结果的客观性。因而要通过主成分分析法对6个指标进行降维，转化成不相关的几个综合指标^[11]。

表 11 数据整合表

	x_1	x_2	x_3	x_4	x_5	x_6
1958布鲁塞尔	186	54	5145	27.66	494	6.07
1970大阪	183	75	6422	35.09	865	391.23
1984新奥尔良	184	26	734	3.99	207	7.34
1992塞维利亚	176	112	4181	23.76	1181	155.39
1998里斯本	132	160	1000	7.58	172	26.75
2000汉诺威	153	172	1810	11.83	395	77.01
2005爱知	185	121	2200	11.89	460	35.61
2010上海	185	242	6448	34.85	1305	42.18

4.3.1 数据标准化

由于主成分分析的结果受量纲的影响，需要将各变量的数据标准化以后进行分析。

本例中进行主成分分析的指标共有6个，共有8个评价对象（8届世博会的不同值）。对第i个评价对象的第j个指标取值为 A_{ij} 。先将各指标值标准化如下：

$$\tilde{x}_{ij} = (x_{ij} - \bar{x}_j) / s_j, (i = 1, 2, \dots, 8; j = 1, 2, \dots, 6) \quad (12)$$

其中 $\bar{x}_{ij} = \frac{1}{8} \sum_{i=1}^8 x_{ij}$, $s_j = \sqrt{\frac{1}{7} \sum_{i=1}^8 (x_{ij} - \bar{x}_j)^2}$, ($j = 1, 2, \dots, 6$) 为第j个指标的样本均值和样本标准差称 $\tilde{x}_i = (x_i - \bar{x}_i) / s_i$, ($i = 1, 2, \dots, 6$)为标准化指标变量。

$$\bar{x}^T = (173.00, 120.25, 3492.50, 19.58, 634.87, 92.69) \quad (13)$$

$$s^T = (19.8853, 70.1727, 2355.6915, 12.3215, 431.8688, 129.9707) \quad (14)$$

指标变量标准化后如表12:

表 12 标准化的数据整合表

	x_1	x_2	x_3	x_4	x_5	x_6
1958布鲁塞尔	0.6536	-0.9441	0.6729	0.5987	-0.3262	-0.5691
1970大阪	0.5028	-0.6449	1.1929	1.1757	0.5329	2.2962
1984新奥尔良	0.553	-1.3432	-1.1233	-1.2392	-0.9907	-0.5596
1992塞维利亚	0.1508	-0.1176	0.2804	0.2955	1.2646	0.5417
1998里斯本	-2.0613	0.5665	-1.015	-0.9607	-1.0718	-0.4153
2000汉诺威	-1.0055	0.7375	-0.6851	-0.6304	-0.5554	-0.0413
2005爱知	0.6033	0.0107	-0.5263	-0.6256	-0.4049	-0.3494
2010上海	0.6033	1.7351	1.2035	1.1571	1.5517	-0.3005

4.3.2 计算相关系数矩阵

相关系数矩阵

$$R = (r_{ij})_{m \times m} \quad r_{ij} = \frac{\sum_{k=1}^8 \tilde{x}_{ki} \cdot \tilde{x}_{kj}}{n-1}, (i, j=1, 2, \dots, 6) \quad (15)$$

其中, $r_{ii} = 1$, $r_{ij} = r_{ji}$, r_{ij} 是第i个指标和第j个指标的相关系数。

得相关系数矩阵如下：

$$\begin{pmatrix} 1. & -0.365483 & 0.543746 & 0.489264 & 0.451134 & 0.156685 \\ -0.365483 & 1. & 0.159866 & 0.193812 & 0.402499 & -0.154575 \\ 0.543746 & 0.159866 & 1. & 0.997593 & 0.809041 & 0.525046 \\ 0.489264 & 0.193812 & 0.997593 & 1. & 0.815472 & 0.544479 \\ 0.451134 & 0.402499 & 0.809041 & 0.815472 & 1. & 0.412514 \\ 0.156685 & -0.154575 & 0.525046 & 0.544479 & 0.412514 & 1. \end{pmatrix}$$

4.3.3 主成分的选择及综合贡献值

易求得R的特征值和标准正交特征向量，并由此求得各个主成分的信息贡献率，如下式

$$b_j = \frac{\lambda_j}{\sum_{k=1}^6 \lambda_k} \quad (j = 1, 2, \dots, 6) \quad (17)$$

特征根及贡献率见表13。可以看出，前三个特征根的累计贡献率已达94%，主成分分析效果很好。

表 13 主成分参数

序号	特征根	贡献率	累计贡献率
1	3.4242	57.0698	57.0698
2	1.3839	23.0644	80.1342
3	0.831	13.8504	93.9846
4	0.2499	4.1651	98.1497
5	0.1108	1.8474	99.9971
6	0.0002	0.0029	100

选取这三个主成分进行综合评价。其对应的特征向量见表14。可得三个主成分分别为

表 14 特征向量

	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4	\tilde{x}_5	\tilde{x}_6
y_1	0.321109	0.087055	0.524998	0.523519	0.481414	0.328415
y_2	-0.49186	0.817399	0.016105	0.054065	0.229687	-0.18438
y_3	0.557034	0.113034	0.033079	-0.0168	0.142427	-0.80949

$$\begin{cases} y_1 = 0.3211\tilde{x}_1 + 0.0871\tilde{x}_2 + 0.5250\tilde{x}_3 + 0.5235\tilde{x}_4 + 0.4814\tilde{x}_5 + 0.3284\tilde{x}_6 \\ y_2 = -0.4919\tilde{x}_1 + 0.8173\tilde{x}_2 + 0.0161\tilde{x}_3 + 0.0541\tilde{x}_4 + 0.2297\tilde{x}_5 - 0.1843\tilde{x}_6 \\ y_3 = 0.5570\tilde{x}_1 + 0.1130\tilde{x}_2 + 0.0330\tilde{x}_3 - 0.0168\tilde{x}_4 + 0.1424\tilde{x}_5 - 0.8095\tilde{x}_6 \end{cases} \quad (18)$$

从主成分系数可以看出，第一主成分主要反映了后四个指标（参与人数、日均人数、占地面积和投资）的信息，体现了世博会的总体规模，第二主成分主要反映了参展国数量，第三主成分主要反映了展会持续时间。

将各次世博会6个指标的标准化数据分别代入主成分表达式，可以得到各次世博会的主成分值（表15）。以三个主成分的贡献率为权重，构造主成分综合评价模型：

$$Z = 0.5706y_1 + 0.2306y_2 + 0.1385y_3 \quad (19)$$

将每届世博会的主成分值代入上式，可以得到综合评价值和排名如表16。

表 15 各次世博会主成分

	y_1	y_2	y_3
1958布鲁塞尔	0.4633	-0.9986	0.7627
1970大阪	2.428	-0.9875	-1.5562
1984新奥尔良	-1.9091	-1.5637	0.5293
1992塞维利亚	1.1364	0.0543	-0.1356
1998里斯本	-2.3608	1.2548	-0.845
2000汉诺威	-1.2701	0.9467	-0.4714
2005爱知	-0.7593	-0.3427	0.6276
2010上海	2.2716	1.6366	1.0886

表 16 各届世博会总评价

	总评价	排名
2010上海	1.8246	1
1970大阪	0.9424	2
1992塞维利亚	0.6423	3
1958布鲁塞尔	0.1398	4
2005爱知	-0.4255	5
2000汉诺威	-0.5718	6
1998里斯本	-1.1749	7
1984新奥尔良	-1.3769	8

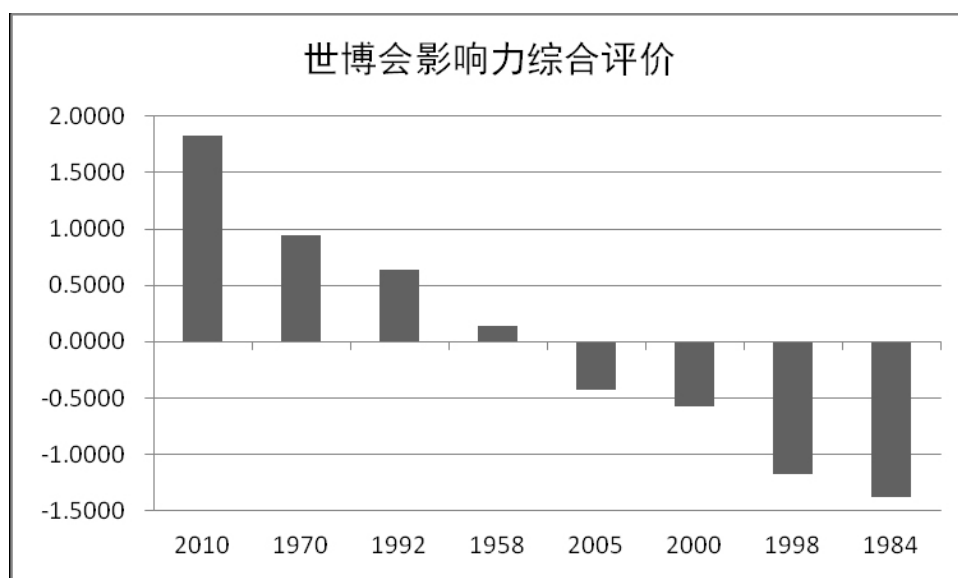


图 15 模型二结果

本过程的matlab代码如下

```
clc,clear
load data.mat %提取原始数据
data=zscore(data); %数据标准化
r=corrcoef(data); %计算相关系数矩阵
[x,eig,z]=pcacov(r); %主成分分析
f= repmat(sign(sum(x)),size(x,1),1);
x=x.*f;
df=data*x(:,1:3)
tf=df*z(1:3)/100;
[stf,ind]=sort(tf,'descend')
```

4.4 结论与有效性分析

通过上述分析可以看出,2010上海世博会的即时影响力综合评价值为1.8246,在选取的样本中最大,主要表现在其三个主成分的值在几届世博会中都处于领先地位。1970年大阪世博会和1992年塞维利亚世博会也有巨大的影响力,尤其是大阪世博会在第一主成分的评分排名第一,说明其规模巨大,从而产生了较为深远的影响力。考虑到大阪和塞维利亚两届世博会分别推动了日本关西经济带和西班牙南部地区的发展以及两届世博会获得的广泛好评,本模型中对这两届世博会的评价与展会后的广泛评价相一致。另外,1984新奥尔良世博会组织不力,出现了严重的亏损,是公认的影响力较弱的一次世博会。这一点与本模型中1984世博会的低评分相一致。

本模型通过展会自身展览期间的数据获得了客观的影响力评价,且提供了可靠的量化标准。

5 模型三 基于Logistic的后续影响模型

模型一和模型二分别通过不同方式研究了世博会筹备期间和开会期间影响力的不同形式及作用大小,而模型三则着重分析并预测了世博会结束后影响力的主要表现形式及作用大小。相对于前两个阶段而言,世博会正式会期结束之后的一段时间内,世博会的影响力由对经济和建设的直接影响,转变为对于社会文化、媒体舆论、民众认识的间接影响^[12]。针对这样一种影响力的作用模式,模型三选取一段时间内的网络相关文章总数(基于Google搜索引擎)为主要标准,建立了Logistic回归模型,对上海世博会的后续影响进行了分析。

随着信息时代的不断发展,网络已经成为文化传播最重要的途径,因此世博会会期结束后的影响可以用网络对世博会的反应来描述。Google作为当今最大的搜索提供商,每时每刻不断抓取网络上的各种信息,并转化为搜索搜索结果条目。通过Google不同时间抓取的同一主题的文章篇数可以比较准确地反应网络的总体趋势。因此,为了表示世博会结束后世博会对文化和理念的影响,选取Google-Timeline历史搜索中不同时间的相关文章篇数作为评价指标。

5.1 模型假设

1. Google提供的搜索结果可有效而可靠地代表真实信息量。
2. 每用户对搜索量做出的贡献相同,每用户传播信息的能力相同。
3. 对世博会、奥运会等”特大事件”,基于信息的后续影响规律相同。
4. 模型中提供的搜索关键字基本涵盖了世博会(奥运会)相关搜索。

5.2 模型的建立

由于每用户传播信息的能力相同，设信息在用户间的传播速率为 r ，建立微分方程^[13]如下

$$\frac{dy}{dx} = pry \quad (20)$$

其中 p 为每用户提供的搜索次数。

当信息传播接近饱和，系统对搜索量的增长显现出明显的阻抑作用，故采用工程师原则引入竞争项 qy ：

$$\frac{dy}{dt} = p(r - qy)y, \text{ 即 } \frac{dy}{dx} = k(K - y)y$$

该初值问题的解为

$$y = \frac{a}{1 + be^{-cx}} \quad (21)$$

为典型的Logistic模型。

5.3 模型的可行性验证

基于假设3，奥运会的后续影响机制可认为与世博会相同。而且考虑到北京奥运会已结束两年，从数据可以看出搜索量的增长已趋于平稳，从而可以获得较为完整的搜索量曲线，做出较完整的可行性验证。

我们搜索关键词“北京OR 中国+奥林匹克运动会”，时间段限制为2007年1月-2010年3月，并对数据进行了非线性回归（代码和数据见附录）。得出的结果如下：由图16可以看出，北京奥运会实际网

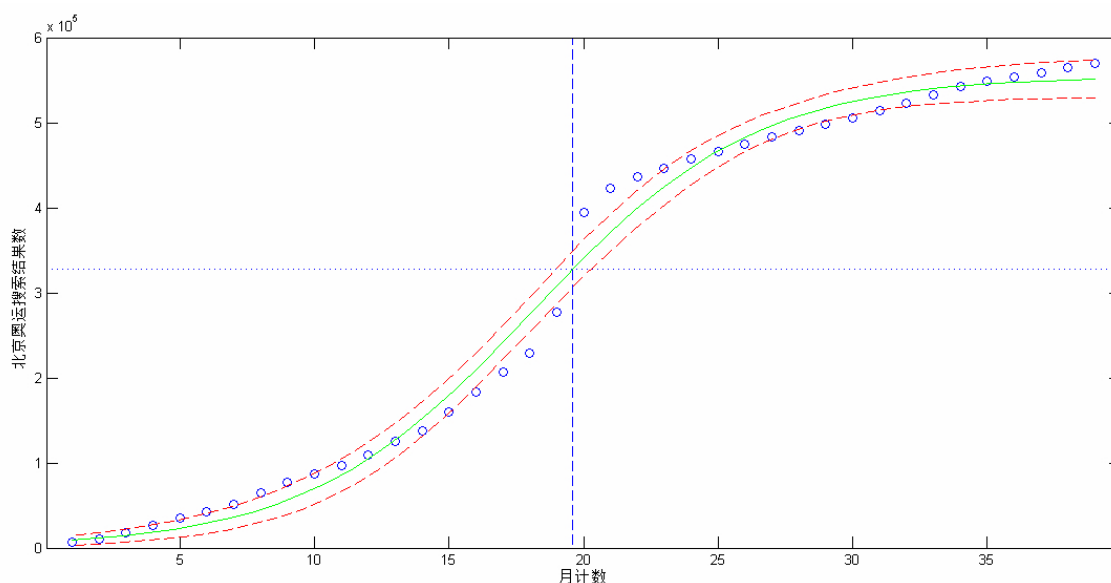


图 16 北京奥运后期影响力Logistic回归

上文章量与Logistic模型回归量基本吻合，绝大多数数据点落在95%置信区间中，可以很好的描述北京奥运会后期文化影响力指数。因此，此模型可以用来研究世博会后期影响。

5.4 模型的应用

在将以上模型应用于上海世博会的后续影响分析之前，首先要获取上海世博会搜索关键字的统计信息。为了选取最有效的搜索关键词，我们在Google上进行了试验和统计，结果见表17。由表17可以看出，“上海OR中国世界博览会”这一组关键词可以最好地涵盖这次世博会的网上相关文章，所以选择这一组关键词在google.com.hk进行历史搜索。同理在google.com使用“Shanghai or China expo”搜

表 17 关键词搜索结果

中文google.com.hk		
关键词	总量(条)	备注
上海+世博	143000	无
中国+世博	106000	无
上海+世界博览会	227000	包含了“上海+世博会”的结果
中国+世界博览会	179000	包含了“中国+世博会”的结果
上海OR中国+世界博览会	267000	包含上述所有结果

索相关的英文文章。为了保证Logistic模型建立的准确性，将时间跨度设置为09年01月-10年08月，并按时间顺序将月份设置为1-20。

表 18 历史搜索结果

月计数	1	2	3	4	5	6	7	8	9	10
中文结果数	1430	3270	6440	10590	14610	18090	22400	25520	30200	35010
英文结果数	201	455	838	1201	1610	1996	2434	2758	3233	3761
月计数	11	12	13	14	15	16	17	18	19	20
中文结果数	40400	45810	54460	59910	73010	100110	125610	136310	149410	161610
英文结果数	4418	4996	6006	6856	8456	12196	16996	18826	21046	22826

现用Logistic模型预测上海世博会后续影响指数。定义 y 为世博会某时刻的后续影响力指数，由其在Google-Timeline中某一时间点的搜索结果总数表示。则根据5.2中推导的Logistic模型， y 与月计数 x 的关系可以表示为：

$$y = \frac{a}{1 + be^{-cx}} \quad (22)$$

其中 a 为常数，表示此网络系统中关键词累计搜索量上限在Matlab中对表X中的数据进行非线性回归(代码见附录：logistic.m)，可以得到模型的参数回归值。

$$\text{针对中文结果: } \begin{cases} \hat{a} = 340200 \\ \hat{b} = 84.7791 \\ \hat{c} = 0.2205 \end{cases} \quad \text{针对英文结果: } \begin{cases} \hat{a} = 51468 \\ \hat{b} = 172.7752 \\ \hat{c} = 0.2504 \end{cases} \quad (23)$$

由回归结果可见，实际值基本上都落在回归值95%的置信区间中，说明用Logistic模型可以很好地描述世博会在会后的影响力变化趋势。下图为中文文章数的回归曲线，英文文章数的回归曲线与中文的基本一致（见附录）。

5.5 模型结果分析

通过对上述模型结果的分析主要可以得到以下结论：

第一，上海世博会对文化的影响力基本可以用Logistic模型来描述。会期趋向中后期时，其影响力指数的增长速度也会逐渐衰减。世博会后期影响力的主要作用是在文化方面，由此可知世博会的影响力总量在会议中后期及会议结束后的一段时间仍会有所增长，但增速会逐渐趋缓，符合Logistic回归曲线。

第二，通过对中文搜索条目数和英文搜索条目数回归结果的比较可以看出，本次上海世博会的影响力主要还是作用于中国国内，对英语国家的影响较小。比较分别对中文搜索和英文搜索结果得到的回归参数可以发现， $\hat{a} = 340200$ ， $\hat{a} = 51468$ ，则 $\frac{\hat{a}}{\hat{a}} = 6.6$ 。所以可以估算得此次上海世博会的影响力尤

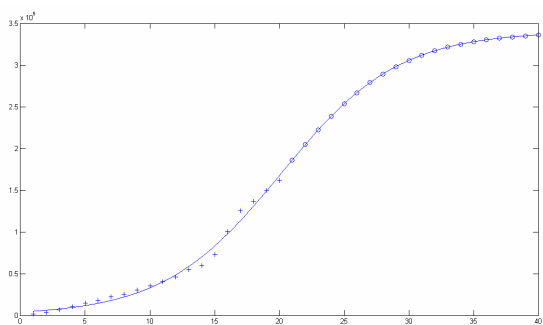


图 17 中文搜索结果量Logistic回归

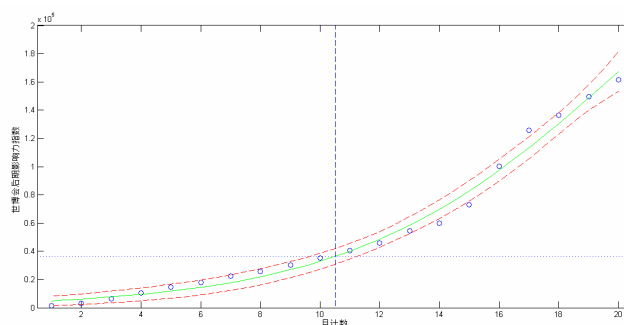


图 18 中文搜索结果量Logistic回归置信区间

其是会后对文化与价值观的影响力方面，对国内的影响力是对英语国家影响力的6倍，且搜索量临界值为340200，即为上海世博会的后续影响力指数。

第三，通过比较上海世博会与北京奥运会的回归结果，可以从文化影响的侧面比较两者影响力的大小。 $\hat{a} = 340200$ ， $\hat{a} = 555050$ 。可以看出，两者相比之下，还是奥运会的影响力较大，但是因为两者都是大型的国际节事，所以影响力并没有太大的差异。

第四，假设 y 值达到0.99a时，搜索量基本达到临界值， y 的梯度值趋近于0。在(22)式中求得 $x=40.98$ ，即大约41个月，以搜索量为标准的后续影响力减弱为0。

6 模型的评价与改进

6.1 优点评述

本文在时间维度下建立了三个模型，分别从不同角度量化衡量了上海世博会的影响力。

模型一着重研究了世博会前期的经济效应，得到其对生产总值、从业人员和固定资产投资的影响力指数分别为5.108,4.767,4.867。其中，世博会对生产总值的贡献率为7.8%，说明世博会对上海经济发展有着巨大的推动作用，尤其在固定资产投资方面。

模型二着重研究了上海世博会的即时影响力，得到影响力指数为1.8246，说明上海世博会远大于邻近几届世博会。

模型三着重研究了世博会的后续影响力，得影响力指数为340200，说明上海世博会的后续影响弱于北京奥运会，并且影响力成S型曲线，在41月后基本消退。

6.2 模型的改进

本文的模型并未考虑世博会的负面影响，例如世博会造成的财政压力和劳动力分配不均等社会问题会不同程度地抵消世博会的正面影响。对于世博会负面影响的估算由于数据的缺乏没有在本文中实现，可以在后续研究中完成。

模型二、三由于事件自身的条件限制，能够获取的数据较少。如果能获得更多数据，进行主成分分析和Logistic非线性拟合时能取得更高的置信度。

参考文献

- [1] 约翰·艾伦, 大型活动项目管理, 2002, 机械工业出版社, p13
- [2] 上海统计局, 2010年上海统计年鉴
<http://www.stats-sh.gov.cn/2004shtj/tjnj/tjnj2010.htm>, 2010.9.10-2010.9.12
- [3] 李霞, 2008年奥运会对北京经济增长影响研究, 天津财经大学: 数量经济学, 硕士学位论文: 26-35, 2008
- [4] 上海世博会事务协调局, 上海世博会即时参观人数
<http://www.expo2010.cn/yqkl/indexn.htm>, 2010.9.10-2010.9.12
- [5] 唐子来, 陈琳, 世博会的经典案例研究之二: 1998年里斯本世博会, 城市规划汇刊, 2 (150): 13-16, 2004
- [6] 唐子来, 付磊, 世博会的经典案例研究之一: 1970年大阪世博会, 城市规划汇刊, 1 (149): 6-9, 2004
- [7] 唐子来, 赵渺希, 世界博览会的经典案例研究之三: 1992年塞维利亚世博会, 城市规划汇刊, 3 (151): 39-42, 2004
- [8] 唐子来, 朱弋宇, 世界博览会的经典案例研究之四: 2000年汉诺威世博会, 城市规划汇刊, 4 (152): 49-52, 2004
- [9] Oanda Currency Converter, 货币转换器,
<http://www.oanda.com/currency/converter/>, 2010.9.10-2010.9.12
- [10] Inflation Calculator, 美国通货膨胀计算器,
<http://data.bls.gov/cgi-bin/cpicalc.pl> 2010.9.10-2010.9.12
- [11] 张德丰, Matlab语言语言高级编程, 机械工业出版社, 2010, P178
- [12] 戴光全, 重大事件对城市发展及城市旅游的影响研究—以'99昆明世界园艺博览会为例, 北京: 中国旅游出版社, 2005, P41
- [13] 杨启帆, 谈之奕, 何勇, 数学建模, 杭州: 浙江大学出版社, 2006, P40-P45

附录

代码

相关图检验，AR模型预测

```
autocorr(g(1:26))
parcorr(g(1:26))
th=ar(g(1:26),12);
p=predict(th,g,6);
```

生产总值一元线性回归

```
x=x';y=y';
x=[ones(26,1),x];
b,bint,r,rint,stats
=regress(y,x);
b,bint,stats,rcoplot(r,rint)
```

单位根检验

```
[h,pValue,stat,cValue,reg] = adftest(y(1:26))
th=ar(y(1:26),12);
p=predict(th,y,6);
```

二次多项式拟合

```
[p,s]=polyfit(x,y,2);
f = polyval(p,x);
table = [x y f y-f]
```

Logistic回归

```
clc;clear;clf;
load dataac.mat
x=data(1,:);
y=data(2,:);
P=@(beta,x)beta(1)./(1+beta(2)*exp(-beta(3)*x));
beta,r,j,COVB,MSE
=nlfit(x,y,P,[200000 1 0.05]);
betaci=nlparci(beta,r,'jacobian',j);%beta回归值95%置信区间
betaa=[beta',betaci] [yhat,delta]=nlpredci(P,x,beta,r,'jacobian',j);%函数回归值及95%置信区间
nlintool(x,y,P,beta);
x0=1:0.02:40;y0=P(beta,x0);
x1=21:40;y1=P(beta,x1);
plot(x0,y0);
```

```
hold on;
plot(x,y,'+');
plot(x1,y1,'o');
```

生产函数模型中回归分析

SAS

```
Data economy; input Y L K X @@;
datalines;
4.695391324 4.562459357 4.819498178 1
4.710574829 4.69163869 4.961129385 2
4.716335606 4.690255777 5.062845228 3
4.733034847 4.755175307 5.175834288 4
4.673748796 4.879467372 5.227583128 5
4.62007536 5.000083459 5.312813173 6
;
proc reg;
model Y= K L X;
run;
```

图片与表格

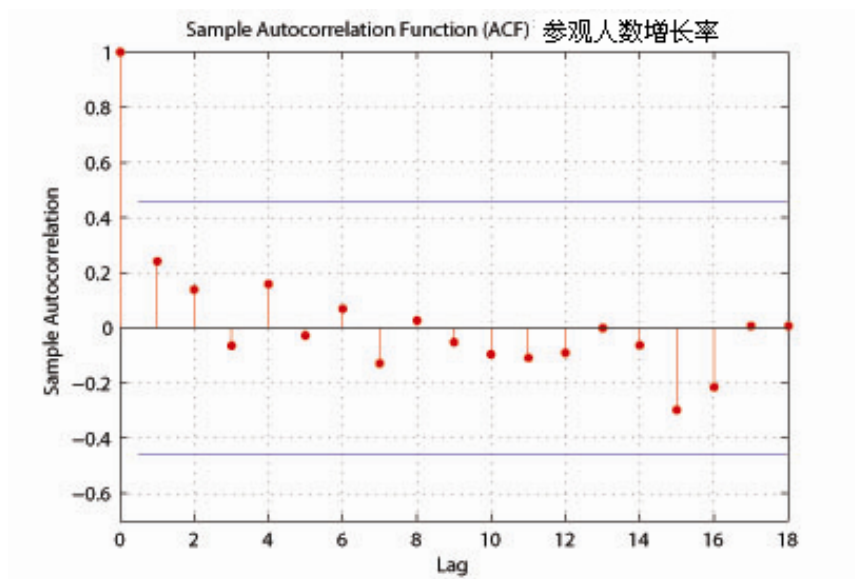


图 19 ACF参观人数增长率

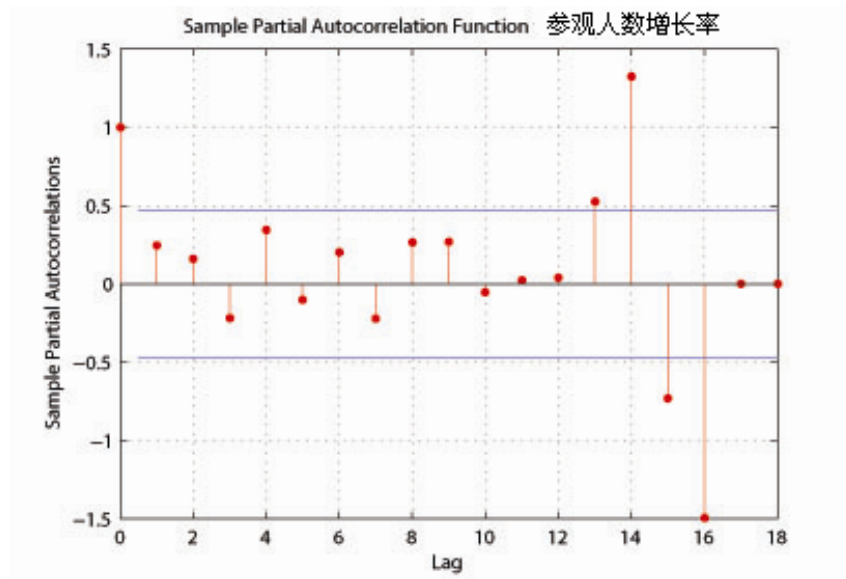


图 20 PCF参观人数增长率

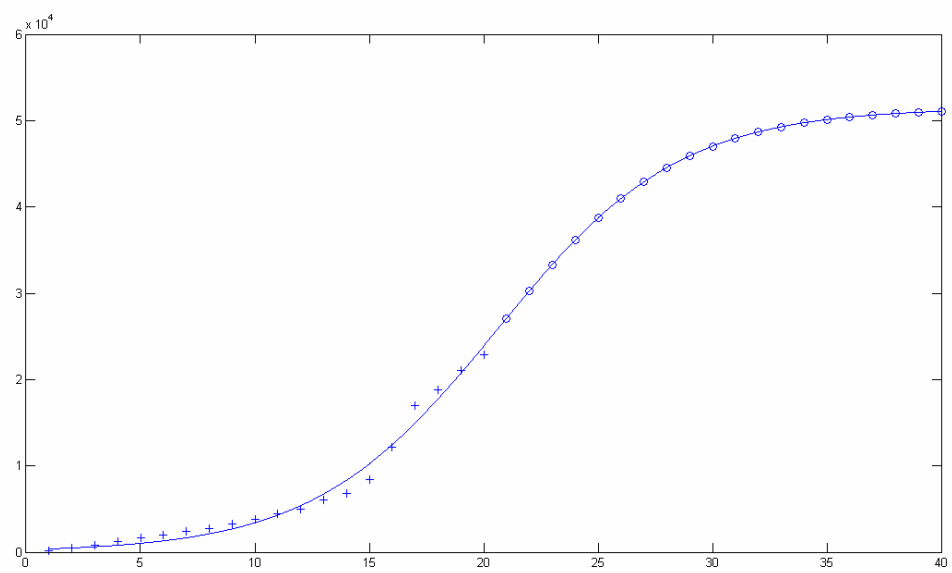


图 21 英语搜索结果数Logistic回归

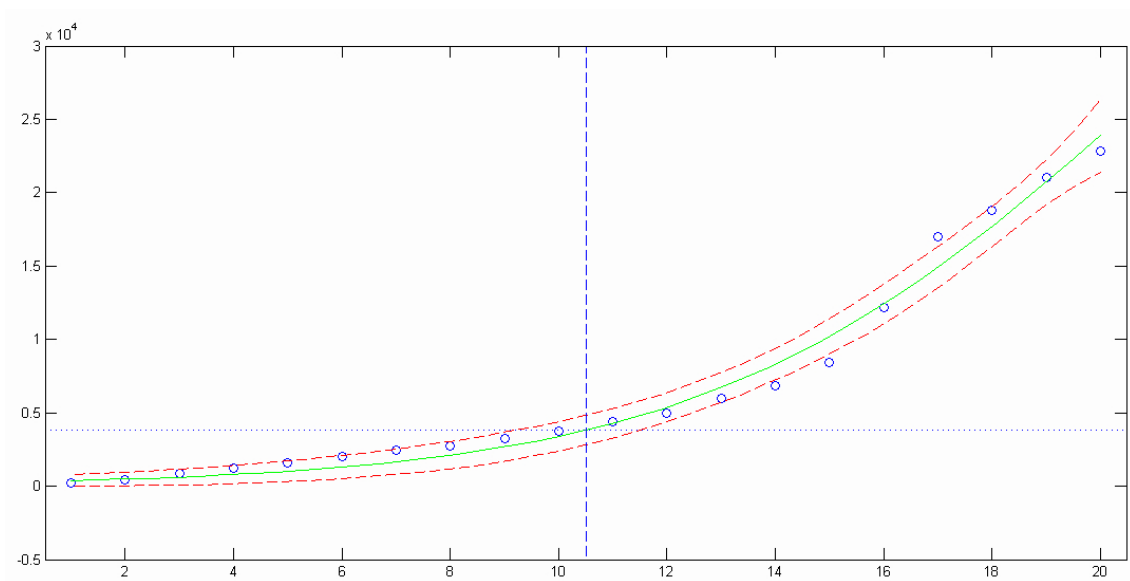


图 22 英语搜索结果数Logistic回归置信区间

表 19 固定资产投资指数拟合分析

商品零售价	固定资产投资	预测固定资产投资	误差
420.9	100	110.7451	-10.7451
441.9	107	104.7471	2.2529
436.6	107.5	106.6214	0.8786
415.2	105.8	111.7138	-5.9138
404	103.8	112.7972	-8.9972
389.5	103.8	112.5853	-8.7853
384	104.5	112.0283	-7.5283
379	104.8	111.2946	-6.4946
375.4	107.4	110.6321	-3.2321
378.8	114.6	111.2607	3.3393
376.7	115.5	110.8843	4.6157
377.4	115.6	111.014	4.586
386.5	119.6	112.314	7.286
407.1	129	112.6061	16.3939
404.8	125.1	112.7559	12.3441

表 20 拟合1978-1994数据(标准化后总数据)

年份	固定资产投资指数	lnzt	lnzt 增长率
1978	100	3.328985	0
1979	102.5806	3.546305	0.065281
1980	119.4016	3.63885	0.026096
1981	123.4226	3.78959	0.041425
1982	124.174	4.050943	0.068966
1983	124.4243	4.111416	0.014928
1984	130.4059	4.259563	0.036033
1985	174.6063	4.218056	-0.00974
1986	194.484	4.324777	0.025301
1987	221.4678	4.432251	0.024851
1988	285.9006	4.451886	0.00443
1989	338.9658	4.148792	-0.06808
1990	354.9399	4.158524	0.002346
1991	385.0576	4.205899	0.011392
1992	414.1715	4.45769	0.059866
1993	457.1291	4.963174	0.113396
1994	482.9243	5.449327	0.097952
1995	472.93	5.8251	0.068958
1996	479.93	6.008165	0.031427
1997	480.465	6.02005	0.001978
1998	478.7444	6.017164	-0.00048
1999	476.73	5.964787	-0.0087
2000	476.734	5.971729	0.001164
2001	477.4606	6.034953	0.010587
2002	477.7742	6.126345	0.015144
2003	480.2904	6.235483	0.017815
2004	487.4836	6.45011	0.03442
2005	488.4	6.586637	0.021167
2006	488.5155	6.688944	0.015532
2007	492.53	6.808207	0.01783
2008	501.9784	6.869101	0.008944
2009	498.03	6.964927	0.01395

表 21 生产函数模型中回归分析各参数统计数据

Variable	DF	ParameterEstimate	Standard Error	t Value	Pr > t
Intercept	1	1.79585	1.48302	1.21	0.3496
K	1	0.91027	0.21847	4.17	0.0531
L	1	-0.30927	0.14867	-2.08	0.173
X	1	-0.07643	0.029	-2.64	0.1189