

文章编号:1005-0523(2000)01-0083-06

我国铁路煤炭运量预测模型的研究

曾佑新¹, 刘海燕¹, 周尚超²

(华东交通大学 1. 经济管理学院 2. 科研处, 江西南昌 330013)

摘要: 本文用一元线性回归、直线趋势、自回归和二次指数平滑等四种预测模型对我国铁路煤炭运量进行了预测, 并对四种预测模型的误差作了比较, 目的在于对煤炭运量作出科学的预测, 提高精度, 实现煤炭的合理运输¹⁹。

关键词: 自回归; 二次指数平滑; 铁路煤炭运量预测; 预测误差

中图分类号: F530.34 **文献标识码:** A

煤炭运量的变化对铁路货运总量的影响是最大的, 自1980年以来煤炭运量占铁路货运总量的比重一直保持在40%左右¹⁹。本文用一元线性回归、直线趋势、自回归和二次指数平滑等四种预测模型对我国铁路煤炭运量进行了预测, 并对四种预测模型的误差作了比较, 目的在于对煤炭运量作出科学的预测, 提高精度, 实现煤炭的合理运输¹⁹。

1 利用一元线性回归与相关分析预测煤炭运量

依据1978—1998年铁路煤炭运量与国内生产总值的统计资料(见表1), 可以看出, 煤炭运量依赖国内生产总值的水平, 二者存在较为明显的直线相关关系¹⁹。因此, 可以用一元线性回归与相关分析预测煤炭运量¹⁹。

1.1 确定煤炭运量与国内生产总值之间的相关系数

利用表1的资料, 计算出煤炭运量与国内生产总值之间的相关系数 $r = 0.8186$, 说明二者是正相关, 而且二者之间线性关系显著¹⁹。

1.2 建立煤炭运量与国内生产总值的一元线性回归方程

以上分析说明煤炭运量与国内生产总值之间关系, 配合一元线性回归方程

$$\hat{y} = a + bx$$

式中: \hat{y} 为各年煤炭运量的回归值; x 表示每年的国内生产总值; a, b 为2个待定参数⁽¹³⁾

确定回归方程参数 a, b 最主要的方法是最小平方法⁽¹³⁾。根据表1的资料, 计算得出: $b = 0.3449$, $a = 47247$ ⁽¹³⁾。回归系数 $b = 0.3449$ 说明国内生产总值每增加1亿元, 煤炭运量平均会增加0.3449万吨⁽¹³⁾。

综上所述, 可以得出我国煤炭运量对国内生产总值的线性回归主程为

收稿日期:1999-07-12; 修订日期:1999-08-31

部级项目:铁道部科技司科研项目(97Y18/A)

作者简介:曾佑新(1962-), 男, 广东兴宁人, 华东交通大学副教授¹⁹。

$$\hat{y} = 47\ 247 + 0.344\ 9x$$

将各年的国内生产总值代入回归方程,就可计算出每年煤炭运量的估计值 \hat{y} ,列在表1中第4栏内,估计值 \hat{y} 与实际值 y 的误差在表1中第5栏内^[13]。如果1998年以后国内生产总值平均增长速度按7%计算,则2000年铁路煤炭运量预测值为78 660万吨,2005年铁路煤炭运量预测值为91 306万吨^[13]。

2 利用直线趋势模型预测煤炭运量

利用表2的统计资料,如果以年份为横坐标,铁路煤炭运量为纵坐标作图形,那么铁路煤炭运量曲线近似一条直线^[13]。因此,预测未来铁路煤炭运量,可以采用直线趋势法^[13]。

直线趋势的方程式是: $Y(t) = a + bt$

其中 t 代表年份, $Y(t)$ 为第 t 年预测运量,在模型中取 $t = 1$ 表示1978年^[13]。方程式中 a 、 b 为待定系数,根据过去若干年的数据确定,这里我们选用了最小平方方法确定^[13]。根据1978年到1999年的煤炭运量算出 $a = 37\ 682$, $b = 1\ 657$,这样预测第 t 年的煤炭运量 $Y(t)$ 由下式计算:

$$Y(t) = 37682 + 1657t$$

说明煤炭运量每年平均增加1657万吨^[13]。利用方程可计算出1978年到2000年铁路煤炭运量的预测值和预测误差,详见表-2^[13]。

3 利用自回归模型预测煤炭运量

任何现象从一个时期到另一个时期的运动,总是要受到多种因素的影响,其中有些因素的影响不是瞬间即逝的,而是持续相当长的时间^[13]。就铁路煤炭运量自身来考虑,随时间的不同,在其前后期的数值之间也表现出一定的依存关系^[13]。因此,可以利用自相关与自回归来预测我国煤炭运量^[13]。

3.1 煤炭运量自相关系数的测定

假设 Y_t 为原数列, Y_{t-k} 表示移后期为 k 的时期数列,则 Y_t 与 Y_{t-k} 之间的自相关系数为 r_k ^[13]。利用历年煤炭运量的实际值,我们计算了 r_1 和 r_5 ^[13]。

$r_1 = 0.973\ 3$,若取显著水平 $\alpha = 0.01$,自由度为 $(n - 1) - 2 = 20 - 1 - 2 = 17$,查相关系数临界值表得 $r_{0.01} = 0.575$,因为 $r_1 > r_{0.01}$,所以 Y_{t-1} 与 Y_t 之间线性关系高度显著^[13]。

$r_5 = 0.927\ 8$,若取显著水平 $\alpha = 0.01$,自由度为 $(n - 5) - 2 = 20 - 5 - 2 = 13$,查相关系数临界值表得 $r_{0.01} = 0.641$,因为 $r_5 > r_{0.01}$,所以 Y_{t-5} 与 Y_t 之间也线性关系高度显著^[13]。

3.2 建立煤炭运量的自回归方程

当 Y_{t-k} 与 Y_t 之间线性关系显著时,就因以 Y_{t-k} 为自变量, Y_t 为因变量,建立自回归方程:

$$\hat{Y}_t = b_0 + b_1 Y_{t-k}$$

式中 b_0 、 b_1 可用最小平方方法求得^[13]。

(1) 当 $k = 1$ 时, $b_1 = 0.923\ 8$, $b_0 = 5\ 487.55$,一阶自回归方程为: $\hat{Y} = 5\ 487.55 + 0.923$

根据自回归方程得到估计值 \hat{Y}_{t-1} , 列在表 3(13)

(2) 当 $k = 5$ 时, $b_1 = 0.7278$, $b_0 = 22800.66$, 一阶自回归方程为: $\hat{Y} = 22800.66 + 0.7278Y_{t-5}$, 根据自回归方程得到估计值 \hat{Y}_{t-5} , 列在表 3(13)

4 利用二次指数平滑法预测煤炭运量的发展趋势

煤炭运量数列存在直线上升趋势, 可以用二次指数平滑法进行预测(13)但二次指数平滑本身并非预测公式, 而是对一次指数平滑值进行调整, 然后结合一次指数平滑值来估计直线方程的 2 个参数(13)

用二次指数平滑法求直线趋势方程的公式如下: $\hat{X}_{t+T} = a_t + b_t T$

式中 T 表示从 t 期起推向未来的期数(13)

4.1 计算步骤及计算公式

第一步: 计算一次指数平滑值 $S_t^{(1)}$ $S_t^{(1)} = \alpha X_t + (1 - \alpha)S_{t-1}^{(1)}$ 式中: $S_t^{(1)}$ 为第 t 期的一次指数平滑值; α 为平滑系数, 取值在 0-1 之间, α 取值小, 对原数列修均作用较强, 但对原数列发生的变化反映迟钝; 若原数列波动较大时, 为了较为灵敏地反映原数列的变化, α 取值可大些(13) X_t 为第 t 期的实际水平; $S_{t-1}^{(1)}$ 为第 $t-1$ 期的一次指数平滑值, 其初始值 $S_0^{(1)}$ 的确定方法下面另有说明(13)

第二步: 计算二次指数平滑值 $S_t^{(2)}$ $S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha)S_{t-1}^{(2)}$ 式中: $S_t^{(2)}$ 为第 t 期的二次指数平滑值; $S_{t-1}^{(2)}$ 为第 $t-1$ 期的二次指数平滑值(13)其初始值 $S_0^{(2)}$ 的确定方法另作说明(13)

第三步: 计算参数 α $\alpha = 2S_t^{(1)} - S_t^{(2)}$

第四步: 计算参数 b_t $b_t = \alpha(1 - \alpha)(S_t^{(1)} - S_t^{(2)})$

第五步: 计算预测值 \hat{X}_{t+T} $\hat{X}_{t+T} = a_t + b_t T$

初始值的确定可用下列方法: $S_0^{(1)} = \hat{a} - (1 - \alpha)\hat{b}/\alpha$ $S_0^{(2)} = \hat{a} - 2(1 - \alpha)\hat{b}/\alpha$

其中参数 \hat{a} 、 \hat{b} 可模拟直线回归方程, 用最小平方求出, 公式如下:

$$\sum X_t = n\hat{a} + b\hat{\Sigma}_t, \quad \sum tX_t = a\hat{\Sigma}_t + b\hat{\Sigma}_t^2$$

4.2 预测煤炭运量

1978 ~ 1998 年铁路煤炭运量, 除个别年份外, 均有增长, 故原数列呈线性增长趋势(13)逐期增量也有一定的波动, 为了较灵敏地反映煤炭运量的变化, 确定平滑系数 $\alpha = 0.01$ (13)

根据表 4(二次指数平滑预测计算表) 资料, 计算得出: $\hat{a} = 37682$ $\hat{b} = 1657$

因此, 二个初始值分别计算如下:

$$S_0^{(1)} = \hat{a} - (1 - \alpha)\hat{b}/\alpha = 22769 \quad S_0^{(2)} = \hat{a} - 2(1 - \alpha)\hat{b}/\alpha = 7856$$

平滑系数和初始值确定后, 便可计算历年的一次指数平滑值和二次指数平滑值以及 a_t 、 b_t , 详见表 4, 利用 a_t 、 b_t 还可得出各年的预测值 \hat{X}_{t+T} , 进而可得误差(13)

利用最后一年的参数即 a_{98} 、 b_{98} 可对 1998 年以后各年的煤炭运量进行预测, 详见表 4(13)

5 各种预测模型预测误差的比较

前面我们已经分别用一元线性相关与回归分析、直线趋势法、自相关与自回归分析和二

次指数平滑法等四种方法对铁路煤炭运量进行了分析和预测⁽¹³⁾。四种预测方法都有预测误差,但误差有大有小,可以分别计算出它们的标准误差,并据以比较四种预测模型的误差大小,详见表 5⁽¹³⁾。

比较结果表明,二次指数平滑预测和移后一期的自相关与自回归预测效果较好,直线趋势预测和移后五期的自相关与自回归预测效果次之,一元线性相关与回归分析的预测效果最差⁽¹³⁾。这是因为铁路煤炭运量除了受经济发展的影响,还要受各种运输方式的竞争、原煤产量等多种因素的影响,而我们在模型中只考虑了经济发展一种主要因素的影响,而未考虑其他因素的影响,因此出现较大误差⁽¹³⁾。

直线趋势预测中采用的是一般最小平方法,将数列中的各项观察值视作同等重要,事实上,数列中各期观察值对未来的影响作用是不能等同看待的⁽¹³⁾。其中近期观察值比远期观察值对未来的影响大得多,一般最小平方法由于未考虑这种影响,使直线趋势预测出现较大误差⁽¹³⁾。

移后一期的自相关与自回归预测误差较小,说明煤炭运量每一年为一个周期,用上一年度的实际数据预测下一年的运量将会有较好的效果,但这只能用于短期的年度预测,难以进行中长期预测⁽¹³⁾。

二次指数平滑预测考虑了以前各年发展水平对未来的影响,根据逐远逐减的原则,给予以前各期发展水平相应的权数,近期权数大、远期权数小,因而有较好的预测效果,当列数水平发展较平稳时,还可以进行中长期预测⁽¹³⁾。

表 1 煤炭运量一元线性回归方程计算表

日期/年	运量/万吨	预测运量/万吨	误差率/%
1978	40 243	48 497	20.51
1979	41 319	48 640	17.72
1980	41 499	48 805	17.61
1981	41 199	48 924	18.75
1982	43 846	49 073	11.92
1983	46 020	49 294	7.11
1984	48 708	49 720	2.08
1985	51 856	50 339	-2.93
1986	53 106	50 766	-4.41
1987	54 343	51 373	-5.47
1988	56 406	52 396	-7.11
1989	60 891	53 079	-12.83
1990	62 870	53 644	-14.67
1991	62 603	54 703	-12.62
1992	64 108	56 434	-11.97
1993	65 336	59 192	-9.40
1994	65 943	63 374	-3.90
1995	67 357	67 416	0.09
1996	72 058	70 661	-1.94
1997	70 264	73 036	3.95
1998	64 081	74 685	16.55

表 2 煤炭运量直线趋势方程计算表

日期/年	t	运量/万吨	预测运量/万吨	误差率/%
1978	1	40 243	39 339	-2.25
1979	2	41 319	40 996	-0.78
1980	3	41 499	42 653	2.78
1981	4	41 199	44 310	7.55
1982	5	43 846	45 967	4.84
1983	6	46 020	47 624	3.49
1984	7	48 708	49 281	1.18
1985	8	51 856	50 938	-1.77
1986	9	53 106	52 595	-0.96
1987	10	54 343	54 252	-0.17
1988	11	56 406	55 909	-0.88
1989	12	60 891	57 566	-5.46
1990	13	62 870	59 223	-5.80
1991	14	62 603	60 880	-2.75
1992	15	64 108	62 537	-2.45
1993	16	65 336	64 194	-1.75
1994	17	65 943	65 851	-0.14
1995	18	67 357	67 508	-0.22
1996	19	72 058	69 165	-4.01
1997	20	70 264	70 822	0.79
1998	21	64 081	72 479	13.11

表 3 自相关与自回归方程计算表

日期 / 年	运量 / 万吨	Y_{t-1}	\hat{Y}_{t-1}	误差	误差率 / %	Y_{t-5}	\hat{Y}_{t-1}	误差	误差率 / %
1978	40 243								
1979	41 319	40 243	40 243	1 345	3.26				
1980	41 499	41 319	41 319	2 159	5.20				
1981	41 199	41 499	41 499	2 625	6.37				
1982	43 846	41 199	41 199	- 299	- 0.68				
1983	46 020	43 846	43 846	- 28	- 0.06	40 243	52 090	6 070	13.19
1984	48 708	46 020	46 020	- 707	- 1.45	41 319	52 873	4 165	8.55
1985	51 856	48 708	48 708	- 1 372	- 2.65	41 499	53 004	1 148	2.21
1986	53 106	51 856	51 856	286	0.54	41 199	52 785	- 321	- 0.60
1987	54 343	53 106	53 106	204	0.38	43 846	54 712	396	0.68
1988	56 406	54 343	54 343	- 718	- 1.27	46 020	56 294	- 114	- 0.20
1989	60 891	56 406	56 408	- 3 294	- 5.41	48 708	58 250	- 2 641	- 4.34
1990	62 870	60 891	60 891	- 1 131	- 1.80	51 856	60 541	- 2 329	- 3.70
1991	62 603	62 870	62 870	964	1.54	53 106	61 451	- 1 152	- 1.84
1992	64 108	62 603	62 603	- 788	- 1.23	54 343	62 351	- 1 757	- 2.74
1993	65 336	64 108	64 108	- 625	- 0.96	56 406	63 854	- 1 482	- 2.27
1994	65 943	65 336	65 336	- 98	- 0.15	60 891	67 117	1 174	1.78
1995	67 357	65 943	65 943	- 951	- 1.41	62 870	68 557	1 200	1.78
1996	72 058	67 357	67 357	- 4 346	- 6.03	62 603	68 363	- 3 695	- 5.13
1997	70 264	72 058	72 058	1 791	2.55	64 108	69 458	- 806	- 1.15
1998	64 081	70 264	70 264	6 316	9.86	65 336	70 352	6 271	9.79
1999		64 081	64 081			65 943	70 794		
2000						67 357	71 823		
2001						72 058	75 244		
2002						70 264	73 939		
2003						64 081	69 439		

表 5 四种预测模型标准误差的比较

	预测模型方程	相关系数	标准误差 / 万吨
一元线性回归模型	$\hat{Y} = 47\ 247 + 0.344\ 9X$	0.818 6	4 736.76
直线趋势模型	$Y_t = 37\ 682 + 1\ 657t$	0.970 4	1 992.62
自回归模型	$\hat{Y}_t = 5\ 487.55 + 0.923\ 8Y_{t-1}$	0.973 3	1 682.75
自回归模型	$\hat{Y}_t = 22\ 800.66 + 0.727$	0.927 8	1 994.07
二次指数平滑法	$8Y_{t-5}$		1 589.21

表4 二次指数平滑预测计算表

日期/年	煤炭运量/万吨	一次指数平滑	二次指数平滑	a_t	b_t	预测运量/万吨	误差/万吨	误差率/%
1978	40 243	24 516	9 522	39 510.00	1 666.00			
1979	41 319	26 196	11 189	41 203.00	1 667.44	42 870	1 551.00	3.75
1980	41 499	27 726	12 843	42 609.00	1 653.67	44 263	2 764.00	6.66
1981	41 199	29 073	14 466	43 680.00	1 623.00	45 303	4 104.00	9.96
1982	43 846	30 550	16 074	45 026.00	1 608.44	46 634	2 788.00	6.36
1983	46 020	32 079	17 676	46 518.00	1 602.33	48 120	2 100.00	4.56
1984	48 708	33 758	19 284	48 232.00	1 608.22	49 840	1 132.00	2.32
1985	51 856	35 568	20 912	50 224.00	1 628.44	51 852	- 4.00	- 0.01
1986	53 106	37 322	22 553	52 091.00	1 641.00	53 732	626.00	1.18
1987	54 343	39 024	24 200	53 848.00	1 647.11	55 495	1 152.00	2.12
1988	56 406	40 762	25 856	55 668.00	1 656.22	57 324	916.00	1.62
1989	60 891	42 775	27 548	58 002.00	1 691.89	59 694	- 1 197.00	- 1.97
1990	62 870	44 785	29 272	60 298.00	1 723.67	62 022	- 848.00	- 1.35
1991	62 603	46 567	31 002	62 132.00	1 729.44	63 861	1 258.00	2.01
1992	64 108	48 321	32 734	63 908.00	1 731.89	65 640	1 532.00	2.39
1993	65 336	50 023	34 463	65 583.00	1 728.89	67 312	1 976.00	3.02
1994	65 943	51 615	36 178	67 052.00	1 715.22	68 767	2 824.00	4.28
1995	67 357	53 189	37 879	68 499.00	1 701.11	70 200	2 843.00	4.22
1996	72 058	55 076	39 599	70 553.00	1 719.67	72 273	215.00	0.30
1997	70 264	56 595	41 299	71 891.00	1 699.56	73 591	3 327.00	4.74
1998	64 081	57 344	42 904	71 748.00	1 604.44	73 388	9 307.00	14.52
1999						73 388		
2000						74 993		

[参 考 文 献]

- [1] 黄良文¹⁹.社会经济统计学原理[M]¹⁹.北京:中国统计出版社,1996,390~427¹⁹.
- [2] 铁道部计划统计司¹⁹.全国铁路统计资料汇编[R]¹⁹.历年¹⁹.
- [3] 北京银冠电子科技有限公司¹⁹.1998年中国统计年鉴[M]¹⁹.北京:中国统计出版社,1998¹⁹.