

# 多元回归分析在中国城镇就业前景预测中的应用

**摘要:** 利用相应数据分析与统计工具, 通过比较中国城镇就业各影响因素的回归平方和增量, 得出各因素对城镇就业人数的影响程度排序, 基于多元回归分析方法和趋势外推法, 针对我国城镇就业人数建立了线性多元回归模型并加以检验, 从分产业、分地区和分就业单位的角度进行模型精化, 结合国家的有关决策和规划, 对 2009 年及 2010 年的我国就业前景进行预测, 根据影响因素重要性和社会现状, 给出模型可能的拓展及解决城镇就业问题的建议。

**关键词:** 城镇就业人数; 多元回归分析; 趋势外推法; 线性规划; 预测

**中图分类号:** 024   **文献标识码:** A

## Using Multiple Regression Analysis in the Prediction of Urban Employment in China

**Abstract:** This paper compared incremental sum of squares of regression of influential factors of urban employment in China with each other by according data analysis and statistics tools, and then concluded the degree of influence of these factors. Based on multiple regression analysis and Trend Extrapolation, linear multiple regression model was founded and checked. Furthermore, this model was refined by industries, regions and employment establishments. Considering relevant national policies and plans, prediction of urban employment in China in 2009 and 2010 was made. Finally, according to the importance of influential factors and current social condition, possible extension of the model and suggestions of solving urban employment problem were given.

**Key words:** urban employment population; multiple regression analysis; Trend Extrapolation; linear programming; prediction

### 0 引言

就业率(失业率)是国民经济中的至关重要的问题。一般来讲, 就业率和失业率反映了人口增长和经济发展之间的协调关系<sup>[1]</sup>。从统计数据可以看出, 社会经济条件、人口变动<sup>[2]</sup>以及地区等因素对我国就业人数或城镇登记失业率有着直接的影响。同时, 政府对宏观经济政策的调控以及对不同地区、不同就业人群的政策导向也极大地发挥着影响作用。自 2008 年以来, 我国的就业形势不容乐观, 中央已采取了一些相应的就业扶持政策来积极地应对这一难关, 并且提出了新一轮的就业目标。

近年的相关统计数据较好地反映了我国就业问题的一般规律, 结合国民经济运行数据, 可以对我国就业人数或城镇登记失业率的核心问题建立相应的数学模型, 并对未来就业前景加以预测。

### 1 中国城镇就业的主要影响因素的统计分析和模型建立

#### 1.1 影响因素的统计分析

在影响我国城镇就业的所有因素中, 我们通过提取政府报告及相关资料中的信息, 确定了城镇总人口、城镇总就业人口、国内生产总值、城镇居民消费总值、投资总额、进出口总额、高校毕业生人数、全国科技机构数及全国工业企业单位数作为研究的主要因素。其中, 所有数据均来自中华人民共和国国家统计局<sup>[3]</sup>以及和讯网<sup>[4]</sup>。

根据中华人民共和国国家统计局编写整理的《国家统计年鉴》, 我们得到我国城镇历年就业人数及影响就业各因素数据统计表(1993-2007), 由于其中的价值量指标均是现价计

算的，所以它们彼此之间无法进行比较，需要用城镇居民消费价格指数进行调整<sup>[5]</sup>。调整公式为：调整后的数据 = 调整前的数据 / 城市居民消费价格指数

去掉一些无用的列，调整后的数据如表 1 所示：

表 1 我国城镇历年就业人数及影响就业各因素数据统计表（1993-2007）（调整后）

年份	城市居民消费价格指数 (1985=100)	城镇净增 就业人数 $y$	城镇净增 总人口 $x_1$	进出口总额 $x_2$	国内生产总值 $x_3$	城镇居民消 费总值 $x_4$
		(万人)	(万人)	(亿元)	(亿元)	(亿元)
1994	367.8	391	996	5541.571506	13104.37738	2483.437
1995	429.6	387	1005	5470.181564	14151.23371	2779.141
1996	467.4	882	2130	5163.414634	15228.19855	3010.524
1997	481.9	859	2145	5596.015771	16387.8398	3258.639
1998	479	835	2159	5605.365344	17620.52192	3509.823
1999	472.8	796	2140	6323.22335	18967.23773	3880.129
2000	476.6	739	2158	8240.285355	20817.16324	4337.683
2001	479.9	789	2158	8790.081267	22849.59367	4749.546
2002	475.1	840	2148	10814.18649	25327.86782	5447.129
2003	479.4	859	2164	14702.44055	28331.83146	6308.15
2004	495.2	837	1907	19293.03312	32285.60178	9346.305
2005	503.1	855	1929	23240.27032	36417.69032	10333.79
2006	510.6	979	1494	27608.97963	41504.79828	11627.54
2007	533.6	1040	1673	31290.23613	46763.47451	13006.65

(续上表)

年份	外商投资金额 $x_5$	政府投资金额 $x_6$	高校毕业人数 $x_7$	科技机构数 $x_8$	工业企业单 位数 $x_9$
	(亿元)	(亿元)	(人)	(个)	(万个)
1994	791.2660223	3842.486017	637000	12499	116.5607
1995	729.3712081	3930.605514	805000	13107	73.6838
1996	742.230015	4160.112732	830000	12033	71.2423
1997	778.5255841	4397.052337	829000	11142	63.6441
1998	785.7885658	5144.518324	839000	10926	68.3145
1999	705.949192	5608.498777	848000	11237	70.2639
2000	707.2074192	6199.576047	950000	7601	69.5891
2001	808.5209544	6945.904968	1036000	7400	68.9513
2002	918.867209	8237.081012	1337000	7192	66.9834
2003	923.7815707	10667.08409	1877000	6841	70.6186
2004	1013.37315	13218.73509	2391000	9083	69.7459
2005	1178.946989	16466.37203	3068000	9352	70.086
2006	1135.271126	20407.65876	3775000	10464	70.8482
2007	1190.205555	24545.15801	4478000	11847	79.4459

## 1.2 多元回归数学模型建立

我们尝试在  $y$  与  $x_i (i=1, 2, \dots, 9)$  之间建立某种函数关系，这是数学中的拟合问题。我

们采用多元线性回归分析方法<sup>[6]</sup>建立此多元回归模型:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 + b_8x_8 + b_9x_9 + e$$

其中  $e : N(0, s^2)$ ,  $b_0, b_1, \mathbf{L}, b_9, s^2$  是未知参数。

由 Excel 中的多元线性回归分析结果, 我们可以得到线性回归方程为:

$$y = -582.525 + 0.335551x_1 - 0.07753x_2 + 0.043727x_3 + 0.023969x_4 - 0.25395x_5 - 0.11412x_6 + 0.000874x_7 + 0.006155x_8 + 3.648911x_9$$

### 1.3 利用多元回归分析进行模型检验

#### 1.3.1 回归方程的显著性检验

利用多元回归分析中的方差分析检验原假设

$$H_0 : b_0 = b_1 = \mathbf{L} = b_9 = 0$$

$$H_1 : \text{至少有一个 } b_i \neq 0 (i = 0, 1, \mathbf{L}, 9)$$

当原假设  $H_0$  成立时, 说明回归方程不显著, 采用线性回归是不合理的;

当备择假设  $H_1$  成立时, 说明回归方程显著, 采用线性回归是有意义的。

查  $F$  分布表, 有:  $F_{0.05}(9, 4) = 6.00$ ,  $F_{0.01}(9, 4) = 14.66$ 。而由表 3 可知  $F = 8.492199$ , 则  $F > F_{0.05}(9, 4)$ 。依据规则, 拒绝原假设  $H_0$ , 即可以认为  $y$  与  $x_1, x_2, \mathbf{L}, x_9$  的线性相关关系显著。

#### 1.3.2 影响因素主次的判别

在变量  $x_1, x_2, \mathbf{L}, x_9$  中, 各个变量对因变量  $y$  的影响程度各不相同, 因此我们用 Excel 对它们做因素主次判别。变量  $x_i$  对应的回归系数  $b_i$  的越小, 则该变量对  $y$  的影响越小。

在表 1 中分别去掉  $x_1, x_2, \mathbf{L}, x_9$  后, 在对改动后的表在 Excel 中进行回归分析, 分别得到新的回归平方和  $S_{R_i}$  ( $i = 1, 2, \mathbf{L}, 9$ ), 记  $\Delta s_i = S_R - S_{R_i}$  ( $i = 1, 2, \mathbf{L}, 9$ ),  $\Delta s_i$  越小, 说明变量  $x_i$  对因变量  $y$  的作用越小。以下是我们按上述方法得到的数据汇总表:

表 2 各因素重要性主次的判别表

$S_{R_i}$	428910.7		
$S_{R1}$	403677.1	$\Delta s_1$	25233.62
$S_{R2}$	427908.8	$\Delta s_2$	1001.879
$S_{R3}$	427901.1	$\Delta s_3$	1009.581
$S_{R4}$	428830.7	$\Delta s_4$	79.96397
$S_{R5}$	428166.2	$\Delta s_5$	744.4744
$S_{R6}$	427698.4	$\Delta s_6$	1212.259
$S_{R7}$	426842.8	$\Delta s_7$	2067.93
$S_{R8}$	428869.5	$\Delta s_8$	41.22939

$S_{R9}$	427824	$\Delta s_9$	1086.739
----------	--------	--------------	----------

从表 2 我们可以清楚地看出对城镇就业人数  $y$  影响最大的是城镇净增总人口  $x_1$ ，其次为高校毕业生人数  $x_7$ 、政府投资金额  $x_6$ 、工业企业单位数  $x_9$ 、国内生产总值  $x_3$ 、进出口总额  $x_2$ 、外商投资金额  $x_5$ 、城镇居民消费总值  $x_4$ 、科技机构数  $x_8$ 。

### 1.3.3 回归系数的显著性检验

$$\begin{aligned} \text{检验假设} \quad & H_0: b_i = 0 \\ & H_1: b_i \neq 0 \quad (i=1,2,\mathbf{L},9) \end{aligned}$$

当原假设  $H_0$  成立时，说明自变量  $x_i$  对因变量  $y$  不起作用，在回归模型中可以去掉；  
当备择假设  $H_1$  成立时，说明自变量  $x_i$  对因变量  $y$  起作用，在回归模型中不可以去掉。

### 1.3.4 复相关系数

我们采用复相关关系： $R^2 = S_R / S_T = 1 - S_e / S_T$  来衡量因变量  $y$  与这组自变量  $x_1, x_2, \mathbf{L}, x_9$  之间的线性相关程度： $R^2$  越接近 1，因变量  $y$  与自变量组  $x_1, x_2, \mathbf{L}, x_9$  之间的线性相关程度越强。其中  $S_e$  为残差平方和， $S_R$  为回归平方和， $S_T$  为总离差平方和。

但是当采用的自变量增多时，其  $S_e$  就会减少，那么  $R^2$  增大，其中有些自变量的引入是多余的，所以调整的复相关系数： $R_a^2 = 1 - [S_e / (n - m - 1)] / [S_R / (n - 1)]$ 。 $R^2$  与  $R_a^2$  越接近 1，因变量  $y$  与自变量组  $x_1, x_2, \mathbf{L}, x_9$  之间的线性关系程度越强。

由表 3 可见， $R^2 = 0.950267$ ， $R_a^2 = 0.838368$ ，比较接近 1，因此，因变量  $y$  与自变量组  $x_1, x_2, \mathbf{L}, x_9$  之间的线性关系程度比较强。

## 1.4 模型改进

我们采用非线性回归方法，由 Matlab 画出的  $x_i (i=1,2,\mathbf{L},9)$  与  $y$  的散点图。有些散点图呈现的形状更接近于曲线。在一般常见的几种可划为直线方程的曲线方程<sup>[7]</sup>中，这里我们选择的是  $y = a + b \ln x$ ，然后利用前面的线性回归分析的方法来确定参数。

假设此多元非线性回归方程为：

$$\begin{aligned} y = & g_0 + g_1 x_1 + g_2 \ln x_2 + g_3 \ln x_3 + g_4 \ln x_4 + g_5 \ln x_5 \\ & + g_6 \ln x_6 + g_7 \ln x_7 + g_8 \ln x_8 + g_9 x_9 + e \end{aligned}$$

其中  $e: N(0, s^2)$ ， $g_0, g_1, \mathbf{L}, g_9, s^2$  是未知参数

令  $u_i = \ln x_i \quad i=2,3,\mathbf{L},8$ ，带入上式，得

$$\begin{aligned} y = & g_0 + g_1 x_1 + g_2 u_2 + g_3 u_3 + g_4 u_4 + g_5 u_5 \\ & + g_6 u_6 + g_7 u_7 + g_8 u_8 + g_9 x_9 + e \end{aligned}$$

此时问题转换为求线性回归方程问题，根据前面已介绍的方法，在 Excel 中用回归分析得到此多元非线性方程为：

$$\begin{aligned} y = & -16215.7 + 0.277867x_1 - 786.304 \ln x_2 + 1991.755 \ln x_3 - 13.2031 \ln x_4 \\ & - 96.3326 \ln x_5 - 1138.23 \ln x_6 + 921.655 \ln x_7 + 125.9858 \ln x_8 + 7.238525x_9 \end{aligned}$$

此非线性回归模型的复相关系数和调整的复相关系数分别为  $R_{new}^2 = 0.957238$ ， $R_{a\ new}^2 = 0.861024$ ，比原线性回归模型  $R^2 = 0.950267$ ， $R_a^2 = 0.838368$  比较更接近 1，因此此非线性回归模型比原线性回归模型有所改进。

## 2 分产业、分地区和分就业单位的精确建模

### 2.1 基于第三产业城镇就业人数的精确模型

根据当今经济产业结构的调整：劳动力逐渐从第一产业向第二、第三产业转移。以及考

考虑到我国国情：第三产业城镇就业人数由 1952 年的 1881 万人增至 2007 年的 24917 万人，由此可以看出：随着经济的发展，第三产业劳动力将继续快速增长，这成为了城镇就业人数的重要来源。因此，我们这里只就第三产业城镇就业人数建立精确模型。

我们首先采用多元线性回归分析方法来观察因变量  $z$  与自变量组  $x_1, x_2, \mathbf{L}, x_9$  之间的线性相关程度如何，来决定是否需要选用上述 4.4 小节中改进后的非线性回归方程。设此多元回归模型：

$$z = h_0 + h_1 x_1 + h_2 x_2 + h_3 x_3 + h_4 x_4 + h_5 x_5 + h_6 x_6 + h_7 x_7 + h_8 x_8 + h_9 x_9 + e$$

其中  $e : N(0, s^2)$ ， $h_0, h_1, \mathbf{L}, h_9, s^2$  是未知参数。

由 Excel 中的多元线性回归分析结果，我们可以得到第三产业城镇就业人数精确模型的回归方程为：

$$z = -1957.8 - 1.22189x_1 - 0.88647x_2 + 0.540994x_3 + 0.899631x_4 - 3.66594x_5 - 1.97197x_6 + 0.010222x_7 - 0.21059x_8 + 57.86x_9$$

由分析结果可知  $F = 18.64729$ ，则  $F > F_{0.01}(9, 4) = 14.66$ ，且由  $R^2 = 0.976721$ ， $R_a^2 = 0.924342$ ，都很接近 1，因此，因变量  $z$  与自变量组  $x_1, x_2, \mathbf{L}, x_9$  之间的线性关系程度很强，说明  $z$  与  $x_1, x_2, \mathbf{L}, x_9$  的线性相关关系特别显著。因此，用 1.2 小节中求得的多元线性回归方程就能很好地表示第三产业净增就业人数  $z$  与自变量  $x_1, x_2, \mathbf{L}, x_9$  的函数关系，无需采用 1.4 小节中改进后的多元非线性回归方程。

## 2.2 基于华东地区就业人数的精确模型

华东地区作为我国沿海经济重点发展地区，在我国经济增长与就业人数中，该地区贡献最为显著，故选取华东地区作为我们建立精确数学模型的对象之一。

这里的华东地区包括上海、江苏、浙江、安徽、福建、江西和山东。我们采用了上述的多元线性回归分析方程和改进后的多元非线性回归方程，发现改进后的非线性回归方程模型更优。

由 Excel 中的多元线性回归分析结果，我们可以得到回归方程为：

$$v = -34706.3 + 0.505274 \ln x_1 + 4173.11 \ln x_2 + 4392.61 \ln x_3 - 2042.99 \ln x_4 - 16.3241 \ln x_5 - 4582.35 \ln x_6 - 230.407 \ln x_7 + 1433.538 \ln x_8 + 0.854173 x_9$$

由分析结果可知  $F = 104.4417$ ，则  $F > F_{0.05}(9, 4) = 6.00$ ，且由  $R^2 = 0.946157$ ， $R_a^2 = 0.825012$ ，比较接近 1，另外，我们同样可算出用多元线性回归模型求出的复相关系数  $R^2$  和调整的复相关系数  $R_a^2$  分别为  $R^2 = 0.875101$ ， $R_a^2 = 0.584077$ ，且  $F = 3.113979 < F_{0.05}(9, 4) = 6.00$ 。说明因变量华东地区城镇净增就业人数  $v$  与自变量组  $x_1, x_2, \mathbf{L}, x_9$  之间的线性关系程度很弱。但改进为多元非线性回归模型后  $v$  与  $\ln x_1, \ln x_2, \mathbf{L}, \ln x_9$  的线性相关关系显著，回归方程较优。

## 2.3 基于外商投资单位就业人数的精确模型

随着对外开放力度的逐年加大，外商投资单位不断增多，外商投资单位中的城镇就业人员在我国城镇就业人员中的比重稳步上升。因此这里暂只考虑此类单位作为精确建模用例。

我们采用了 4.2 小节中的多元线性回归分析方程和 4.4 小节中的改进后的多元非线性回归方程，发现改进后的非线性回归方程模型更优。

由 Excel 中的多元线性回归分析结果，我们可以得到回归方程为：

$$w = -92.1501 + 0.029848x_1 + 337.0317 \ln x_2 - 235.589 \ln x_3 - 4.32808 \ln x_4 + 56.95156 \ln x_5 - 97.6395 \ln x_6 - 57.4289 \ln x_7 + 77.93857 \ln x_8 - 0.49423x_9$$

由分析结果可知  $F = 25.96657$ ，则  $F > F_{0.01}(9, 4) = 14.66$ ， $R^2 = 0.983172$ ， $R_a^2 = 0.945309$ ，非常接近 1。另外，我们同样可算出用多元线性回归模型求出的复相关系数  $R^2$  和调整的复相关系数  $R_a^2$  分别为  $R^2 = 0.977874$ ， $R_a^2 = 0.92809$ 。经比较，显而易见改进后的多元非线性回归模型更优。

### 3 2009 及 2010 年我国就业前景预测

本节我们将利用以上建立的数学模型预测得出 2009 及 2010 年我国城镇就业人数，并通过与业界广泛运用的趋势外推法求出的预测就业人数作对比，同时利用线性规划预测了 1994—2010 年期间几个时间段的城镇就业人数范围，来检验我们模型的合理性与正确性。

#### 3.1 基于多元非线性回归模型的预测

由 4.4 小节中的分析可知，对于求解城镇净增就业人数  $y$  来说，改进后的非线性模型将比改进前的线性模型更接近现实值，因此我们选用前者作为求解  $y$  的最终模型。

根据国家有关政策文件及相关预测，我们得到以下预测数据：

表 3 我国城镇历年就业人数及影响就业各因素数据预测表 (2008-2010)

年份	城市居民消费价格指数 (1985=100)	城镇净增就业人数 $y$	城镇净增总人口 $x_1$	进出口总额 $ox_2$	国内生产总值 $ox_3$	城镇居民消费总值 $ox_4$
		(万人)	(万人)	(亿元)	(亿元)	(亿元)
2008	540.6	1100	1481	204859.8	287372.1	62175.05
2009	545.5	940	1335	229762.7	299443	58842.7
2010	552.1	990	1200	243545.9	326650.5	82271.51

(续上表)

年份	外商投资金额 $ox_5$	政府投资 $ox_6$	高校毕业生人数 $x_7$	科技机构数 $x_8$	工业企业单位数 $x_9$
	(亿元)	(亿元)	(人)	(个)	(万个)
2008	6863.431	174947.2	5990000	10563	80.3792
2009	7170.336	217069.5	7100000	11072	80.7269
2010	7473.772	223878.8	7000000	14205	82.5482

用城镇居民消费价格指数对  $ox_2$ 、 $ox_3$ 、 $ox_4$ 、 $ox_5$ 、 $ox_6$  进行调整后的数据为：

表 4 我国城镇历年就业人数及影响就业各因素数据预测表 (2008-2010) (调整后)

年份	进出口总额 $x_2$	国内生产总值 $x_3$	城镇居民消费总值 $x_4$	外商投资金额 $x_5$	政府投资 $x_6$
	(亿元)	(亿元)	(亿元)	(亿元)	(亿元)
2008	37894.89	53158	11501.12	1269.595	32361.67
2009	42119.66	54893.3	10786.93	1314.452	39792.76
2010	44112.64	59165.1	14901.56	1353.699	40550.41

将 2009 及 2010 年的相关数据带入 1.4 小节中的改进模型计算，得 2008 年城镇净增就业人数  $y = 1078.47$  万人；2009 年城镇净增就业人数万人  $y = 946.119$  万人；2010 年城镇净增就业人数  $y = 1024.452$  万人。

那么 2008 年城镇就业人数为  $y_c = 29350 + 1078.47 = 30428.47$  万人；2009 年城镇就业人数为  $y_c = 30428.47 + 946.119 = 31374.589$  万人；2010 年城镇就业人数为  $y_c = 31374.589 + 1024.452 = 32399.041$  万人，即为最终预算结果。

### 3.2 基于趋势外推法的检验及预测

趋势外推法 (Trend extrapolation) 是根据过去和现在的发展趋势推断未来的一类方法的总称<sup>[8]</sup>。它的基本假设是未来即过去和现在连续发展的结果。事物发展过程一般都是渐进式的变化，而不是跳跃式的。依据这种规律推导，就可以预测出它的未来趋势和状态。目前，趋势外推法在技术预测和市场预测中得到了广泛的应用。

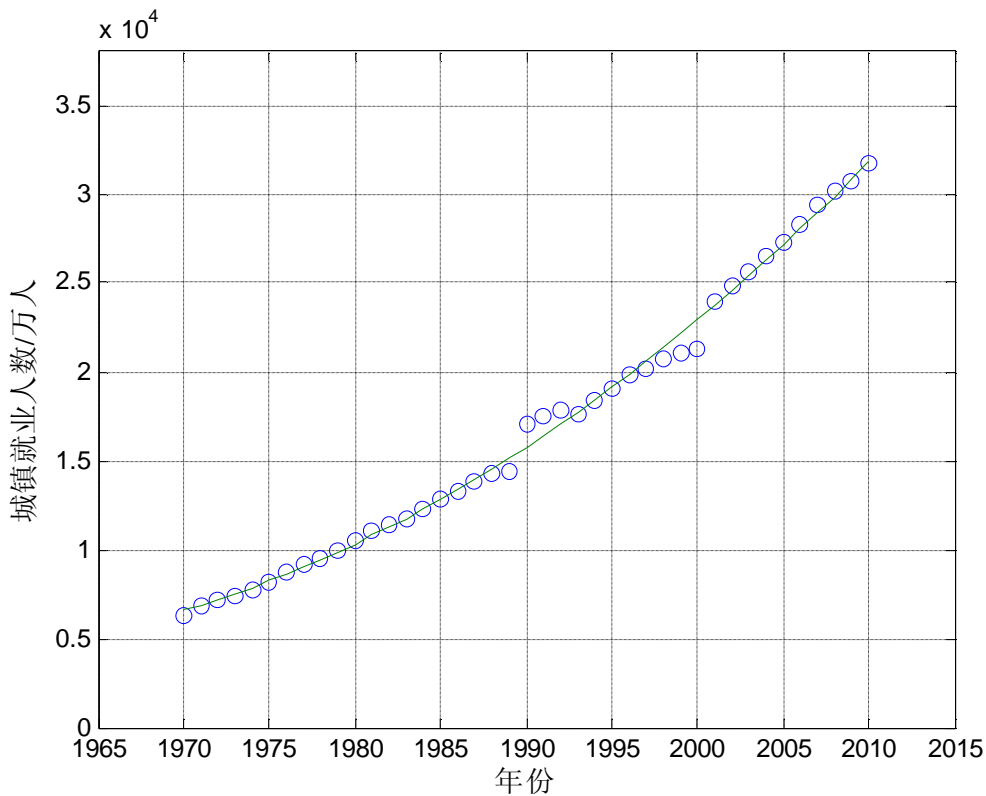
根据趋势外推法的求解步骤，可列出下表来分析预测 2009 及 2010 年的城镇就业人数：

表 5 基于趋势外推法的我国城镇历年就业人数数据统计表 (1970-2008)

年份	时序 $t$	城镇就业人数 (万人) $y_c$	$\ln y_c$	年份	时序 $t$	城镇就业人数 (万人) $y_c$	$\ln y_c$
1970	1	6312	8.750207863	1990	21	17041	9.743377
1971	2	6868	8.834628222	1991	22	17465	9.767954
1972	3	7200	8.881836305	1992	23	17861	9.790375
1973	4	7388	8.907612341	1993	24	17589	9.775029
1974	5	7687	8.947285869	1994	25	18413	9.820812
1975	6	8222	9.014568767	1995	26	19093	9.857077
1976	7	8692	9.070158341	1996	27	19815	9.894195
1977	8	9127	9.118992333	1997	28	20207	9.913784
1978	9	9514	9.160519677	1998	29	20678	9.936826
1979	10	9999	9.210240367	1999	30	21014	9.952944
1980	11	10525	9.261508659	2000	31	21274	9.965241
1981	12	11053	9.310457163	2001	32	23940	10.08331
1982	13	11428	9.343821763	2002	33	24780	10.11779
1983	14	11746	9.371268036	2003	34	25639	10.15187
1984	15	12229	9.411565459	2004	35	26476	10.18399
1985	16	12808	9.457825255	2005	36	27331	10.21578
1986	17	13292	9.494917629	2006	37	28310	10.25097
1987	18	13783	9.531191228	2007	38	29350	10.28705
1988	19	14267	9.565704457	2008	39	30210	10.31593
1989	20	14390	9.5742888				

对数据画如下折线图来分析，以城镇就业人数为  $y$  轴，年份为  $x$  轴：

图 1 基于趋势外推法的我国城镇历年就业人数曲线拟合图 (1970-2008)



从上图可以看出大致的曲线增长模式,较符合的模型有二次曲线和指数曲线模型,但无法确定哪一个模型能更好地拟合该曲线,则我们将分别对这两种模型进行参数拟合。

对于  $y_c = b_0 + b_1t + b_3t^2$  与  $y_c = ae^{bt}$  这两种模型,按趋势外推法在 Excel 中计算如下:

(1)  $y_c = b_0 + b_1t + b_3t^2$  模型

计算结果为:

$$b_0 = 6234.06, b_1 = 267.9553, b_2 = 8.63547$$

$$\text{则 } y_c = 6234.06 + 267.9553t + 8.63547t^2.$$

且有  $R^2 = 0.997492$ ,  $R_a^2 = 0.994991$  都很接近 1,因此该二次曲线能较为真实地反映城镇就业人数。

(2)  $y_c = ae^{bt}$  模型

先将  $y_c = ae^{bt}$  划为线性的:令  $u = \ln y_c$ ,  $r = \ln a$ , 则原式化为  $u = r + bt$ , 计算结果为:

$$r = 8.792063, b = 0.040195, \text{则 } a = e^r = 6581.8, y_c = 6581.8e^{0.040195t}$$

且有  $R^2 = 0.992735$ ,  $R_a^2 = 0.992539$  都很接近 1,因此该指数曲线也能较为真实地反映城镇就业人数。

以上两种模型相互比较,  $y_c = b_0 + b_1t + b_3t^2$  模型所对应的复相关系数  $R^2$  和调整的复相关系数  $R_a^2$  更接近于 1,则选择该模型得出的预测结果更优。因此采用  $y_c = 6234.06 + 267.9553t + 8.63547t^2$  模型来预测 2009 及 2010 年的城镇就业人数,对应时序分别为  $t = 40$  和  $t = 41$ ,带入模型计算可得:2009 年城镇就业人数为 30769 万人,2010 年城镇就业人数为 31736 万人。

下面将我们在 3.1 和 3.2 小节中分别采用多元非线性回归分析方法和趋势外推法所得出的 2009 及 2010 年的城镇就业人数预测结果作一个比较:

表 6 多元非线性回归法和趋势外推法的预测结果对比表(单位:万人)

年份	非线性回	趋势外推	误差率
----	------	------	-----



	归分析	法	
2009	31374.59	30769	0.019302
2010	32399.04	31736	0.020465

由上表可以看出用我们设计的模型得出的预测人数与当前较为成熟的趋势外推法得出结果的相对误差率在 2% 左右，反映我们模型的预测精度较高，因此我们以上所建立的多元非线性回归模型用于城镇就业人数预测是可行的。

### 3.3 基于线性规划法的检验及预测

在我们利用多元回归分析法建立的最初的模型中，因变量  $y$  与自变量组  $x_1, x_2, \dots, x_9$  之间具有线性关系；此外，自变量组  $x_1, x_2, \dots, x_9$  历年数据均已知，故可以确定每个变量在 1994—2007 年的范围。据此，我们可建立线性规划模型并利用 Matlab 计算出  $\min y'$  和  $\max y'$ （其中  $y' = y - 582,525$ ，单位万人）。

建立的目标函数如下：

$$\min y' = 0.335551x_1 - 0.07753x_2 + 0.043727x_3 + 0.023969x_4 - 0.25395x_5 - 0.11412x_6 + 0.000874x_7 + 0.006155x_8 + 3.648911x_9$$

约束条件：

$$\begin{cases} 996 \leq x_1 \leq 2164 \\ 5163.414634 \leq x_2 \leq 31290.23613 \\ 13104.37738 \leq x_3 \leq 46763.47451 \\ 2483.437 \leq x_4 \leq 13006.65 \\ 705.949192 \leq x_5 \leq 1190.205555 \\ 3842.486017 \leq x_6 \leq 24545.15801 \\ 637000 \leq x_7 \leq 4478000 \\ 6841 \leq x_8 \leq 13107 \\ 63.6441 \leq x_9 \leq 116.5607 \end{cases}$$

在 Matlab 中编写源代码（参见附录 4）计算可得  $\min y' = 779.7252$ ，进而  $\min y = 197.2002$ 。同理可得  $\max y = 3006.975$ ，于是 1994—2007 年中国城镇净增就业人口的范围为 (197.2002, 3006.975)。

对照最初统计数据，中国城镇净增就业人口历年真实数据均在该范围之内，说明模型具有可靠性。

考虑到 1994—2007 年年限较长，其间数据变化可谓纷繁复杂。观察实际统计数据发现，各个因素在 2000—2007 年间呈现较为平缓的发展态势，故依照上述方法得到 2000—2007 年中国城镇净增就业人口的范围 (523.875, 2713.275)，再次对照最初统计数据，历年真实数据仍在所求的范围之内，进一步说明模型具有可靠性。据此，我们预测了以下几个时间阶段的范围：

表 7 2000 至 2010 年各时间阶段中国城镇净增就业人口范围预测表

年份	$y$	$\min y$	$\max y$
2000—2007		523.875	2713.275
2000—2008		523.875	3028.375
2000—2009		523.875	3087.675

2000—2010	523.875	3368.175
-----------	---------	----------

显然根据非线性回归分析的方法算出的 2009 年城镇净增就业人数万人  $y = 946.119$  万人，和 2010 年城镇净增就业人数  $y = 1024.452$  万人均在此范围内，结果相互符合。

但是，以上各阶段的范围跨度太大，我们继续计算了如下阶段：

表 8 2007 至 2010 年各时间阶段中国城镇净增就业人口范围预测表

年份 \ y	min y	max y
2007—2010	724.275	978.175
2008—2010	1043.17	1665.175

此时，根据非线性回归分析的方法算出的 2009 年城镇净增就业人数万人  $y = 946.119$  处在 2007—2010 年的预测范围之间，2010 年城镇净增就业人数  $y = 1024.452$  与两个预测范围均有偏差，但相差不是很大。此外，由于我们 2009—2010 年的数据均为各网站公布的预测值，所以这样的偏差也是在所难免的。总之，从以上分析和比较可见，我们的模型是有效可行的。

#### 4 问题拓展及解决建议

##### 4.1 城镇登记失业率模型

城镇就业问题和失业问题是同步存在的，故而将本文中的因变量由城镇就业人数换为城镇登记失业率，即可建立起相应模型，从而预测未来城镇登记失业率。

##### 4.2 建立加入干扰项的多元回归模型

2008 年是极不寻常的一年。从特大自然灾害到国际金融危机，中国经受了巨大的历史考验，而其中的城镇人口的就业问题也不可避免地受到了影响。在这样的社会背景之下，我们的城镇就业人数预测模型应该加入金融危机等重要干扰项，应该能使预测值更接近实际。

##### 4.3 分行业、分地区、分就业人群就业容量预测

就理论上来说，预测全国范围内的就业趋势一般从两个方面展开，一是根据国民经济发展趋势及经济结构变化，对就业需求进行预测，二是根据人口发展过程中劳动力数量与结构变化，结合就业结构对劳动力的需求进行预测，然后通过比较二者的预测结果，综合评定，做出未来就业容量的预测<sup>[9]</sup>。本文主要做了第一个方面的工作，也对产业、行业、就业人群中的一类选择了较为精确的模型，但关于第二个方面的工作没有做到位。相关文献就服务业就业容量进行了预测且结果良好<sup>[10]</sup>。如果时间允许，我们可以仿效其做法预测各产业、各地区的就业容量，进而与本文中精确模型所作的结果进行比较，从而得出更为准确的预测。

##### 4.4 中国城镇就业问题解决建议

就业率（失业率）是反映一个社会是否健康有效运转的重要指标，政府应加大力度，采取积极可行的政策调控，对提高我国城镇就业人数或减少城镇登记失业率做出重要举措。根据我们得出的数学模型和预测结果，现提出以下咨询和建议，用以缓解当前的就业压力：

###### 4.4.1 明确界定农村人口或地区就业

中国是个农业大国，农村人口可谓中国人口构成中的半边天，而我们的失业率规定为城镇登记失业率，根本没有对农村人口就业与否给一个较为明确的定义，更没有就其对整体国民就业状况的影响做出过相关讨论，这是有一定缺陷的。我们在建立模型的过程中不时会受到相关问题的困扰，我们相信在明确界定农村人口或地区就业的基础上进行模型预测并制定相关政策法规，中国整体的就业状况将会有很好的前景。

###### 4.4.2 稳定政府投资，扶持青年就业

通过多元回归分析, 我们清楚地看出对城镇就业人数影响最大的首先是城镇净增总人口, 其次即为高校毕业人数和政府投资金额。把握城镇净增人口的难度应该高于稳定政府投资和扶持高校毕业生就业, 因此稳定政府投资和加大力度扶持青年就业成为解决城镇人口就业问题的重要途径, 正如财政部经济建设司司长王保安表示: 积极财政政策第一个政策导向还是要稳定政府投资。

#### 4.4.3 充分利用第三产业的强拉动力

通过 2.1 小节中的分析, 明确了第三产业和城镇就业具有较强的相互关系, 即说明了第三产业对就业具有较强的绩效作用<sup>[11]</sup>。因此, 积极扶持并有效规划第三产业使其健康、有效发展必将创造更为广阔的就业前景。

#### 参考文献

- [1] 郝晓辉. 中国可持续发展指标体系探讨[J]. 科技导报, 1998, (11).
- [2] 刘楠. 2003—2010 年我国人口变动对就业的影响及应对措施[J]. 河南社会科学, 2006, (05).
- [3] 中华人民共和国国家统计局. 年度统计数据. <http://www.stats.gov.cn/tjsj/ndsj/>, 2009.
- [4] 和讯网. 宏观数据. <http://mac.hexun.com/>, 2009.
- [5] 王馥. 中国城镇就业影响因素的计量分析 [EB/OL]. <http://public.tjufe.edu.cn/Office/jiaowuchu/netcourse/jiliang/kcjx/sjhj/3.htm>, 2004.
- [6] 胡细宝, 王丽霞. 概率论与数理统计[M]. 北京: 北京邮电大学出版社, 2004.
- [7] 钱振伟. 消费对就业增长影响比较研究——以边疆少数民族地区与东部发达地区比较为例[J]. 云南财贸学院学报, 2005, (01).
- [8] 李栓, 刘莉, 刘阳. 趋势外推法在电力负荷预测中的应用[J]. 沈阳工程学院学报(自然科学版), 2005, (Z1).
- [9] 刘社建. 改革开放以来的就业结构演变及其发展趋势[A]. 第七届全国经济学、管理学博士后学术大会, 2003, (07).
- [10] 马利彪, 高文博. 大力发展吉林省服务业以促进就业[J]. 经济纵横, 2008, (03).
- [11] 张车伟. 中国的劳动就业形势、挑战与前景[A]. 2002 年中国青年农业经济学者年会论文集, 2002.

## 中国城镇就业的影响因素统计分析及其前景预测

谭力<sup>1</sup>, 谢瑾霞<sup>2</sup>, 阿磊<sup>2</sup>

(1. 华东师范大学 计算机科学技术系, 上海 200241; 2. 华东师范大学 数学系, 上海 200241)

## Statistical Analysis of Influential Factors and Prediction of Urban Employment in China

Li TAN<sup>1</sup>, Jinxia XIE<sup>2</sup>, Lei A<sup>2</sup>

(1. Department of Computer Science and Technology, East China Normal University, Shanghai 200241; 2. Department of Mathematics, East China Normal University, Shanghai 200241)

第一作者: 谭力 (1986-), 男, 硕士研究生. E-mail: darkwhite29@gmail.com

通讯作者：谭力（1986-），男，硕士研究生. E-mail: darkwhite29@gmail.com

通讯地址：上海市闵行区虹梅南路 5800 号研究生公寓 2 舍 509 室

电话/手机：13761641588

**2009 年 10 月 7 日**