

基于拉开档次法组合模型在物流需求预测中的应用

——以江苏省为例

李磊, 易琼坚, 单丹

(江南大学商学院, 江苏无锡, 214122)

[摘要] 为了对物流需求进行科学预测, 文章在主成分分析的基础上, 选取了对物流需求影响较大的3个主要因素: 人均社会商品消费额、人均GDP和商品零售价格指数为自变量, 分别建立了多因素灰色预测模型、ARIMA预测模型、多元回归预测模型, 并运用拉开档次法建立了组合模型。通过该组合模型对江苏省物流需求进行了预测。结果表明, 江苏省物流需求每年以近9.5%快速增长。与其他预测模型相比, 该预测模型具有更高的预测精度。

[关键词] 拉开档次法; 物流需求; 多因素灰色模型; 预测

[中图分类号] F224

[文献标识码] A

[文章编号] 1671-6973(2015)01-0087-05

一、引言

当今社会快速发展, 人们生活更加快捷, 物流活动逐渐渗透到生产、流通、消费等整个社会经济活动过程之中, 与社会经济的发展息息相关, 物流对经济发展起着基础性的作用, 影响和制约经济的运行效率, 是社会经济活动的不可缺失的组成部分; 而且根据国家统计局最新统计数据显示, 我国物流总费用占GDP的比率维持在18%左右。以江苏省为例, 江苏省交通便利, 公路、铁路、水运均发达, 为经济发展提供了重要的条件。随着江苏省经济的不断发展, 对物流的依赖性也越加明显, 现有的交通条件以及基础设施等能否支撑高速发展的经济, 成为快速发展不得不面临的问题。为寻求物流供应与需求之间的平衡, 物流需求预测则为不可或缺的环节。物流需求预测是进行物流系统规划与设计的依据, 对物流需求进行有效的预测, 能够为政府提供经济活动对物流的需求信息, 从而制定正确的物流行业发展政策, 科学规划物流系统, 合

理配置物流资源, 有效协调物流需求与物流供给的关系, 促进物流对经济的有效发展。^[1]

物流预测始于20世纪90年代, 随后许多国内外学者对区域物流需求预测做了大量研究。Michaelw Babcock等(1999)介绍了三种预测模型: 一元线性回归模型、多元线性回归模型以及时间序列分析模型, 并且利用时间序列分析模型对铁路货物运输量进行了预测;^[2] Mudit Kulshreshtha, AarnaliNag等(2001)根据印度铁路1960年至1995年的时间序列数据, 利用多元正交向量自回归模型推导出货物需求量之间的关系, 通过VAR模型利用数据进行预测, 指出铁路运输物流与经济的关联度。^[3] Real Carbonneau等(2008)运用神经网络、递归神经网络和支持向量机的方法, 对物流需求量进行了预测, 并且与趋势预测法、移动平均法和线性回归方法进行了比较, 认为递归神经网络方法和支持向量机方法预测效果较好。^[4] 方威、肖衡等(2009)利用线性回归对物流需求进行预测, 并且按时间分段, 进行逐步回归后得出物流需求对GDP的贡献逐渐增大。^[5] 后锐, 张毕西(2005)构建了基

[收稿日期] 2014-10-10

[基金项目] 浙江省高校人文社科重点研究基地支撑子项目(RWSKZD04-2012ZB2); 研究生创新项目(SJLX_0540)。

[作者简介] 李磊(1959-), 男, 黑龙江哈尔滨人, 教授, 博士。研究方向: 预测理论及方法。

易琼坚(1988-), 男, 江西宜春人, 硕士研究生。

于 MLP 神经网络的物流需求预测模型,为物流需求预测提供了新方法,揭示了经济与物流需求之间的非线性映射关系。^[6]杨树果、王新利(2010)和陈森、周峰(2006)运用灰色系统理论对物流需求进行了预测。^[7-8]朱帮助(2008)分别采用一元线性回归模型、灰色 GM(1,1)模型和组合预测模型对物流需求进行预测,进行对比之后认为组合预测模型精确度较高。^[9]物流需求预测研究不断完善,精度不断提高,对后来者的研究有一定参考价值。但是美国也存在小小的缺陷,大多数学者运用的是单一模型或者单因素进行预测对比,且在自变量因素选取时并未进行详细分析,有的虽用组合模型预测但以主观赋权的较多,对预测结果精度产生一定影响。

本文将在参考以上学者研究的基础上,首先根据本人在江苏省物流需求影响因素分析及预测一文的研究成果,利用主成分分析的方法提取了社会人均商品消费额、人均 GDP 和商品零售价格指数作为自变量,以货运量为因变量,并作为物流需求预测指标,选取江苏省相关数据。其次运用多因素灰色模型等多种创新方法进行单一模型预测,然后运用拉开档次法客观赋权,克服主观因素的影响,对各模型预测结果对比分析,最后运用拉开档次法组合模型对物流需求进行预测,有效降低非平稳变量之间虚假因果、非平稳数据随机扰动等因素的影响,充分实现组合模型高拟合度和预测结果稳定的效果,准确描述物流需求,为相关部门制定行业战略决策提供参考。

二、单一预测模型的预测

1. 多因素灰色预测模型

(1) 多因素灰色预测模型建立

根据多元回归和灰色预测模型原理建立多因素灰色预测模型:

$$y_{(t)} = b_0 + b_1 x_{1(t)} + b_2 x_{2(t)} + \cdots + b_m x_{m(t)} \quad (1)$$

其中 $y_{(t)}$ 作为此事件在 t 时刻的预测值, b_i 为该模型的估计参数, $x_{i(t)}$ ($i = 1, 2, 3, \dots, m$) 则为在 t 时刻时第 i 个因素预测值。要解上述模型,首先,确定参数 $b = (b_0, b_1, \dots, b_m)$ 的值,根据历史统计数据 y_1, y_2, \dots, y_m 和 $x_{i(1)}, x_{i(2)}, \dots, x_{i(n)}$, 运用最小二乘法即可得参数值:

$$b = (b_0 b_1 \cdots b_m)^T = (X^T X)^{-1} X^T Y \quad (2)$$

其中,

$$X = \begin{pmatrix} 1 & x_{1(1)} & \cdots & x_{m(1)} \\ 1 & x_{1(2)} & \cdots & x_{m(2)} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1(n)} & \cdots & x_{m(n)} \end{pmatrix} \quad (3)$$

$$Y = (y_1 y_2 \cdots y_n)^T \quad (4)$$

将参数值代入:

$$y_{(t)} = b_0 + b_1 x_{1(t)} + b_2 x_{2(t)} + \cdots + b_m x_{m(t)} \quad (5)$$

其次,估计在 t 时刻时第 i 各因素的估计值 $x_{i(t)}$ ($i = 1, 2, 3, \dots, m$), 在此建立灰色预测模型,设此模型存在 m 个因素, $y = (y_1, y_2, y_3, \dots, y_m)$ 表示此事物特征因素时间序列,用

$$X_i^{(0)} = (x_i^{(0)}(0), x_i^{(0)}(1), \dots, x_i^{(0)}(m)) \quad (6)$$

$$(i = 1, 2, \dots, m)$$

表示单因素时间序列,做 1-AGO 可得:

$$X_i^{(1)} = (x_i^{(1)}(1), x_i^{(1)}(2), \dots, x_i^{(1)}(n)) = (x_i^{(0)}(1), x_i^{(0)}(1) + x_i^{(0)}(2), \dots, x_i^{(0)}(n-1) + x_i^{(0)}(n)) \quad (7)$$

背景数列为:

$$Z_i^{(1)} = (z_i^{(1)}(1), z_i^{(1)}(2), \dots, z_i^{(1)}(n)) \quad (8)$$

而且,

$$z_i^{(1)}(k) = \frac{1}{2}(x_i^{(1)}(k) + x_i^{(1)}(k-1)), \quad k = 2, 3, \dots, n \quad (9)$$

对 $X_i^{(0)}$, $i = 1, 2, \dots, m$ 分别进行建立 GM(1,1) 模型,得出第 i 因素在 t 时刻的估计值为 $\hat{x}_{i(t)}^{(0)} = (1 - e^a)(x_i^{(0)}(1) - \frac{b}{a})e^{-a(t-1)}$, $t = 1, 2, \dots$ 其中, a 为发展灰数, b 为内控制灰数。

$$\hat{a} = [a, t]^T = (B^T B)^{-1} B^T Y \quad (10)$$

$$Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T \quad (11)$$

$$B = \begin{pmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \cdots & -z^{(1)}(n) \\ 1 & 1 & \cdots & 1 \end{pmatrix}^T \quad (12)$$

然后根据平均相对误差和残差对此模型进行检验,可用该模型进行预测^[10]。

(2) 物流需求的多因素灰色预测

根据物流需求影响因素的主成分分析,得出物流需求的三个主要影响因素:人均社会消费品零售额(元) x_1 , 价格指数(以 1978 年的为基数) x_2 , 人均 GDP(亿元) x_3 。^[1]客观来说衡量物流需求的指标很多,本文选取货运量(y)作为物流需求的衡量指标。这是因为考虑到研究过程中选取指标的代表性,货运量在一定程度上能够较直接反映出区域物流需求量的变化规律和特点。

首先根据江苏省 1997 年—2013 年 X_1 、 X_2 、 X_3 的数据,通过建立的模型,运用灰色系统理论建模软件 3.0 对各因素值分别预测,预测结果如表(1)所示。

其次根据 y 、 x_1 、 x_2 、 x_3 的历史数据,运用 Eviews6.0 软件计算,经过方差检验,该模型在显著水平为 0.05 时是显著的。于是,得到多因素灰

色预测模型

$$y_{(t)} = 190.63x_{1(t)} - 0.89x_{2(t)} + 2.68x_{3(t)} - 15690.64 \quad (13)$$

代入各因素的预测值得出 2014 年—2018 年的物流需求预测结果。

表 1 各因素 2014—2018 年预测值

Tab. 1 Predicted values of various factors from 2014 to 2018

年份	商品价格指数 (1978 年=100)	均 GDP (元)	人均社会消费品 金额(元)
2014	465.4	84195.7	30682.28
2015	476.5	94002.9	35545.96
2016	488.0	104952.6	41180.63
2017	499.6	117177.7	47708.48
2018	511.6	130826.7	55271.12

数据来源:江苏省统计年鉴 1997—2013

2. ARIMA 模型

(1) ARIMA 模型基本原理

ARIMA 模型是将预测对象随时间推移而形成的数据序列视为一个随机序列,用一定的数学模型来近似描述这个序列,确定后就可以从时间序列的过去值及现在值对未来进行预测。^[1] ARIMA 模型主要包括 4 种基本类型:MA 模型(移动平均模型)、AR 模型(自回归模型)、ARMA 模型(自回归移动平均模型)和 ARIMA 模型(差分自回归移动平均模型)。^[11] 其中 ARIMA(p,d,q)称为差分自回归移动平均模型,p 称为自回归项,d 作为时间序列平稳时所做的差分次数,q 为移动平均项数,AR 是自回归,MA 为移动平均。^[12]

(2) ARIMA 模型的建立

首先为避免数据本身差异因素,对数据进行简单处理,运用 ADF 单位根对序列平稳性进行检验,若非平稳,进行一次差分,观察其是否进入平稳状态,若还未进入平稳状态,进行二次差分,从而确定 d 的值^[13];本文检验结果如表 1 所示,ADF 检验结果表明原数列经过两次差分后进入平稳状态,所以 d=2。

表 2 江苏省物流需求时间序列的平稳性检验

Tab. 2 Stationarity test of time series of Jiangsu province's logistics demand

序列	ADF 统计量	临界值			P 值
		1%	5%	10%	
y	1.5015	-4.1219	-3.1449	-2.7137	0.9978
dy	-1.5335	-4.2000	-3.1753	-2.7289	0.4804
dyy	-6.0265	-4.2000	-3.1753	-2.7289	0.0001

注:dy 为 y 的一阶差分,dyy 为 y 的二阶差分

通过 Eviews6.0 软件对时间序列 y 进行自相关系数和偏相关系数分析,得出两系数均具截尾性,且移动平均参数 q=1,自回归参数 p=1,经检验残差数列是白噪音序列,检验后的效果较好,所以可以运用 ARIMA(1,2,1)模型对时间序列货运量进行预测,由 1997 年至 2013 年货运量数据得出预测模型,先预测 2009 年至 2013 年五年数据,与实际数据对比,验证模型的精度,预测与实际数据误差分别为 1.7%,0.8%,2.97%,0.42%,2.16%。

表 3 ARIMA 预测对比(万吨)

Tab. 3 ARIMA predictions (tonnes)

年份	货运量	预测货运量	相对误差
2009	160966	158229.57	1.7%
2010	188565	187056.48	0.8%
2011	212594	206455.10	2.97%
2012	231295	230312.11	0.42%
2013	251691	257147.48	2.16%

数据来源:江苏省统计年鉴 1997—2013

3. 多元回归模型

本文承接前人的研究成果及根据物流需求影响因素分析一文,选取价格指数(以 1978 年的为基数) x_1 、人均社会消费品零售额(元) x_2 、人均 GDP(亿元)等三个变量作为物流需求的自变量,以货运量作为因变量 Y,组建线性回归预测模型;为了避免纲量不同和数值差异造成的误差,在进行预测前对相关自变量数据进行标准化处理;根据影响因素和自变量的相关样板数据,运用最小二乘法建立回归模型,^[13]得到回归模型如下:

$$y = 190.63x_1 - 0.89x_2 + 2.68x_3 - 15690.64$$
$$t = (2.729) \quad (-2.66) \quad (-2.728) \quad (-2.726)$$

(14)

表 4 多元回归计算结果

Tab. 4 Multiple regression results

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X3	2.679811	96.83015	-2.728907	0.0183
X2	-0.885124	105.5387	-2.664237	0.0206
X1	190.6228	49544.94	2.729243	0.0183
C	-15690.64	18723099	-2.726552	0.0184

通过 eviews6.0 软件检验可知,可决定系数和调整的可决定系数分别为 $R^2 = 0.987$ 、 $R^2 = 0.983$ 均接近 1,该模型拟合优度较高;其次进行回归平方和、残差平方和、总平方和、F 值以及对 F 值的显著性检验,显著性水平小于 0.05,都具有较好显著性

即可应用该模型进行预测,预测如图1中可以看出,模型的实际值与预测值非常接近,拟合程度很好,残差基本上也在 $-0.1 \sim 0.1$ 之间波动,再次说明了模型的拟合度较好,具有较好的显著性水平,可以用来预测江苏省的物流需求情况。

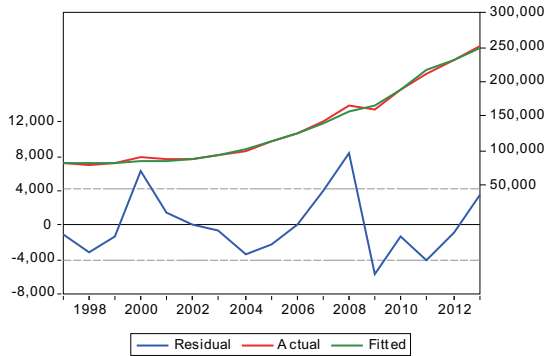


图1 预测值与真实值拟合图

Fig. 1 Fitting figure of predicted values and real values

三、组合预测模型的预测

1. 拉开档次法赋予权重

拉开档次法赋权可从几何角度来看,将 n 个被评价对象看成是由 m 个评价指标构成的 m 维评价空间中的 n 个点(或向量)。寻找 n 个被评价对象的评价值(标量)就相当于把这 n 个点向某一维空间做投影。选择指标权系数使得各被评价对象之间的差异尽量拉大,也就是根据 m 维评价空间构造一个最佳的一空间,使得各点在此一维空间上的投影点最为分散,即分散程度最大。^[14]取极大型评价指标 $x_1, x_2, x_3, \dots, x_m$ 的线性函数

$$y = w_1 x_1 + w_2 x_2 + \dots + w_m x_m = w^T x \quad (15)$$

作为被评价对象的综合评价函数。式(15)中 $W = (w_1, w_2, \dots, w_m)^T$ 是 m 维待定向量(作用相当于权系数向量), $X = (x_1, x_2, \dots, x_m)^T$ 为被评价对象的状态向量。如果将第 i 个被评价对象 s_i 的 m 个标准观测值 $x_{i1}, x_{i2}, \dots, x_{im}$ 代入式中即得

$$y_i = w_1 x_{i1} + w_2 x_{i2} + \dots + w_m x_{im} \quad i=1, 2, \dots, n \quad (16)$$

若记

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, A = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} \quad (17)$$

则式(16)可写成 $Y = AW$ 。要确定权系数向量 w ,则应能最大限度地体现出质量不同的评价对象之间的差异。例如用数学语言来描述,也就是求指标向量 x 的线性函数 $W^T X$,并使此函数对 n 个被评价对

象取值的方差或分散程度尽可能地大。

将 $Y = W^T$ 按 n 个评价对象取值构成样本的方差为

$$s = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{Y^T Y}{n} - \bar{y}^2 \quad (18)$$

将 $y = AW$ 代入式(18)中,并注意到原始数据的标准化处理,可知 $\bar{y} = 0$,于是有

$$ns^2 = W^T A^T A W = W^T H W \quad (19)$$

式中, $H = A^T A$ 是对称矩阵。

显然,对 w 不加以限制时,式子(19)中可取任意大的值,在这限定 $W^T W = 1$,并求式中的最大值,也就是选择 w 使得

$$\begin{aligned} \max \quad & W^T H W \\ \text{s. t. } & W^T W = 1 \\ & W > 0 \end{aligned} \quad (20)$$

若取 ω 为 H 的最大特征值,则所对应的标准特征向量将其归一化就可得权重系数向量;若 H 为正方形,则存在唯一正的最大特征值 λ_{\max} 及唯一与 λ_{\max} 相对应的正特征向量。^[14-15]

2. 组合模型建立与物流需求预测

通过拉开档次法计算对多因素灰色模型、ARIMA预测模型、回归预测模型赋予权数分别为0.4471、0.3372、0.2157,令多因素灰色模型预测值为 y_1 ,ARIMA预测模型预测值为 y_2 ,回归预测模型预测值 y_3 ;由此构成物流需求预测组合模型为

$$y = 0.4471 y_1 + 0.3372 y_2 + 0.2157 y_3$$

四种模型建立可以分别对与江苏省2009年—2013年的物流需求进行预测,然后与实际值比较,计算其相对误差,对比分析结果如表5所示。

表5 预测精度的比较(万吨)

Tab. 5 Comparison of prediction accuracy (tonnes)

年份	多因素灰色预测	ARIMA预测值	多元回归预测	组合模型预测
2009	159648.4	158229.6	166743.7	160700.4
2010	185070.5	187056.5	189774.5	186754.8
2011	205474.0	206455.1	216696.9	208225.6
2012	229874.3	230312.1	232073.6	230496.3
2013	257977.2	257147.5	248270.8	255603.8
误差	1.5%	1.7%	2.2%	1.01%

数据来源:江苏省统计年鉴1997—2013

结果表明拉开档次法赋权组合模型相对于其他预测模型具有更高的精度,通过此预测模型预测2009—2013年江苏省物流需求与实际比较平均误差仅为1.01%,因此,基于此模型对江苏省2014年一

2018 年五年的物流需求进行预测,结果如表 6 所示

表 6 2014—2018 年江苏省物流需求预测(万吨)

Tab. 6 Prediction of logistics demand of Jiangsu province from 2014 to 2018

年份	物流需求预测值
2014	273585.84
2015	299129.47
2016	327287.86
2017	358073.62
2018	391754.88

四、结论

①本文在江苏省物流需求影响因素分析及预测一文的基础上选取 x_1 :人均社会消费品零售额(元), x_2 :价格指数(以 1978 年的为基数), x_3 :人均 GDP(亿元)为自变量因素,最后组建的组合预测模型与其他预测模型对比可知,运用拉开档次法将多因素灰色预测、ARIMA 预测模型、多元回归预测模型客观赋权,大大的提高物流需求预测的精度。

②从预测结果可知,江苏省物流需求较大,2014 年—2018 年每年将以近 9.5%速度快速增长,到 2018 年物流需求近 40 亿吨。庞大的物流需求,要求江苏政府应做较好的规划,改善交通基础设施,大力培养物流高素质人才,经济发展与物流紧密联系,只有满足物流才能更好地服务于经济发展,服务于人民生活。

[参 考 文 献]

[1] 李磊,单丹. 江苏省物流需求影响因素分析及预测[J]. 科技与经济,2013,(6).

[2] MiChaelw Babcock, Xiaohua Lu, Jerry Norton. Time series forecasting of quarterly railroad grain carloadings [J]. *Transportation Research part E: Logistics and Transportation Review*, 1997, 35(1): 43—57.

[3] Mudit Kulshreshtha, Aarnali Nag, Mukul Kulshreshtha. A Multivariate Cointegrating Vector Auto-regressive Model of Freight Transport Demand: Evidence from Indian Railways[J]. *Transportation Research Part A*, 2001, 35(2): 29—45.

[4] Real Carbonneau, Kevin Laframboise, Rustam Vahidov. Application of machine learning techniques for supply chain demand forecasting[J]. *European Journal of operational Research*, 2008, 184(1): 1140—1154.

[5] 方威,肖衡,任湘郴. 基于线性回归模型的物流需求预测分析[J]. 生产力研究, 2009, (12).

[6] 后锐,张毕西. 基于 MLP 神经网络的区域物流需求预测方法及其应用[J]. 系统工程理论与实践, 2005, (12).

[7] 杨树果,王新利. 偏最小二乘回归与灰色模型耦合在物流需求预测中的应用[J]. 农业技术经济, 2010, (7).

[8] 陈森,周峰. 基于灰色系统理论的物流需求预测模型[J]. 统计与决策, 2006, (2).

[9] 朱帮助. 组合预测模型在区域特流需求预测中的应用[J]. 经济地理, 2008, 28(11).

[10] 王小丽. 基于多因素灰色模型的物流需求量预测[J]. 统计与决策, 2013, (14).

[11] CONEJO A J, PLAZAS M A, ESPINOLA R, et al. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models[J]. *IEEE Transactions on Power Systems*, 2005, 20(2): 1035—1042.

[12] TANG Z Y, ALMEIDA C, FISHWICK P A. Time series forecasting using neural networks vs Box-jenkins methodology[J]. *Simulation*, 1991, 57(5): 303—310.

[13] 孟凡生,李美莹. 组合模型在能源需求预测的应用[J]. 哈尔滨工业大学学报, 2013(11): 33—36.

[14] 郭亚军. 综合评价理论、方法及应用[M]. 北京: 科学出版社, 2007: 67—71.

[15] 郭亚军. 一类多层次大系统的递阶综合评价方法[J]. 管理工程学报, 1994, 8(1): 8—12.

[16] 权丽. 我国终端能源消费因素分析及实证研究[J]. 技术经济, 2011, 8(1): 83—86.

(责任编辑:程晓芝、蒋萍)