

DOI: 10.13954/j.cnki.hduss.2022.01.001

## 基于组合预测模型的快递需求预测研究

陈畴镛, 高明镜

(杭州电子科技大学 管理学院, 浙江 杭州 310018)

**摘要:** 考虑到快递需求受电子商务环境和季节性特征影响,文章将电子商务因素纳入快递需求预测指标体系中,采用灰色关联分析法确定关联度显著指标,采用GM(1,N)模型进行趋势预测,应用时间序列ARIMA模型预测季节性差异,并建立上述两类模型的GM(1,N)-ARIMA权重组合模型以提升预测精度。运用组合预测模型对杭州市2021年4季度—2022年4季度的快递需求进行预测。研究结果表明,组合预测模型的预测误差显著优于所选单一模型;电子商务的发展对快递需求的影响将会逐渐增加;快递需求的季节性差异出现新特征。

**关键词:** 快递需求; 电子商务; 灰色关联分析; ARIMA模型; 组合预测

中图分类号: F252.21

文献标志码: A

文章编号: 1001-9146(2022)01-0001-09

电子商务崛起以来,凭借其庞大的用户规模和产业关联能力,带动了快递需求的快速增长。电子商务交易以网络为载体,交易时间自由、交易量庞大且退换货较多;加之网络贸易本身存在的淡旺季以及促销活动的影响,特别是农村电商、跨境电商、社交电商等新业态的快速发展,使得快递需求的季节性差异不断增大,旺季爆仓、淡季资源过剩的不平衡问题愈发突出。对快递需求做出较为准确的预测,可为政府和企业快递资源的合理布局、提升快递服务质量、安全性和响应能力等方面提供决策依据。

### 一、相关概念与文献研究

快递需求是指在一定时间范围和行业发展水平下,快递业对快件(信函、商业文件或包裹)从揽收、包装到送达消费者手中的各个环节中起到的资源配置功能,包括时间效用、安全效用和信息查询等附加服务的实现。我国的快递需求分为现实快递需求和潜在快递需求。现实快递需求通常用“快递业务量”衡量,即一定时期内社会活动中企业和个人所需要的总的快递服务,是快递作业过程中有形物资数量的总计。近年来快递业务量的八成以上均来自于电子商务快递<sup>[1]</sup>。潜在快递需求指消费者对快递服务产生的尚未满足的确定意愿。一旦条件许可,潜在的快递需求就可以形成现实的快递活动。目前我国快递业的寄递设施和网络布局还未完善,快递服务能力较弱,而随着快递业的进步,末端配送网络规模化、专业化程度的提高,潜在的快递需求将会随之转变。

在快递需求的季节性特征方面,康阅春<sup>[2]</sup>认为该季节性表现在快递业务量在春节、五一、国庆等节假日会显著增长,而在节假日过后会回落;汤炎非和黄静<sup>[3]</sup>则认为快递业淡季较常出现在春节至3月期间,旺季则在9月至春节期间;魏邦艳<sup>[4]</sup>注意到快递业务量在春节时期的错峰效应,电商企业的节日促销使得春节的快递业务量暴涨,通常2月份后业务量开始出现回落;邓欣伟和刘丽娜<sup>[5]</sup>认为天气会影响人们的社会活动范围及顾客的购物心理,而由于人们的购物传统、商家的促销活动等因素,节假日是人们产生购物需求的高峰期;李贞贞<sup>[6]</sup>采用ARIMA季节模型对河南省快递业业务量进行预测,该模型能够较好拟合业务量变化趋势且精度较高。

收稿日期: 2021-12-07

基金项目: 国家自然科学基金-浙江省两化融合联合基金项目(U1509220);浙江省哲学社会科学规划重大课题(20NDYD53ZD)

作者简介: 陈畴镛(1955—),男,浙江绍兴人,教授,博士生导师,信息化战略、管理系统分析与决策优化。

在电子商务对快递业的影响方面,Tang 和 Deng<sup>[7]</sup>认为经济的快速增长、互联网的普及和网上购物的兴起是促使快递业务快速增长的主要原因;Han 等<sup>[8]</sup>研究得出宏观经济、电子商务发展和运输水平三方面因素影响了快递业发展;Lv 和 Weng<sup>[9]</sup>采用电子商务企业营业收入和快递组合两个关键指标,对电子商务与快递业协调程度进行协整检验;张文法<sup>[10]</sup>使用灰色关联分析法对山东省电子商务的发展对物流业的影响做出研究,结果显示两行业总体业态关联度较高,电子商务从业人数对物流业发展有较强影响;石咏梅<sup>[11]</sup>采用协整模型分析影响快递业务量的因素,研究表明网上购物和互联网发展对快递业发展起到了促进作用。

在快递需求预测模型方面,Yu 等<sup>[12]</sup>将蚁群算法引入到建模过程中,建立改进的 ACO-SVM 预测模型对青岛市的物流需求进行预测;Huang 等<sup>[13]</sup>为缓解电子商务与物流需求增长匹配度较低的问题,从电子商务的角度构建广东省物流需求预测指标体系,并采用 GM(1,1)模型和 BP 神经网络模型预测物流需求;许荣斌等<sup>[14]</sup>研究物流公司日均快递业务量的预测问题,并对粒子群优化算法进行惯性权重改进,结合反向传播神经网络模型的组合模型进行快递业务量预测;周杨等<sup>[15]</sup>基于某快递企业的日度业务量数据,使用优化后的 SARIMA 模型与滑动窗口模型共同预测该企业的快递业务量;许闯来等<sup>[16]</sup>研究不确定需求下快递配送网络鲁棒优化问题,使用改进的蚁群算法在满足时间窗的基础上对每条路径的使用时间进行精确计算;薛蓉娜等<sup>[17]</sup>根据 GRU 深度学习算法在不同日期的有效性对我国日均快递业务量进行预测,该算法的抗干扰性与预测准确率均较优。

上述学者的研究成果丰富了快递需求的研究内容,但仍存在一些问题,如对电子商务影响的定量研究和季节性特征的研究不足,数据往往为快递业务量单一数据且预测周期较长,预测精度也不够理想。因此本文从扩充数据多样性及提升模型预测精度出发,探讨电子商务及季节性差异影响下的快递需求预测问题,并使用 GM(1,N)模型与 ARIMA 模型进行组合预测。

## 二、组合模型的构建

### (一) 快递需求预测指标体系

在相关文献研究的基础上,将影响快递需求的指标分为宏观指标和电子商务指标,构建快递需求预测指标体系(表 1),指标说明如下。

#### 1. 宏观指标

地区生产总值。经济发展程度较高的地区往往具有人口密集、产业集群、基础设施完善等特征,这些地区的服务业发展水平通常也较高,快递服务作为服务业中的基础产业在该地区具有一定的发展优势。故区域经济的发展水平可以很好地反映该区域快递服务水平。

进出口总额。国内贸易在快递业发展初期产生了大量的快递需求,使我国快递业在短时期内得以实现规模性的增长。而随着我国国际贸易的蓬勃发展,跨境电商产业也如雨后春笋般茁壮成长,近年来跨境电商产业进出口总额高速增长,随之产生的进出口业务快递需求同样不容忽视。故选用进出口总额衡量国际贸易和国际快递需求的发展潜力。

第三产业增加值。快递业是为消费者提供运输服务及相关信息服务的基础服务业,从属于第三产业。快递需求的增长会直接推动第三产业的发展,第三产业发展水平的提高也会带动快递业的转型升级,两者具有相互促进的正相关关系。故本文选用第三产业增加值来反映包括快递业在内的第三产业发展水平。

社会消费品零售总额。近年来网络购物和快递送货的新型消费方式逐渐成为居民消费的主要途径。在实体购物中,人们也往往选择将体积较大或难以搬运的商品通过快递等方式送货上门,人们的消费行为越来越多地与快递需求产生联系。故本文选用社会消费品零售总额反映快递需求的潜在增长。

固定资产投资总额。我国主要的快递业基础设施包括道路、港口、机场、快递网点、网络及通信设备等,基础设施的完善对于快递业的发展起到关键作用。对其建设的投资力度越大,快递业的发展规模和潜力就越大,进而促进潜在快递需求的增加。由于目前缺少针对快递业的固定资产投资额统计指标,故

采用固定资产投资总额这一宏观指标进行替代。

公路和水路货运量。运输作为快递业的核心环节,任何快件都必须经过运输过程送达消费者手中,快递需求的升高与货运量的增加有着直接联系。目前我国快递的运输方式主要有水路、公路、铁路、空运四种,其中公路和水路是快递运输的频率最高的方式。由于目前缺少针对快递业的货运量统计指标,故采用公路和水路货运量这一宏观指标进行替代。

## 2. 电子商务指标

网络零售额。电子商务的发展直接推动了网络购物行为的普及,已成为人们购物的主要方式之一。网购商品设置快递和自取两种取货方式,其中绝大多数商品均为快递送货,网络购物群体也逐渐成为了使用快递服务的主力军。网络零售额的增加往往意味着相应的快递服务的增加,即快递需求上升。

跨境网络零售出口额。跨境出口电商打开了国外消费市场,目前主要的跨境电商物流模式有邮政小包、国际快递、国内快递、专线物流、海外仓储等。跨境网络零售出口额反映了跨境电商国际贸易的规模,也影响了国内外快递服务的变化。

电子商务产业增加值。电子商务产业增加值是电子商务产业发展规模的数据表现,国内外学者对快递业与电子商务产业的协同情况已做出了较多的研究。电子商务产业规模的增大,一方面会带动快递业产业规模扩大,另一方面会促进快递业经营模式的更新和快递服务的转型升级,均对快递需求产生积极影响。

居民人均可支配收入和居民网络消费额。居民消费能力的提升与潜在快递需求的提升有着直接关系,居民人均可支配收入作为潜在快递需求的主要来源,是反映居民消费能力的重要指标;而居民网络消费额是指居民在网络购物过程中实现的商品零售总额,是居民消费的数据表现,直接反映了居民网络购物的规模。

城镇和农村居民电脑拥有量。电脑作为电子商务实现的基础设施,是人们在互联网进行娱乐、消费等的主要工具,居民电脑拥有量能较好地衡量人们的网络消费基础。但应说明,在寻找相关网络基础设施统计指标(如互联网普及率)时,发现其他较普遍的指标往往为年度数据,不符合所需的数据要求,暂不纳入考虑范围。

表1 快递需求预测指标体系

因变量	一级指标	二级指标	指标作用/单位
Y: 快递需求	X1: 宏观指标	X11: 地区生产总值	反映区域经济水平对快递业的影响,亿元
		X12: 进出口总额	反映国际贸易状况和潜在的国际快递需求,亿元
		X13: 第三产业增加值	反映包括快递业在内的第三产业发展水平,亿元
		X14: 社会消费品零售总额	反映国内快递需求的潜在增长,亿元
		X15: 固定资产投资总额	反映快递业基础设施建设投入,亿元
		X16: 水路货运量	反映有水路运输需求的快递业务总量,万吨
		X17: 公路货运量	反映有公路运输需求的快递业务总量,万吨
	X2: 电子商务指标	X21: 网络零售额	反映国内通过网络购物实现的快递需求的潜在增长,亿元
		X22: 跨境网络零售出口额	反映国外通过网络购物实现的快递需求的潜在增长,亿元
		X23: 电子商务产业增加值	反映快递需求的市场规模,亿元
		X24: 居民网络消费额	反映了居民网络购物的规模,亿元
		X25: 居民人均可支配收入	反映居民消费能力,元
		X26: 每百户城镇居民电脑拥有量	反映了城镇居民的网路消费基础,台
		X27: 每百户农村居民电脑拥有量	反映了农村居民的网路消费基础,台

## (二) 基于灰色关联分析的指标关联度排序

### 1. 设置参考序列和比较序列

设参考序列为  $X_0 = \{X_0(t) \mid t = 1, 2, \dots, n\}$ ; 比较序列为:  $X_i = \{X_i(t) \mid t = 1, 2, \dots, n\}$ , 其中  $i = 1, 2, \dots, m$ 。

## 2. 数据的无量纲化处理

$$x_0(t) = \frac{x_i(t)}{x_i(1)} \quad t=1, 2, \dots, n; i=0, 1, 2, \dots, m \quad (1)$$

其中  $t$  为时间  $i$  为其他指标序列中的一行。

## 3. 计算灰色关联系数

$\xi_i(t)$  为其他指标序列对业务量序列在第  $t$  时刻的关联系数:

$$\xi_i(t) = \frac{\min_i \min_t |x_0(t) - x_i(t)| - \rho \max_i \max_t |x_0(t) - x_i(t)|}{|x_0(t) - x_i(t)| + \rho \max_i \max_t |x_0(t) - x_i(t)|} \quad (2)$$

其中  $\rho \in [0, 1]$  为分辨系数, 通常取  $\rho = 0.5$ ;  $\min_i \min_t |x_0(t) - x_i(t)|$  为最小两级差,  $\max_i \max_t |x_0(t) - x_i(t)|$  为最大两级差。

## 4. 计算灰色关联度

$r_i$  为其他指标序列对快递业务量序列在第  $t$  时刻的灰色关联度:

$$r_i = \sum_{k=1}^n w_k \xi_i(t) \quad (3)$$

$r_i$  值越接近 1, 说明相关性越好。

## 5. 关联度排序

将各关联度按照大小顺序排列, 从而更直观地判断其他因素序列对快递业务量序列的影响程度。

## (三) GM(1, N) 预测模型及其应用

### 1. 模型构建

将所预测地区的快递业务量序列记为  $x_1^{(0)}$ ; 其他相关指标序列记为  $x_i^{(0)}$ :

$$\begin{aligned} x_1^{(0)} &= \{x_1^{(0)}(1), x_1^{(0)}(2), \dots, x_1^{(0)}(n)\} \\ x_i^{(0)} &= \{x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n)\} \quad i=1, 2, \dots, N \end{aligned}$$

对序列  $x_i^{(0)}$  进行一次累加, 得到一阶累加生成序列  $x_i^{(1)}$ :

$$\begin{aligned} x_i^{(1)} &= \{x_i^{(1)}(k) \quad i=1, 2, \dots, N\} \\ x_i^{(1)}(k) &= \sum_{j=1}^k x_i^{(0)}(j) \quad i=1, 2, \dots, N \quad k=1, 2, \dots, n \end{aligned} \quad (4)$$

生成序列  $x_i^{(0)}$  的紧邻均值序列  $Z_1^{(1)}$ :

$$Z_1^{(1)}(k) = \frac{1}{2} [X_1^{(1)}(k) + X_1^{(1)}(k-1)] \quad k=2, 3, \dots, n \quad (5)$$

得到快递需求 GM(1, N) 预测模型的白化方程如下:

$$x_1^{(0)}(k) + aZ_1^{(1)}(k) = \sum_{i=2}^N b_i x_i^{(1)}(k) \quad (6)$$

### 2. 模型求解

由最小二乘参数估计法, 求得 GM(1, N) 预测模型的参数为:

$$\hat{\beta} = (B^T Y)^{-1} B^T Y \quad (7)$$

其中, 以下指标的解为:

$$B = \begin{bmatrix} -Z_1^{(1)}(2) & X_2^{(1)}(2) & \cdots & X_N^{(1)}(2) \\ -Z_1^{(1)}(3) & X_2^{(1)}(3) & \cdots & X_N^{(1)}(3) \\ \cdots & \cdots & \cdots & \cdots \\ -Z_1^{(1)}(n) & X_2^{(1)}(n) & \cdots & X_N^{(1)}(n) \end{bmatrix}; Y = \begin{bmatrix} X_1^{(0)}(2) \\ X_1^{(0)}(3) \\ \cdots \\ X_1^{(0)}(n) \end{bmatrix}$$



GM(1, N) 预测模型白化方程的解为:

$$x^{(1)}(k) = e^{-at} \left[ x_1^{(1)}(0) - k \sum_{i=2}^N b_i X_i^{(1)}(0) + \sum_{i=2}^N \int b_i X_i^{(1)}(t) e^{at} dt \right] \quad (8)$$

此时 GM(1, N) 模型的时间响应函数为:

$$\hat{x}_1(k+1) = e^{-ak} \left[ X_1^{(1)}(0) - \frac{1}{a} \sum_{i=2}^N b_i X_i^{(1)}(k+1) + \frac{1}{a} \sum_{i=2}^N b_i X_i^{(1)}(k+1) \right] \quad (9)$$

使用累减还原法, 求出快递业务量序列  $X_1^{(0)}$  的灰色预测值:

$$\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k) \quad (10)$$

#### (四) 时间序列 ARIMA 预测模型及其应用

考虑到灰色 GM(1, N) 预测模型具有线性拟合的特征, 模糊了快递需求的季节变化趋势, 因此使用 ARIMA 模型对快递需求的季节性特征做进一步预测。模型表达式为:

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1-L)^d X_t = \left(1 + \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t \quad (11)$$

ARIMA 模型预测步骤如下:

1. 对输入数据进行预处理。输入数据为所预测区域的历史快递业务量数据。
2. 识别序列平稳性。使用 ARIMA 模型需满足数据序列为平稳序列, 因此对历史快递业务量输入数据进行平稳性检验。通常同时使用时序图检验法、自相关图(ACF)与偏自相关图(PACF)检验法和单位根检验法。
3. 序列平稳化。若检验得出快递业务量原始数据为非平稳数列, 一般使用差分法使之平稳化。
4. 模型识别。参数  $p, q$  分别代表模型的自回归项数和滑动平均项数。对差分后的快递业务量序列画出自相关与偏自相关图, 确定  $p, q$  值。
5. 模型定阶。确定参数  $p, q$  时往往会产生多个符合条件的模型组合, 为选出精确度最高的预测模型, 可使用 AIC、BIC、HQ 最小准则进行模型比较。

$$AIC = 2k - 2\ln(L) \quad (12)$$

$$BIC = k\ln(n) - 2\ln(L) \quad (13)$$

$$HQ = -2\ln(L) + \ln(\ln(n)) * k \quad (14)$$

其中  $L$  为最大似然函数,  $n$  为数据数量,  $k$  为变量个数。

6. 残差检验。对该模型的有效性进行检验, 即检验该模型是否充足而有效地将数据中所包含的信息提取完全。

7. 得到预测值。ARIMA 模型具有静态或动态两种预测方法。静态预测即滚动地进行向前一步预测, 需手动调节预测数据范围, 精确度高; 动态预测步骤简单, 但精确度较低。本文选用静态方法进行预测。

#### (五) GM(1, N) - ARIMA 组合预测模型

为克服单一模型局限, 综合利用两模型优势, 建立 GM(1, N) - ARIMA 组合预测模型。选用方差倒数加权法进行权重分配, 利用误差与权系数成反比关系原理, 对于误差和较大的预测方法赋予较小的权数, 反之赋予较大的权数。设可确定第  $i$  种单项预测模型的权重为  $w_i$ , 该模型可表示为:

$$Y_t = \sum_{i=1}^2 w_i y_{ti} \quad (15)$$

其中, 权重系数的计算公式为:

$$w_i = \frac{(1/S_i)}{\left(\sum_{i=1}^2 \frac{1}{S_i}\right)} \quad S_i = \sum_{i=1}^2 (Y_t - y_{ti})^2 \quad (16)$$

式中:  $Y_t$  为  $t$  时间段内的组合预测值;  $y_{ti}$  为  $t$  时间段内 GM(1, N) 模型与 ARIMA 模型预测值;  $w_i$  为

GM(1 N) 模型与 ARIMA 模型权重系数; $S_i$  为 GM(1 N) 模型与 ARIMA 模型得到的拟合值序列方差。

为选出最具适用性的快递需求预测模型,使用模型评估指标 MAE、RMSE、MAPE 对比预测结果。指标计算公式如下:

$$\text{平均绝对误差 } MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_1 - y_i| \quad (17)$$

$$\text{均误差平方根 } RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{y}_1 - y_i|^2} \quad (18)$$

$$\text{平均误差率 } MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_1 - y_i|}{y_i} \quad (19)$$

### 三、杭州市快递需求预测的实证分析

#### (一) 数据来源

杭州市作为“电子商务之都”,是阿里、网易等电子商务平台和众多电商创业者崛起的发源地,多年来快递业务量排在全国前五之列,对其进行实证分析具有较强的借鉴意义。数据来源为《杭州市国民经济和社会发展统计公报》《杭州市邮政行业发展统计公报》。

#### (二) 灰色关联分析

使用 MATLAB2018a 对数据序列进行处理得到各指标的灰色关联度结果(图 1、表 2)。从整体来看,14 个指标序列与快递业务量序列灰色关联度均大于 0.55,表明所选指标与快递需求存在较强的相关性,可有效预测杭州市需求,但各指标的关联性并不均衡,波动幅度较大。电子商务因素指标中的四个指标与快递需求均有较强的相关性,有力证明了电子商务因素是快递业良性发展、快递业务量快速增长的有力支撑。

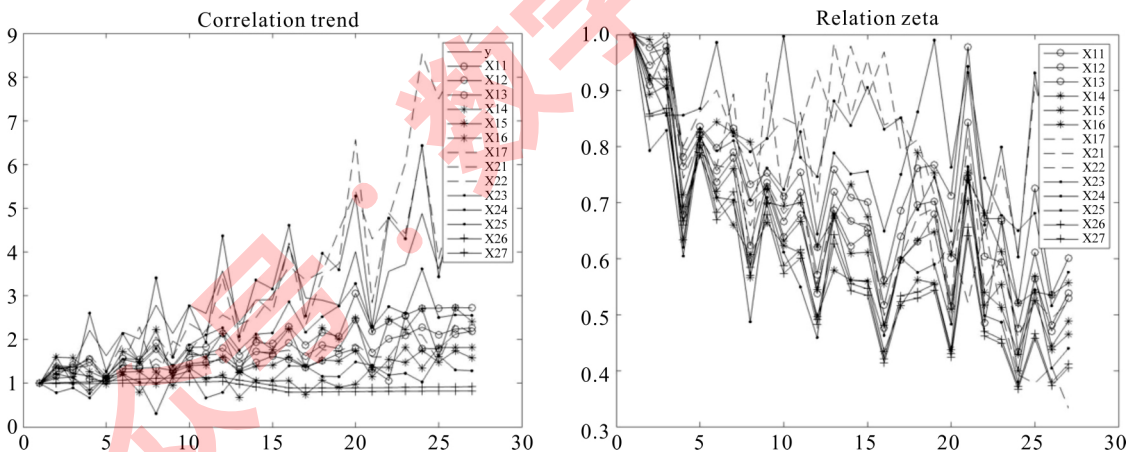


图 1 各指标结果与关联趋势图

表 2 各指标灰色关联度结果

指标	X11	X12	X13	X14	X15	X16	X17	X21	X22	X23	X24	X25	X26	X27
关联度	0.70	0.66	0.75	0.64	0.71	0.62	0.68	0.82	0.73	0.76	0.83	0.60	0.58	0.59

#### (三) GM(1 N) 预测模型应用分析

选取关联度大于 0.70 的指标作为因素行序列,建立基于 GM(1 6) 的快递业务量预测模型,使用 MATLAB2018a 进行计算。使用 2015 年 1 季度至 2020 年 4 季度数据作为模型训练数据以验证模型准确性,结果表明该训练结果的总体相对误差为 -0.38%(表 3)。

表 3 GM(1,6) 模型模拟预测值

季度	实际值	预测值	残差	总体相对误差
2021Q1	75 750.28	84 287.38	8 537.10	-0.38%
2021Q2	100 249.72	87 308.92	-12 940.80	
2021Q3	90 000.00	90 438.76	438.76	

对 2015 年 1 季度至 2021 年 3 季度的模型拟合值、残差值、平均相对误差值进行计算并作图(图 2)。结果表明,GM(1,6) 模型平均相对误差为 5.52%, 但该模型拟合值存在明显的线性趋势, 消除了快递需求季节性特征, 需对快递需求的季节性特征做进一步预测。

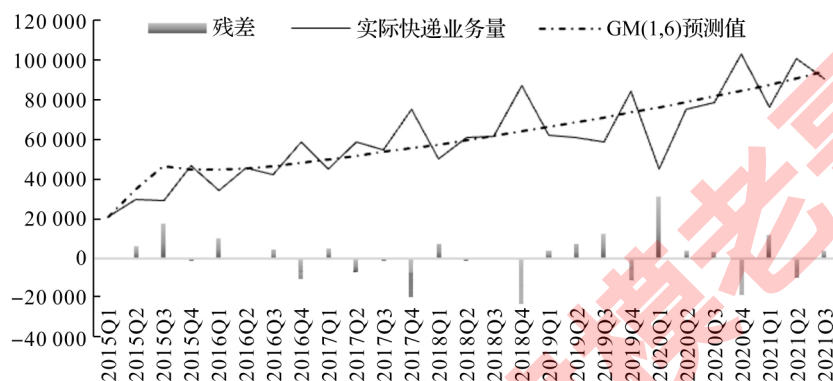
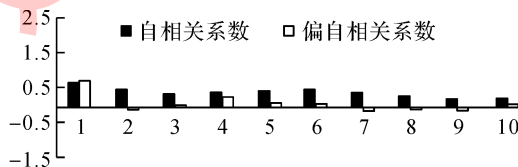


图 2 GM(1,6) 模型预测值与真实值对比

#### (四) ARIMA 预测模型应用分析

使用 Eviews11.0 软件建立 ARIMA 预测模型,原序列的 ADF 检验结果为 0.980 1,大于检验临界值(10% 显著性水平),证明该序列为非平稳序列。一阶差分后序列的 ADF 检验统计量为 0.025 3,小于检验临界值(5% 显著性水平),且消除了大部分的季节性趋势,证明一阶差分序列平稳。画出一阶差分序列的自相关与偏自相关图(图 3),

图 3  $Z_t^{(1)}$  序列自相关与偏自相关系数图

图中自相关系数拖尾,偏自相关系数截尾,且在 1 阶、4 阶较显著,因此可以尝试  $q=1,4$ 。反复比较不同参数组合后所产生的模型效果,最终确定 ARIMA(2,1,1)、ARIMA(1,1,4)、ARIMA(4,1,3)、ARIMA(4,1,4) 符合模型要求,采用 AIC、SIC、HQ 最小准则和残差的纯随机性检验确定最优模型(表 4)。

表 4  $Z_t^{(1)}$  序列各模型的检验参数值对比

模型	AIC	SIC	HQ
ARIMA(2,1,1)	19.87	19.99	19.92
ARIMA(1,1,4)	19.82	19.99	19.89
ARIMA(4,1,3)	19.67	19.86	19.72
ARIMA(4,1,4)	19.59	19.86	19.70

最后对 ARIMA(4,1,3) 与 ARIMA(4,1,4) 模型进行残差检验。结果显示,ARIMA(4,1,4) 模型的残差序列为非白噪声序列,排除此模型。选用 ARIMA(4,1,3) 模型进行杭州市快递需求预测,模型的平均相对误差为 -3.87%(图 4),季节性趋势的拟合效果较好。

#### (五) GM(1,N)-ARIMA 组合预测模型应用及比较分析

根据式(15)、式(16) 计算得出各预测模型权重系数为:  $w_1=0.564\ 3$ ,  $w_2=0.435\ 7$ , 则组合模型预测值为:  $\hat{Y}_t=0.564\ 3 \cdot y_{t1}+0.435\ 7 \cdot y_{t2}$ 。各模型预测值见表 5。

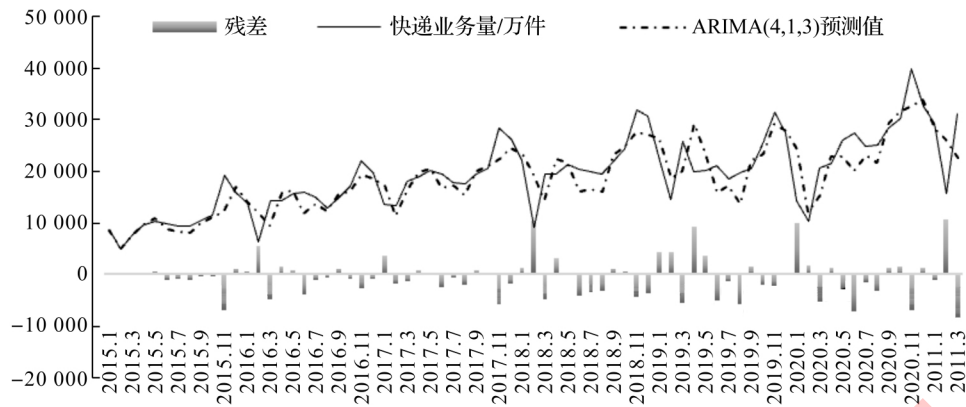


图4 ARIMA(4,1,3)模型月度预测拟合值与实际值对比

表5 杭州市2021Q4—2022Q4 快递业务量预测值

季度	GM(1,N)模型预测值	ARIMA模型预测值	组合模型预测值
2021Q4	93 927.95	104 522.37	98 702.63
2022Q1	97 147.81	79 664.10	89 268.27
2022Q2	100 478.04	99 088.41	99 851.76
2022Q3	103 922.44	74 630.84	90 721.34
2022Q4	107 484.91	93 309.25	101 096.25

各模型误差对比见表6。由于疫情影响,2020Q1的预测误差值显著偏高,剔除该特殊值影响后,61.5%的预测误差值在10%以内,其余在20%左右浮动,认为该模型精度在可接受范围内。采用评价指标综合比较各模型预测结果,GM(1,N)-ARIMA组合模型平均误差(1.30%)显著优于GM(1,6)模型(5.54%)和ARIMA(4,1,3)模型(-3.87%);平均绝对误差(3 557)、均方根误差(22 596)与平均误差率(7%)也小于两种单一模型,组合模型显著提高了预测精度。

表6 三种模型误差对比

评价指标	GM(1,N)模型	ARIMA模型	组合模型
平均相对误差	5.54%	-3.87%	1.30%
平均绝对误差 MAE	8 544	5 531	3 557
均方根误差 RMSE	59 275	34 986	22 596
平均误差率 MAPE	15%	9%	7%

## 四、结论

1. 快递需求受多种因素影响。电子商务因素中网络零售额、跨境网络零售出口额、电子商务产业增加值、居民网络消费额的灰色关联度均大于0.70,说明快递业与电子商务的协同发展状况良好;宏观因素中固定资产投资总额的灰色关联度大于0.70,说明基础设施的完善对潜在快递需求的影响较大。

2. 杭州市快递需求产生新特征。快递业务量将保持高速增长,2022年全年快递业务总量将达389 293.44万件,同比增长6%,并在第4季度达到861.89万件/天的高峰,迎来继疫情危机后一波新的业务量增长。季节性差异出现新特征,业务量次高值与最高值分别出现在2022年第2季度和第4季度;第1、2季度的业务量增速显著升高,同比增速分别为9.71%、9.87%。

3. GM(1,N)-ARIMA组合模型具有优越性。该组合模型集合了GM(1,N)模型和ARIMA模型的优势,同时考虑到电子商务环境和季节性特征两个重要因素,且预测精度最高,是研究快递需求预测较理想的方法。

本文的研究不足之处是,在进行预测时,由于数据获得的难度,预测指标的选择不够全面、精准;并且由于部分数据统计的缺失,对快递需求影响因素的归纳不够完善,也使得预测精度受到影响。



## 参考文献

- [1] 陆澜清. 2021 - 2026 年中国电商物流行业发展模式与投资战略规划分析报告 [EB/OL]. [2021 - 01 - 10]. <https://bg.qianzhan.com/report/detail/9e265a11045a4edc.html>.
- [2] 康阅春. 快递企业“爆仓”问题的解决方案[J]. 物流技术(装备版) 2011(16): 40 - 41.
- [3] 汤炎非, 黄静. 快递业季节性问题的研究[J]. 经济研究参考 2013(5): 46 - 51.
- [4] 魏邦艳. 探析我国快递业发展的现状以及发展对策[J]. 管理观察 2013(19): 50 - 52.
- [5] 邓欣伟, 刘丽娜. 基于神经网络的快递业务量预测分析研究[J]. 辽宁经济 2017(12): 82 - 83.
- [6] 李贞贞. 基于 ARIMA 简单季节模型的河南省快递业务量预测[J]. 价值工程 2019 38(17): 271 - 273.
- [7] Tang S, Deng G. Based on the Theory of Grey System to Forecast China's Business Volume of Express Services[J]. Modern Economy, 2015 6(2): 283 - 288.
- [8] Han S, Liu J, Wu H J. Research on the Regional Development of China's Express Delivery Industry and its Influencing Factors[J]. Journal of Applied Science and Engineering Innovation, 2015 2(9): 338 - 342.
- [9] Lv X H, Weng X G. Research on the Development of China's Express Industry Under the Background of E-commerce[J]. Journal of World Economic Research, 2016 5(4): 26 - 30.
- [10] 张文法. 基于灰色关联分析的山东省电子商务与物流业协同发展研究[J]. 数学的实践与认识 2019 49(17): 121 - 131.
- [11] 石咏梅. 快递业务量影响因素的实证分析[J]. 统计与管理 2018(6): 54 - 57.
- [12] Yu N, Xu W, Yu K L. Research on Regional Logistics Demand Forecast Based on Improved Support Vector Machine: A Case Study of Qingdao City under the New Free Trade Zone Strategy[J]. Quality Control, Transactions, 2020 8: 9551 - 9564.
- [13] Huang L J, Xie G J, Zhao W, et al. Regional logistics demand forecasting: a BP neural network approach[J]. Complex & Intelligent Systems, 2021( C31): 1 - 16.
- [14] 许荣斌, 王业国, 王福田 等. 基于改进 PSO - BP 算法的快递业务量预测[J]. 计算机集成制造系统 2018 24(7): 1871 - 1879.
- [15] 周杨, 甘陆军, 金龙. 基于 R 语言建立 SARIMA 滑动窗口模型的日度快递业务规模预测[J]. 物流技术 2020 39(10): 103 - 108.
- [16] 许闯来, 胡坚堃, 黄有方. 不确定需求下快递配送网络鲁棒优化[J]. 计算机工程与应用 2020 56(3): 272 - 278.
- [17] 薛蓉娜, 张明敏, 南玉婷 等. 基于 GRU 深度学习算法的日均快递业务量预测模型[J]. 统计与决策 2021 37(13): 176 - 179.

## A Research on Express Demand Forecast Based on Combination Forecasting Model

CHEN Chou-yong, GAO Ming-jing

(School of Management, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China)

**Abstract:** Considering that the express demand is affected by the e-commerce environment and seasonal characteristics, this paper brings the e-commerce factors into the express demand prediction index system, using the grey correlation analysis method to determine the significant correlation index, the GM (1, N) model to predict the trend and the time series ARIMA model to predict the seasonal difference, and establishing the GM (1, N) of the above two types of models-Arima weight combination model to improve the prediction accuracy. The first mock exam of Hangzhou's express demand in the 4th quarter of 2021 and the 4th quarter of 2022 is carried out by using the combined forecasting model. The results show that the prediction error of the combined forecasting model is significantly better than that of the selected single model, the impact of the development of e-commerce on the demand for express delivery will gradually increase and the seasonal difference of express demand will appear some new characteristics.

**Key words:** express demand; e-commerce; grey correlation analysis; ARIMA model; combination forecast