

城市物流运输网络优化与需求预测研究

摘 要

本题是关于我国现代物流业发展中,快递需求分析的问题。分析目标包括:对城市的重要性量化、对未来发货情况及发货量的预测、优化城市之间的运输网络,从数学理论的角度给出固定需求常数和非固定需求常数的计算。这些对于我国物流业的工作者来说具有重要的现实意义。

首先,题中所给的物流运输数量的附件数据具有明显的缺失、异常。考察数据的特征,对缺失和异常进行合理的数据填补。一定程度上提升了时间序列的平稳性。

针对问题一,选择收发货量的年均值与最大值、上下游城市数量、日均上下游收发量、快递数量日月变化率均值等 8 个指标来量化城市重要程度。在求解权重时,本文选取优序图法和熵权法综合的**加法集成赋权法**,将主观赋权与客观赋权综合起来,提高模型的可信度。最终给出重要性排名前五的城市依次为:L、G、V、X、R。

针对问题二,题中数据为严格的时间序列数据,考虑到数据本身的特点,对问题二采用**ARIMA 时间序列模型**,推测目标日期中,城市间的快递运输数量以及快递运输总量,得出所有“发货-收货”城市之间的快递运输数量。

针对问题三,题中数据均是密度较大的时间序列,且并未给出对突发事件的考量,这导致我们无法通过已有的时间序列来精确推得下一个时间序列的状态,因此无法使用传统的时间序列预测或神经网络的方法。考虑到突发事件的影响可能具有周期性,我们筛选了过去 3 年同时期的数据,根据同期数据的发货情况进行预测。在判定能发货的前提下,提取所有城市对近 4 周的每周日平均快递运输数量。建立**灰色预测模型**,将模型值作为后两日快递运输数量的预测值。

针对问题四,考虑实际情况与运费,规定实际装货量不超过额定装货量,并且城市对之间的运输尽可能选择较少的路径数,在此基础上得出 2023 年 4 月 23 日-4 月 27 日的最低运输成本分别为:1110.89、1166.22、1189.88、1033.78、1399.40。

针对问题五,查阅文献后,将 $\mu - 2\sigma$ 作为固定需求成本,通过 XGBoost 对已有数据训练预测,检验其准确性,并通过非参数估计中的**核密度估计**给出非固定需求概率分布估计方法。

总之,对于未来快递运输数量的预测,本文采用 ARIMA 时间序列和灰色预测模型,分别对应不同的场景,模型精度高。在目标优化模型中,充分利用城市对之间的铁路运输网络,大大降低运输成本。查阅文献后建立固定需求常数的计算模型,理论依据强。另外,本文没有使用复杂的或可解释性差的模型,而是注重结果的可解释性和现实意义。以上是本文的核心优点。

关键词: 加法集成赋权法 ARIMA 时间序列模型 灰色预测 核密度估计

一、问题重述

1.1 问题背景

网络购物作为互联网技术和电子商务发展的产物，已经深入人们的日常生活。这一现象带来了快递服务需求的激增，对我国经济发展产生了重大影响。在快递公司的运输过程中，准确预测快递运输需求数量以及制定合理的运输策略对于仓库布局、存储成本节约和运输线路规划具有重要意义。然而，由于受到诸如突发事件、季节性因素和市场竞争等多种因素的影响，快递需求呈现出较大的波动性，使得预测和规划变得更加困难。为了解决这个问题，可以通过建立数学模型或使用机器学习算法来分析现有数据。例如，可以使用传统的统计方法或新工具来准确预测并保持优秀的性能。这些模型将能够学习许多超越传统统计模型能力的关系。本文旨在为快递公司提供有效支持，帮助其优化运营策略，降低运输成本，提高服务质量。

1.2 问题要求

1、通过对附件 1 数据的处理，从发货量、收货量、快递数量增长变化趋势等多角度因素的分析，建立数学模型对各站点城市的重要程度进行综合排序，比较各个城市对快递公运营的重要性。

2、利用附件 1 数据，建立数学模型来预测未来各“发货-收货”站点城市之间的快递运输数量。给出 2019 年 4 月 18 日和 2019 年 4 月 19 日所有城市对之间的快递运输数量，这有助于快递公司制定更为合理的运输计划，从而降低成本并提高运营效率。

3、利用附件 2 数据，考虑突发事件影响的现实背景下，利用数学模型预测可正常发货的站点城市对以及对应的快递运输数量。这将有助于快递公司在面临不确定性因素时，做出更为稳妥的运输决策。

4、结合铁路运输网络和运输成本计算公式，建立数学模型找出成本最低的运输方案。这将有助于快递公司协调运输方案，提高社会经济效益。

5、从固定需求和非固定需求的角度分析快递需求，并按季度估计固定需求常数和非固定需求概率分布。这将有助于快递公司更好地了解需求变化的规律，为制定合适的运输策略提供依据。

二、问题分析

2.1 总体分析

本题是一个关于我国现代物流业发展中，快递需求分析的问题。

从分析目的来看，需要对城市在物流网中的重要程度进行排名。根据已有数据预测未来快递物流的变化，提升快递公司运输计划的合理性。考虑实际运输方案后，建立最低运输成本方案，提升快递物流的社会经济效益。从固定需求和非固定需求的角度解释物流变化规律。

从数据特征来看，附件中所给的数据量丰富，包含各个时间点，各个物流链。剔除了因

为电商的购物促销活动容易导致数据不精确的 6 月、11 月、12 月。由于数据量大，需要在解决每一个问题时对数据做一定的预处理。

从选择模型来看，由于物流业具有很强的现实意义，且分析目的与实际情景息息相关。因此构建的模型应当追求模型的可解释性，不宜使用过于复杂的模型。

2.2 针对问题一的分析

问题一的核心目标在于构建城市重要性的评估模型。在收货量指标中，选取日最大收货量和日平均收货量；在发货量的指标中，选取日最大发货量和日平均发货量；在相关性的指标中选取，上下游城市数量和日均收发货量；在快递数量增长或减小趋势中选取日变化量的均值和月变化量的均值。通过赋权，构建综合评价模型，计算每个城市的得分。

2.3 针对问题二的分析

问题二的核心在于利用已有数据对未知数据进行推测。附件 1 给出的数据为严格的时间序列数据，考虑到数据本身的特点。对问题二采用 ARIMA 时间序列模型，推测 2019 年 4 月 18 日-2019 年 4 月 19 日两天，城市间的快递运输数量以及快递运输总量。

2.4 针对问题三的分析

问题三的核心目标在于对城市能否发货进行判别预测，若能发货再对发货量预测。但已有数据均是密度较大的时间序列，且并未给出对突发事件的考量，这导致我们无法通过已有的时间序列来精确推得下一个时间序列的状态，因此无法使用传统的时间序列预测或神经网络的方法。考虑到突发事件的影响可能具有周期性，我们筛选了过去 3 年同时期的数据，根据同期数据的发货情况进行预测。在判定能发货的前提下，提取所有城市对近 4 周的每周日平均快递运输数量。建立灰色预测模型，将模型值作为后两日快递运输数量的预测值。

2.5 针对问题四的分析

问题四的核心在于根据所给的站点城市间铁路运输网络，以最低运输成本为目标，优化 2023 年 4 月 23 日至 4 月 27 日的快递运输路线。考虑实际情况与运费，规定实际装货量不超过额度装货量，并且城市对之间的运输尽可能选择较少的路径数。

2.6 针对问题五的分析

问题五的核心在于对同一季度中城市对的固定需求常数和非固定需求常数进行量化。查阅文献后，将 $\mu - 2\sigma$ 作为固定需求成本，通过 XGBoost 回溯预测法检验其准确性，并通过非参数估计中的核密度估计给出非固定需求概率分布估计方法。

三、模型假设

针对上述问题，我们提出如下假设：

- 1、假设 2023 年同时期会发生和往年类似的突发事件。
- 2、假设城市对之间共用铁路线时，成本不会发生相互影响。
- 3、假设每件快递的重量为单位 1，且仅考虑运输成本。
- 4、假设实际装货量可以超过额定装货量，但会导致成本增加。
- 5、假设城市间铁路运输选择尽可能少的路径数量。

四、符号说明

符号	说明
I_i	站点城市重要城市
y_t	t 时刻原始序列
n_{path}	城市对之间需求的路径数
c_i	第 i 组城市对的运输成本
$m_{package}$	城市对需要快递运输的数量

其余表中未给出符号在相应部分会有说明。

五、数据预处理

5.1 对无法正常收发货的数据处理

赛题中提到“无数据表示无法正常收发货”。因此，对时间序列中无法正常收发货的城市对，其实际收发货量定为 0。

5.2 对无发货需求数据的处理

赛题中提到“0 表示无发货需求”。在此快递公司的数据提供的集中，对于无发货需求的数据统一用 21 填充，其合理性体现在三方面：

其一，查阅文献后发现，部分快递公司会将一个运输数量作为阈值，不对运输数量很小的数据进行统计。若运输数量小于其阈值，则会被定义为无发货需求。考察所给的数据集，发现附件 1、附件 2 所给的每月所有城市对的快递运输数量最小值均为 42，且在大部分情况下 42 为当月所有值的众数。推测该快递公司对发货需求的阈值设置为 42。

其二，在后续解决问题时，对数据进行时间序列分析，大量的异常值的出现，会降低时间序列的平稳性，从而影响模型的精度。

其三，根据题中所给数据，无法得知实际的发货数量，故采用 42 与 0 的平均值 21，对无发货需求的值进行填补，最大程度上弥补了数据本身的缺陷。

六、模型的建立与求解

6.1 问题一的模型建立与求解

6.1.1 站点城市的重要程度的量化

评估站点城市的重要程度对物流企业至关重要，可提高物流效率和客户满意度，增强企业竞争力。其受到地理位置、收发货量、快递数量增长趋势和与其他站点城市的联系等多方面因素的影响。本文从收发货量、快递数量变化率和相关性四个方面对站点城市的重要程度进行了量化。

1、收发货量

站点城市的收发货量是衡量其重要程度的重要指标之一，通常表现为该城市的物流需求量和物流服务能力。本文分别计算每个站点城市一年内的收货量和发货量的平均数和最大值。用 \bar{r}_{ij} 表示 j 城市的年均收货量， \bar{s}_{ij} 表示 j 城市的年均发货量， $\max r_{ij}$ 表示 j 城市的年最大收货量， $\max s_{ij}$ 表示 j 城市的年最大发货量， i 代表年份。

2、相关性

站点城市的相关性也是评估其重要程度的重要指标。重要程度较高的站点城市通常有较多的上下游通过路径数，并且与上下游城市的收发货量也有较强的相关性。故本文从站点城市的上下游通过路径数和日均上下游收发量评估相关性。具体计算公式如下：

(1) 站点城市上下游路径数

$$n_j = \sum_i n_{ij} + n_{jk} \quad (1)$$

式中， n_j 是 j 城市上下游路径数， n_{ij} 是由 i 城市到 j 城市的路径数， n_{jk} 是由 j 城市到 k 城市的路径数。

(2) 日均上下游收发量

$$\bar{r}_j = \frac{\sum_k \sum_i r_{ij}^k + s_{jm}^k}{k_{\max}} \quad (2)$$

式中， \bar{r}_j 是城市日均上下游收发量； r_{ij}^k 是第 k 天 j 城市的收货物量； s_{jm}^k 是第 k 天 j 城市的发货量。

3、快递数量日（月）变化率均值

快递数量日（月）变化率均值可以反映城市的快递业务增长速度、商业潜力和在物流网络中的地位。具体计算公式如下：

$$\overline{dc} = \frac{\sum_k [(\sum_j r_{ij}^k + s_{jl}^k) - (\sum_j r_{ij}^{k-1} + s_{jl}^{k-1})]}{k_{\max} - 1} \quad (3)$$

式中， \overline{dc} 是日变化率均值

$$\overline{mc} = \frac{\sum_m [(\sum_j r_{ij}^m + s_{jl}^m) - (\sum_j r_{ij}^{m-1} + s_{jl}^{m-1})]}{m_{\max} - 1} \quad (4)$$

式中， \overline{mc} 是月变化率均值； r_{ij}^m 是第 m 月 j 城市的收货量； s_{jl}^m 是第 m 月 j 城市的发货量。

6.1.2 站点城市重要性模型建立与求解

站点城市重要性从三方面考虑：站点繁忙程度，站点城市的地理位置与交通便利程度。站点进出货物越多该站点越繁忙。而站点城市的地理位置与交通便利程度可从该站点的上下游城市数与日均上下游收发量进行衡量。

对上述四个指标进行简单的线性加权，建立出企业生产重要性模型如下：

$$I_i = \lambda_1 \bar{r}_{ij} + \lambda_2 \bar{s}_{ij} + \lambda_3 \max r_{ij} + \lambda_4 \max s_{ij} + \lambda_5 n_j + \lambda_6 \bar{r}_j + \lambda_7 \overline{dc} + \lambda_8 \overline{mc} \quad (5)$$

对于权值的求解，我们选定优序图法和熵权法综合的加法集成赋权法，将主观赋权法和客观权重法综合，得到最优权重：

$$\lambda_j = k_1 \lambda_j^1 + k_2 \lambda_j^2 \quad (6)$$

式中， λ_j^1 和 λ_j^2 分别表示优序图法和熵权法对第 j 项指标的权重， k_1 和 k_2 分别为主观和客观权重的重要程度系数， λ_j 表示第 j 项指标的最优权重。

（一）优序图法求解 λ_j^1

Step1：画出棋盘图，进行对比评分

优序图法通过棋盘图对指标两两对比，选用 0，0.5，1 三个数字来描述指标优劣性，进而计算指标权重的主观赋权法。由于数据指标较多，逐一赋权可能会降低模型的精确度。故团队对上述指标进行分类归纳后赋权，各个指标的优先程度如下：

日均上下游收发量>上下游路径数>年均收货量=年均发货量>年最大收货量=年最大发货量>日变化率均值=月变化率均值

指标评价情况如下表所示：

表 1 指标类型评分棋盘

	$\bar{s}r_j$	n_j	\bar{r}_{ij}	\bar{s}_{ij}	$\max r_{ij}$	$\max s_{ij}$	\overline{dc}	\overline{mc}
$\bar{s}r_j$	0.5	1	1	1	1	1	1	1
n_j	0	0.5	1	1	1	1	1	1
\bar{r}_{ij}	0	0	0.5	0.5	1	1	1	1
\bar{s}_{ij}	0	0	0.5	0.5	1	1	1	1
$\max r_{ij}$	0	0	0	0	0.5	0.5	1	1
$\max s_{ij}$	0	0	0	0	0.5	0.5	1	1
\overline{dc}	0	0	0	0	0	0	0.5	0.5
\overline{mc}	0	0	0	0	0	0	0.5	0.5

Step2：计算出指标权重

将上述 Excel 表格导入 SPSSAU 软件中进行运算，得出各指标权重 λ_j^1 。

（二）熵权法求解 λ_j^2

熵权法是一种客观赋权法，其根据指标离散程度的信息熵值来确定指标对综合评价的权重。该方法简单易行，具有较高的可信度和精确度，可在实践中广泛应用。

Step1：确定指标体系，清洗指标极值

选择日均上下游收发量、上下游路径数、年均收货量、年均发货量、年最大收货量、年最大发货量、日变化率均值和月变化率均值作为评价指标。为了避免极值数据对指标的熵值产生影响，对各指标中的极值进行了剔除处理。

Step2：归一化指标处理

采用归一化处理对指标进行评价，可以将指标的实际值转化为不受量纲影响的指标评价。具体的计算公式如下：

$$x_{ij} = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j} \quad (7)$$

式中， x_{ij} 表示归一化后的城市 i 的第 j 个指标； $\min x_j$ 表示第 j 个指标的最小值； $\max x_j$ 表示第 j 个指标的最大值； x_{ij} 表示城市 i 的第 j 个指标。

Step3: 计算指标熵和权重

首先需要计算出城市 i 在第 j 个指标上的权重比例，然后再通过计算该指标的信息熵来确定其对综合评价的权重。最终可以得到第 j 个指标的权重。具体的计算公式如下：

$$y_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \quad (8)$$

$$e_j = -k \sum_{i=1}^m y_{ij} \ln y_{ij} \quad (9)$$

$$\lambda_j^2 = \frac{1 - e_j}{\sum_j (1 - e_j)} \quad (10)$$

式中， $k = \ln m$ 为常数， y_{ij} 表示城市 i 第 j 个指标的比重， e_j 表示第 j 个指标的信息熵， λ_j^2 表示第 j 个指标的权重。

6.1.3 模型结果

熵权法的客观赋权法主要由表格中的信息熵确定，具有统计学意义。而优序图的主观赋权法的权重则通过人为评判所得，更具有现实意义。将优序图的主观赋权的权重设定为 0.3，熵权法的客观赋权法的权重定为 0.7，得出最终权重结果。

表 2 最终权重结果图

序号	指标	主观赋权法	客观权重法	最优权重
		优序图法	熵权法	
1	日均上下游收发量	0.2344	0.2105	0.2177
2	上下游路径数	0.2031	0.1067	0.1356
3	年均收货量	0.1563	0.1486	0.1509
4	年均发货量	0.1563	0.1910	0.1806
5	年最大收货量	0.0938	0.1064	0.1026
6	年最大发货量	0.0938	0.1321	0.1206
7	日变化率均值	0.0313	0.0733	0.0607
8	月变化率均值	0.0313	0.0313	0.0313

因此站点城市的重要程度可量化为：

$$I_i = 0.1509\bar{r}_{ij} + 0.1806\bar{s}_{ij} + 0.1206 \max r_{ij} + 0.1026 \max s_{ij} + 0.1356n_j + 0.2177\bar{s}_j + 0.0607\bar{dc} + 0.0313\bar{mc}$$

将计算得到的站点城市的指标值分别代入上述公式中，求出站点城市重要性的值，并对其进行排序。重要性排名前 5 名的站点城市如下表所示：

表 3 站点城市重要性排名

排序	1	2	3	4	5
城市名	L	G	V	X	R
重要性	64.0535	63.3711	58.3700	56.2029	55.7972

此外，后三名的城市分别为 E、P、C。重要性排名前五的城市在年均收货量、年均发货量、年最大收货量、年最大发货量、路径数、日均上下游收发量上排名皆很高，后三名的城市在这些指标上排名较差，尤其在日均上下游收发量和年均发货量上，证明该分析所选权重是合理的。

6.2 问题二模型的建立与求解

6.2.1 选用 ARIMA 模型原因

本题预测目标是根据附件 1 给出的一年间各站点城市之间的快递运输数据预测 2019 年 4 月 18 日和 2019 年 4 月 19 日各站点城之间快递运输数量。本文采用 ARIMA 模型进行预测。选用原因从样本特征及 ARIMA 模型优点两方面考虑：

对附件 1 数据进行预处理，得出附件 1 数据有如下特征：

存在季节性变化趋势：快递运输数量的数值在不同的时间段内呈现出规律性的波动，即存在季节性，快递运输数量在某些月份呈现出周期性的波动。且“基于 ARIMA 模型的快递业季节性预测研究”^[1]利用 ARIMA 模型对快递运输量的季节性进行了预测研究，结果表明快递业存在明显的季节性变化趋势。

ARIMA 模型在预测快递运输数量时的优点包括：

- 1、能够捕捉到时间序列数据的趋势和季节性变化，使得预测结果更加准确。
- 2、可以处理非平稳时间序列数据，使得预测结果更加可靠。
- 3、可以通过调整模型的参数来提高预测精度，如增加 AR 部分的阶数、MA 部分的阶数。
- 4、综上，选用 ARIMA 模型进行快递运输数量的预测是合理的。

6.2.2 ARIMA 模型求解

ARIMA 模型是一种时间序列分析方法，其内部的具体流程如下：

- 1、确定时间序列是否平稳。如果时间序列不平稳，需要进行差分处理，直到时间序列变为平稳序列。
- 2、通过自相关函数（ACF）和偏自相关函数（PACF）分析，确定 ARIMA 模型的参数 p 、 d 和 q 。其中， p 表示自回归项的阶数， d 表示差分次数， q 表示移动平均项的阶数。
- 3、根据确定的 ARIMA 模型参数，建立 ARIMA 模型。
- 4、对 ARIMA 模型进行拟合，得到模型的参数估计值。
- 5、对模型进行检验，包括残差分析、模型诊断等步骤。如果模型不符合要求，需要重新调整模型参数，重新拟合模型。
- 6、利用已经拟合好的 ARIMA 模型进行预测。

ARIMA 模型是一种时间序列预测模型，它可以用于预测未来的值。使用 ARIMA 模型进行城市间的快递运输数量的预测的具体公式如下：

$$ARIMA(p, d, q): y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (11)$$

8

它的预测过程可以分为以下步骤：

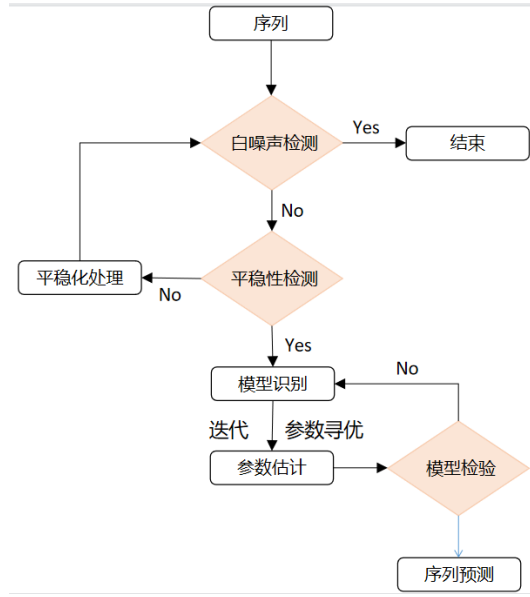


图 1 ARIMA 参数寻优流程图

1、差分操作

将原始时间序列进行差分操作，得到一个平稳的时间序列，即 ARIMA 模型中的 I（差分）项。具体公式如下：

$$y'_t = y_t - y_{t-1} \quad (12)$$

y'_t 是 t 时刻差分后的序列； y_t 是 t 时刻原始序列； y_{t-1} 是 $t-1$ 时刻原始序列。

2、确定模型阶数

确定 ARIMA 模型的阶数，包括 AR（自回归）项、MA（移动平均）项和差分项。

AR（自回归）项：

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} \quad (13)$$

MA（移动平均）项：

$$y_t = \varepsilon_t + \sum_{j=1}^q \theta_{t-j} \quad (14)$$

差分项：

$$y'_t = \Delta^d y_t \quad (15)$$

式中， d 为差分次数。

3、估计模型参数

根据已知数据，通过极大似然估计或其他方法，估计模型的参数，包括 AR（自回归）系数、MA（移动平均）系数和方差项。

使用时间序列数据进行参数估计，常用的方法是极大似然估计（ALE），具体公式如下：

$$L(\phi_1, \phi_2, \dots, \phi_m | y) = \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{y_t - \sum_{i=1}^p \phi_i y_{t-i}}{2\sigma^2} \quad (16)$$

$$L(\theta_1, \theta_2, \dots, \theta_m | y) = \prod_{t=q+1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{y_t - \varepsilon_t - \sum_{i=1}^p \theta_i y_{t-i}}{2\sigma^2} \quad (17)$$

4、模型检验

对估计得到的模型进行检验，常用的方法是检验残差序列是否为白噪声，分为 Ljung-Box 检验和 Jarque-Bera 检验，具体公式如下：

Ljung-Box 检验：

$$Q(k) = n(n+2) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{n-j} \quad (18)$$

式中， n 为时间序列长度； k 为检验的滞后阶数； $\hat{\rho}_j^2$ 为样本自相关系数。

Jarque-Bera 检验：

$$JB = n \left(\frac{S^2}{6} + \frac{(K-3)^2}{24} \right) \quad (19)$$

式中， n 为时间序列长度； S 为样本偏度； K 为样本峰度。

5、模型预测

利用已估计的 ARIMA 模型，对未来的时间序列值进行预测，具体公式如下：

$$\hat{y}_{t+h} = y_t + \sum_{i=1}^h \hat{\varepsilon}_{t+i} \quad (20)$$

在进行模型检验时，采用均方根误差对模型的预测效果进行评估，具体公式如下：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (21)$$

6.2.3 模型结果

由于篇幅受限，只展示部分“发货-收货”城市之间的模型参数、拟合优度、预测值等信息，其余城市之间的预测结果已在附件中给出。

表 4 指定城市对预测值及拟合优度

城市站点对	2019 年 4 月 18 日 预测值	2019 年 4 月 19 日 预测值	拟合优度	模型参数
M-U	142	143	0.812	ARIMA (0,1,0)
Q-V	41	33	0.744	ARIMA (2,1,2)
K-L	66	80	0.734	ARIMA (1,0,1)
G-V	565	534	0.663	ARIMA (1,0,2)

V-G	535	485	0.686	ARIMA (1,0,1)
A-Q	128	128	0.76	ARIMA (0,1,0)
D-A	43	44	0.635	ARIMA (1,1,1)
L-K	293	293	0.765	ARIMA (0,1,0)
城市总运输量	10659	10676	0.837	ARIMA (0,1,2)

注：考虑实际情况，已对快递运输数量数据进行取整处理。

对问题 2 结果的汇总表如下：

表 5 问题 2 结果表

日期	“发货-收货”城市之间的快递运输数量		所有“发货-收货”城市之间的总快递运输数量
2019 年 4 月 18 日	M-U	142	10659
	Q-V	41	
	K-L	66	
	G-V	565	
	V-G	485	
2019 年 4 月 19 日	A-Q	128	10676
	D-A	44	
	L-K	293	

注：考虑实际情况，已对快递运输数量数据进行取整处理。

6.3 问题三的模型的建立与求解

6.3.1 建立能否正常发货的预测模型

本模型的预测目标是：推测在突发事件的影响下，站点城市对之间能否正常发货。但是，已有数据均是密度较大的时间序列，且并未给出对突发事件的考量，这导致我们无法通过已有的时间序列来精确推得下一个时间序列的状态，因此无法使用传统的时间序列预测或神经网络的方法。

基于以上原因，考虑到突发事件的影响可能具有周期性，我们筛选了过去 3 年同时期的数据，根据同期数据的发货情况进行预测。查阅相关文献后发现，中长途物流的周期在 3-5 天，省内物流周期在 1-3 天。^[2]取历史四年数据中的 4 月 23 日至 5 月 4 日，共 36 天作为同期数据，具体预测方法如下：

考察同期历史数据中，能发货的城市对占比约 84%，则可认为所有城市对中能发货的占比为 84%。考察在同时期历史数据中，所有城市对的总发货次数，若其发货次数排名>84%，

则预测未来两天均会发货，否则认为不会发货。

这样做的合理性和好处有以下五点：

- 1、考虑了同期历史数据中的突发事件。
- 2、在突发事件不明确的条件下，防止盲目套用时间序列或神经网络模型导致预测错误。
- 3、预测方法相对直观，解释性强。
- 4、不容易受历史数据中疫情期间持续性不发货的影响。
- 5、保证了正常发货的城市对的频率不变。

根据上述预测模型，预测表格中的城市对不能正常发货的有：I-S，S-Q，Q-O，U-O。其余均能正常发货。其余“发货-收货”站点城市对能否发货结果已在附录中展示。

6.3.2 ARIMA 模型的缺陷

附件 2 所给数据为 2020 年 4 月 28 日-2023 年 4 月 27 日的快递运输数量数据。结合时代背景，该时间序列受疫情影响，造成不能正常发货的异常值多、极端高值多，故用时间序列预测易造成拟合精度差、时间序列平稳性差等。下图为对“发货-收货”站点城市对 Y-L 的时间序列采用自动寻优算法，后根据 ARIMA (1,0,1) 模型拟合后，拟合值与真实值对比，其拟合精度为 0.047，无法满足对未来快递运输数量的预测。

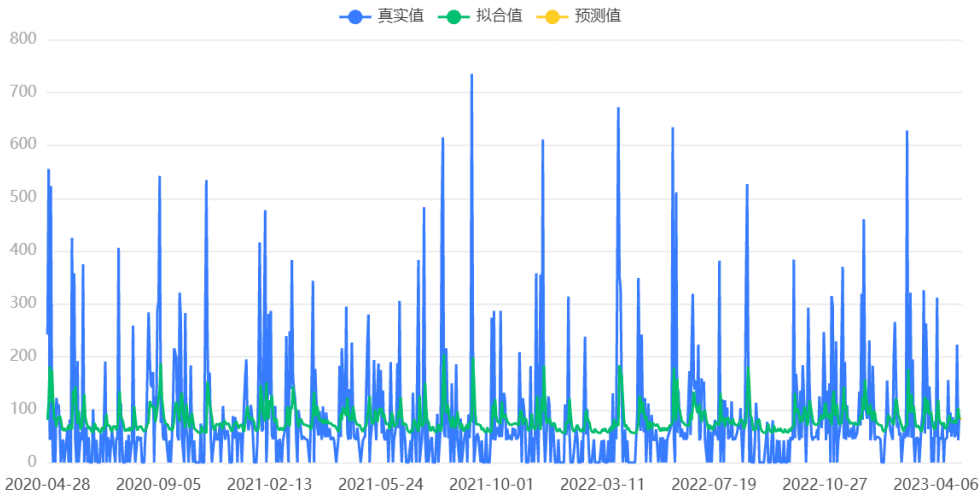


图 2 ARIMA(1,0,1)模型对 Y-L 时间序列的拟合

6.3.3 基于灰色预测对模型的改进

6.3.3.1 GM(1, 1) 灰色预测模型

一般情况下，平稳的数列容易进行预测。然而很多情况下原数列不平稳，数据波动较大而较难预测。灰色预测模型的优点为不需要很多的数据，就能解决历史数据少，序列完整性及可靠性低的问题，且 GM(1,1)模型适用于中短期预测^[3]。

具体建模思路为：提取所有城市对近 4 周的每周日平均快递运输数量。建立灰色预测模型，将模型值作为后两日快递运输数量的预测值。模型的优点在于：对每周数据进行计算，一定程度上减小了时间序列的波动，且利用每周日均值的运输数量，预测数据具有平稳性，

且有很强的现实意义。

6.3.3.2 灰色预测建模过程

GM(1,1)模型首先对原数列进行累加得到新的平稳数列，灰色预测模型具体建模步骤是：

(1) 令原始数据为：

$$X^{(0)} = (x^{(0)}(t), t = 0, 1, 2 \dots n) \quad (22-1)$$

$$X^{(0)} = (X^{(0)}(1), X^{(0)}(2), \dots X^{(0)}(n)) \quad (22-2)$$

(2) 对原始数据进行累加使数据趋于平滑，生成新数列，记为：

$$X^{(1)} = (X^{(1)}(1), X^{(1)}(2), \dots X^{(1)}(n)) \quad (23)$$

(3) 建立相应的微分方程：

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (24)$$

(4) 用最小二乘法得到 a 和 u ：

$$[au]^T = (B^T B)^{-1} B^T Y \quad (25)$$

其中

$$Y = (X^{(0)}(2), X^{(0)}(3) \dots \dots X^{(0)}(n))^T \quad (26)$$

$$B = \begin{pmatrix} -\frac{1}{2}(X^{(1)}(1) + X^{(1)}(2)) & 1 \\ -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(X^{(1)}(n-1) + X^{(1)}(n)) & 1 \end{pmatrix} \quad (27)$$

(5) 求出 a 和 u 后建立微分方程所对应的时间响应函数，即：

$$X^{(1)}(t+1) = \left(X^{(0)}(1) - \frac{u}{a}\right)e^{-at} + \frac{u}{a} \quad (28)$$

(6) 对模型进行检验

残差值： $\varepsilon^{(0)}(t) = X^{(0)}(t) - X^{(0)}(t)$ ，其中 $X^{(0)}(t)$ 为模拟值

相对误差： $q(t) = \frac{\varepsilon^{(0)}(t)}{X^{(0)}(t)} \times 100\%$

(7) 将实测值代入时间相应函数，进行预测

6.3.3.3 模型结果

由于篇幅限制，此处仅展示部分站点城市对的预测结果，其余站点城市对结果已在附录中给出。

表 6 问题 3 结果表

日期	“发货-收货” 站点城市对	是否能 正常发货	快递运输 数量	灰色预测相 对误差 (%)
2023 年 4 月 28 日	I-S	否	——	——
	M-G	是	218	19.762
	S-Q	否	——	——
	V-A	是	121	11.591
	Y-L	是	239	15.772
2023 年 4 月 29 日	D-R	是	167	18.218
	J-K	是	135	1.870
	Q-O	否	——	——
	U-O	否	——	——
	Y-W	是	142	4.929

注：考虑实际情况，已对快递运输数量数据进行取整处理。

由上表可知，每个城市对之间的灰色预测相对误差均小于 20%，说明模型拟合效果良好。

6.4 问题四的模型建立与求解

6.4.1 模型建立

（一）目标函数

由题设可知，需要分别求出每一天的最小运费量，对于单独的一天，目标函数可表示为：

$$\min z = \sum_i c_i (i = 1, 2, 3) \quad (29)$$

其中 c_i 表示该组城市对的运输成本， i 表示当天需要快递运输的城市对总数。

（二）模型建立

1、每个城市对之间，尽可能选择较短的路径数：

$$n_{path} = \left\lfloor \frac{m_{package}}{200} \right\rfloor \quad (30)$$

式中， n_{path} 为该组城市对之间需求的路径数； $m_{package}$ 表示该城市对需要快递运输的数量。

2、求解出结点间的多个最短路径

颜氏算法最短路径是 Dijkstra 的一个扩展，允许找到多条最短路径，即前 k 条最短路径。适用于具有非负边权重的有向无环图。利用任意最短路径算法（如 Dijkstra 算法）找到最优路径，然后再找到 $k - 1$ 条与最优路径不同的次优路径。具体步骤如下：

（1）初始化一个候选路径集合 B 和一个结果路径集合 A ，使用 Dijkstra 算法计算起点到终点的最短路径 A^1 ，并将其加入 A 中；

（2）对于 A 中的每条路径 A^k ，找出它所有可能的偏离路径 A_i^k ，并将它们加入 B 中。偏离路径是指从 A^k 中的第 i 个节点开始，沿着不同于 A^k 的边走到终点的路径；

(3)从 B 中选出最短的偏离路径,并将其加入 A 中。如果 B 为空或者 A 中已经有 k 条路径,则停止算法, 否则重复步骤 2 和 3。

3、计算每一组“发货-收货”的站点城市对运输成本

4、对每一天的所有城市对运输成本求和, 得到总成本

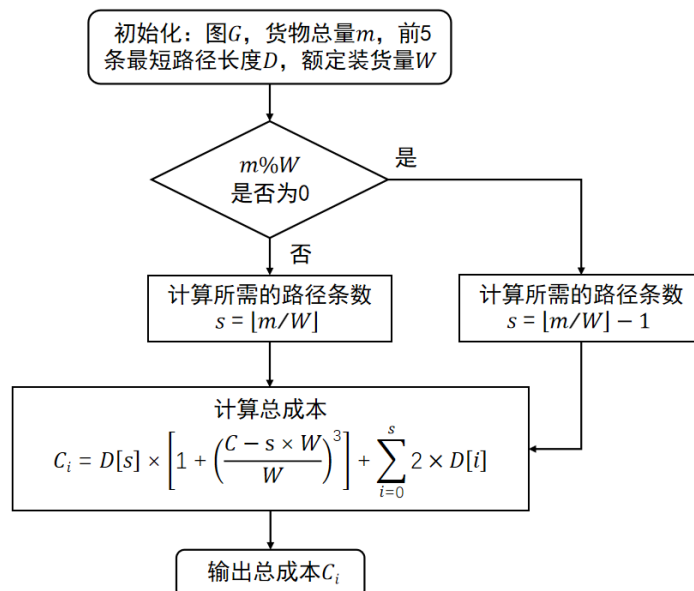


图3 求解出结点间的多个最短路径流程图

6.4.2 模型结果

1、G-V 城市对间的最短路径展示

G-V 的五天运输量分别为: 616、737、728、606、601。考虑模型的现实意义, 为类似运货量远大于 200 的城市对分配多条运输路线, 有很强的经济效益, 以下为 G-V 的多条成本最小路径展示:

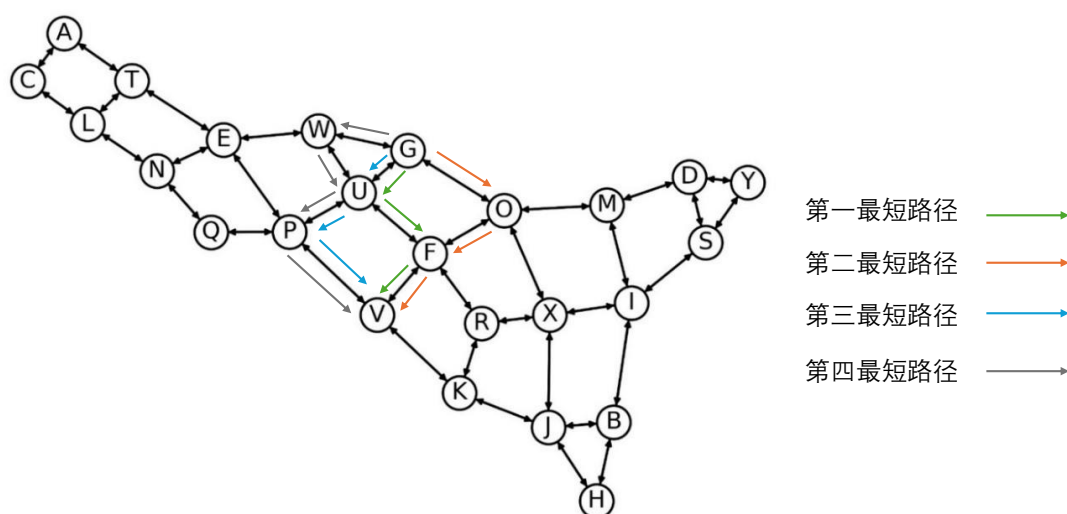


图4 G-V 的多条成本最小路径

表 7 G-V 的多条成本最小路径所经过路径及固定成本和

经过的结点	路线的固定成本之和
G→U→F→V	3.6
G→U→P→V	4.4
G→O→F→V	4.8
G→W→U→P→V	6.2
G→W→E→P→V	7.6

2、部分城市链条费用成本

由于篇幅限制，以下为城市对 CN、DL、GV 三条单链每一天的最终运输成本，分别代表低运输量、中等运输量，高运输量的费用值。其余城市对的结果将在附件中展示。

表 8 指定城市对每天的最终运输成本

城市对	2023-04-23	2023-04-24	2023-04-25	2023-04-26	2023-04-27
CN	4.54	4.38	3.78	3.64	3.63
DL	15.86	15.84	15.78	15.76	15.74
GV	31.80	33.79	33.43	31.80	31.80

3、大量货运优化前后对比

部分城市结点对之间的运货数量远大于 200，考虑其成本成幂函数上升，必须选择多条路线完成该次运输，则可大幅度降低成本。下表为 GV 城市结点对仅选择最小固定成本的路线运输，和选择多条最短路线运输的最终成本对比图。大幅的成本下降，说明对多条路线进行运输的合理性。

表 9 指定城市对优化后成本

城市结点对	日期	运货数量	单条路线运输成本	多条路线运输成本
GV	4.24	737	183.74	33.79
GV	4.25	728	177.22	33.43
GV	4.23	616	108.79	31.80

4、模型的最终结果

表 10 模型最终结果

日期	最低运输成本
2023 年 4 月 23 日	1110.89
2023 年 4 月 24 日	1166.22
2023 年 4 月 25 日	1189.88
2023 年 4 月 26 日	1033.78
2023 年 4 月 27 日	1399.40

6.4.3 模型的稳定性

对于每个城市间铁路运输网络,实际装货量的最大值为 200 的基础上,给予每条路线 5% 的微小波动,比较最终成本与之前成本的差异,最终成本差异曲线图如下图所示,最终成本相差不大,证明该优化模型具有较好的稳定性。

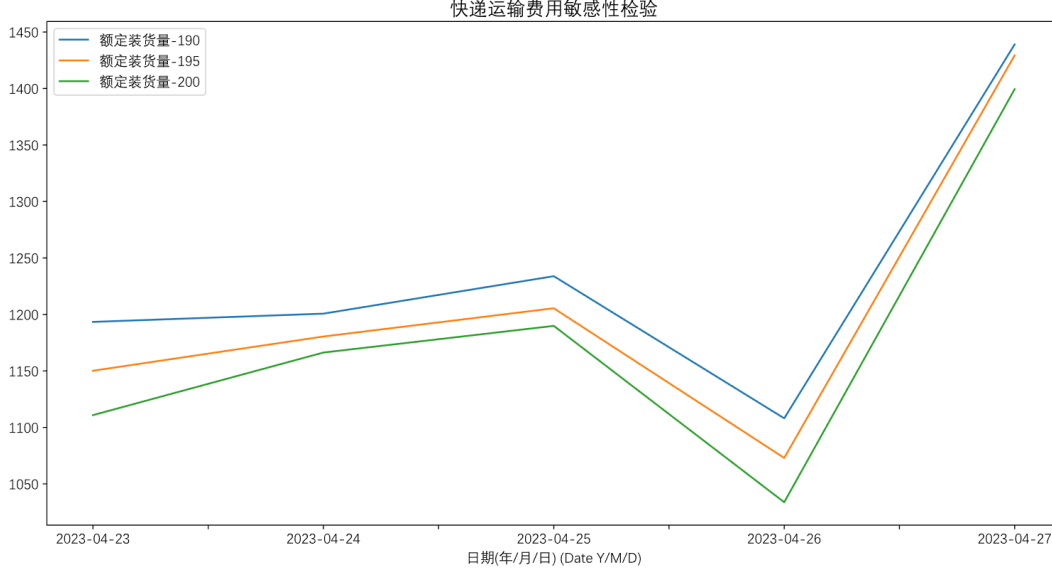


图 5 成本差异曲线图

6.5 问题五的模型建立与求解

6.5.1 按季度估计固定需求常数

6.5.1.1 求解固定需求常数

由频率直方图及文献分析知固定需求常数知其特性如下：固定需求常数是在同一季度内确定的，因此不随时间变化而变化且具有稳定性；固定需求常数是基于历史数据进行估计得出的，不受外部因素的影响；固定需求常数通常是在整个需求中所占比例较小的一部分，因此对预测结果的影响相对较小，基于此其对后续估计非固定需求影响不大。

《量化投资：策略与技术》^[4]提出对于一个时间序列，可将其分解为固定需求、季节性需求和随机波动三个部分。其中，固定需求是时间不变的、长期的趋势部分，季节性需求是周期性的波动部分，随机波动是不可预测的随机部分。

文中指出为估计固定需求的大小，可采用 $\mu - 2\sigma$ 的方法，即计算序列的均值和标准差，然后将均值加减两倍标准差得到一个大致区间，该区间内的数值可以认为是固定需求的可能取值范围。具体来说，如果序列的均值为 μ ，标准差为 σ ，则固定需求的大致区间为 $[\mu - 2\sigma, \mu + 2\sigma]$ ，且其通常小于需求的最小值。综合考虑其特性，我们以每季度快递运输量的均值和标准差拟合固定化需求，具体公式如下：

$$u_{jk} = \frac{\sum_k \sum_j x_{jk}}{k_{max}} \quad (31)$$

$$\sigma_{jk} = \sqrt{\frac{\sum_k \sum_j (x_{jk} - u_{jk})^2}{k_{max}}} \quad (32)$$

$$x_{jkNF} = u_{jk} - 2\sigma_{jk} \quad (33)$$

式中， u_{jk} 是第 j 条线路第 k 个季度的均值； σ_{jk} 是第 j 条线路第 k 个季度的标准差； x_{jkNF} 是第 j 条线路第 k 个季度的固定需求常数。若计算结果小于 0，则规定其等于 0。

6.5.1.2 验证固定需求常数准确性

(一) 和表格中每季度需求最小值对比发现，固定需求常数大部分小于该阶段历史数据的最小值，满足题意要求，具有一定的准确性。

(二) 根据 XGBoost 回溯预测

固定需求源于日常的必要网购消费，根据题意会呈现明显的季节相关性，我们对计算出的固定需求常数，构建 XGBoost 对该数据所在年季度进行分类检验，训练后混淆矩阵热力图如下：

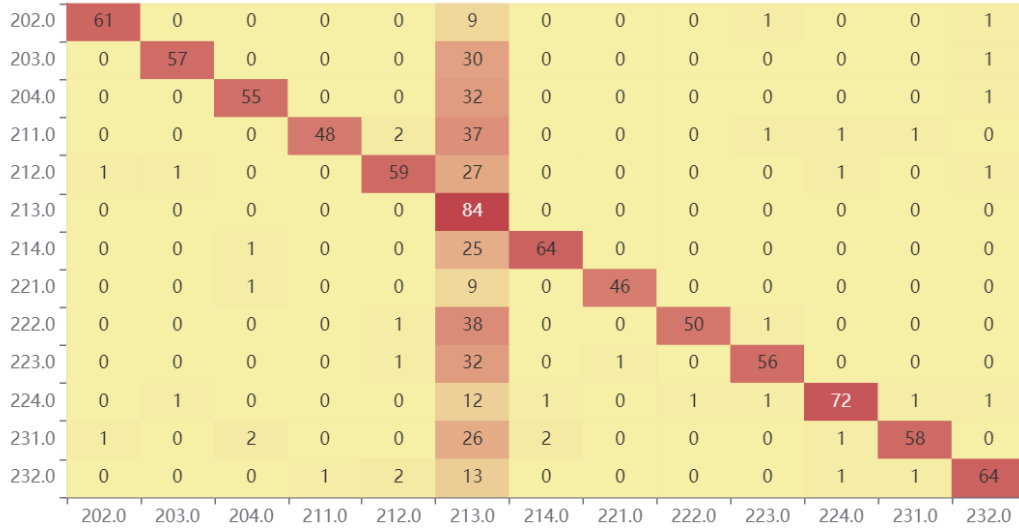


图 6 XGBoost 回溯预测得出的混淆矩阵热力图

由预测检验结果可知，除 2021 年第 3 季度外，其余分类结果较好，说明季节固定需求常数的估计具有较强的现实意义，具有较高的准确性。推测 2021 年第 3 季度，城市站点受新冠疫情影响，没有表现出明显的特征。

6.5.2 利用核密度估计法对非固定需求常数计算

6.5.2.1 非固定需求成本的计算

问题 5 的背景中可知，非固定需求常数的计算方法为：

$$x_{jkRF} = x_{jk} - x_{jkNF} \quad (34)$$

即非固定需求常数为实际值与固定需求常数之差。

6.5.2.2 基于核密度法估计非固定需求的概率分布

核密度估计是一种非参数估计方法，用于估计一组数据的概率密度函数。其基本思想是将每个数据点作为一条正态分布曲线的中心，然后将这些曲线相加形成一条平滑的概率密度曲线。

和参数估计不同，非参数估计并不加入任何先验知识，而是根据数据本身的特点、性质来拟合分布，这样能比参数估计方法得出更好的模型。高斯核函数的特点是具有连续性、平滑性和对称性，适用于大多数数据集，本文通过高斯核函数构建核密度估计模型，该函数形式如下：

$$K(x) = \frac{1}{\sqrt{2\pi}h} e^{-\frac{x^2}{2h^2}} \quad (35)$$

式中， x 是待估计的数据点， h 是带宽参数， $K(x)$ 是该点处的估计概率密度。根据公式，求解出指定季度，指定城市对的核密度估计的概率分布图像。

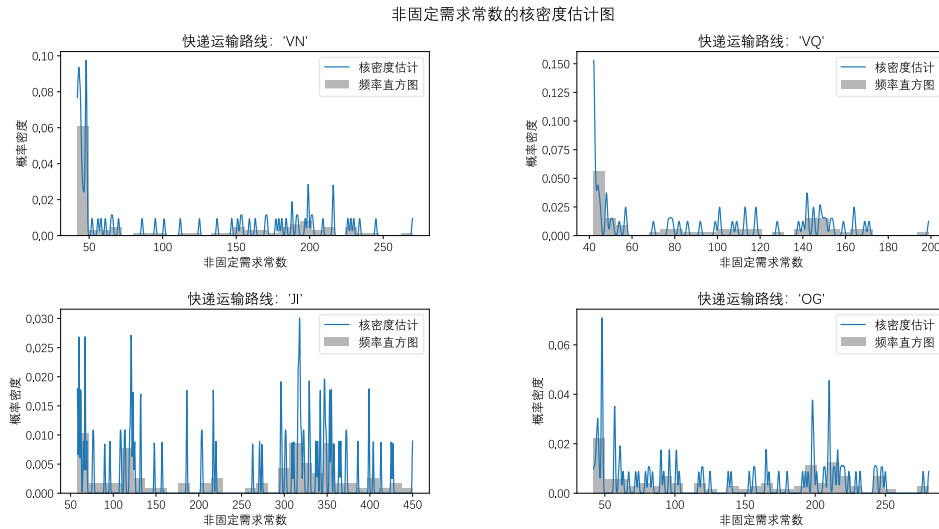


图 7 指定城市对的核密度估计的概率分布图像

6.5.3 表格数据的汇总

综合以上计算，问题 5 的结果如下表所示：

表 11 问题 5 结果表

季度	2022 年第三季度 (7~9 月)		2023 年第一季度 (1~3 月)	
“发货-收货” 站点城市对	V-N	V-Q	J-I	O-G
固定需求常数	0	0	3	0
非固定需求均值	107.93	93.27	242.40	141.01
非固定需求标准差	72.94	47.91	121.23	74.31
固定需求常数总和	1589.22		1741.62	
非固定需求均值总和	6376.60		4272.13	
非固定需求标准差总和	5557.32		4907.08	

针对上述求解结果可知，2023 年第一季度的固定需求常数总和大于 2022 年第三季度，说明我国近期以来快递需求数量有增加趋势，网络购物在生活中的角色愈发重要。

2023 年第一季度的非固定需求的均值总和与标准差均小于 2022 年第三季度，说明在新冠疫情形势趋于平稳后，快递需求数受社会因素的影响逐渐变小。

七、模型的评价、改进与推广

7.1 模型的优点

- 1、本文没有使用复杂的或可解释性差的模型，而是注重结果的可解释性和现实意义。
- 2、考虑了多类、多项指标，对站点城市的重要程度准确排序。
- 3、对于未来快递运输数量的预测，本文采用 ARIMA 时间序列和灰色预测模型，分别对应不同的场景，模型精度高。
- 4、目标优化模型中，充分利用城市对之间的铁路运输网络，大大降低运输成本。
- 5、查阅文献后建立固定需求常数的计算模型，理论依据强。

7.2 模型的缺点

- 1、未对无发货需求的异常值其真实情况做进一步推测。
- 2、模型对样本的依赖性较大，权重会随着样本的变化而发生改变。

7.3 模型的改进与推广

- 1、进一步查阅城市代号所对应的具体城市名称，推广此模型，提升我国物流业的社会经济效应。

八、参考文献

- [1]张文静、李鑫，“基于 ARIMA 模型的快递业季节性预测研究”，《物流技术》，2017 年第 23 期，pp. 55-58.
- [2]尹小峰，黄璐. 基于滚动优化的城市间物流运输装载优化研究[J]. 物流技术，2018,37(6): 100-102.
- [3]司守奎，姜启源. 数学建模算法与应用[M].3 版.国防工业出版社.
- [4] Ernest P. Chan, 《量化投资：策略与技术》（第二版），北京市：中国人民大学出版社，2017 年.