

# 基于ARIMA模型的上海港货物吞吐量和进出口总额预测分析

徐颖双\*, 胡文伟

上海工程技术大学, 上海

收稿日期: 2022年5月21日; 录用日期: 2022年6月11日; 发布日期: 2022年6月21日

## 摘要

选取上海市1978~2019年外贸进出口总额数据和港口货物吞吐量数据进行分析, 本文首先通过构造一元时间序列模型ARIMA(3,2,0)对上海市未来5年外贸进出口总额的变化趋势进行预测; 其次, 为分析上海市外贸进出口发展对港口货物吞吐量的影响, 本文先对二者进行协整检验, 并拟合协整动态回归模型和外贸进出口总额对数模型, 再根据进出口总额预测值去预测上海市未来5年港口货物吞吐量的趋势变化。预测结果显示, 未来5年的上海市进出口总额和港口货物吞吐量总体上都将呈上升趋势, 且通过对比真实值与预测值之间的偏差率发现模型预测的误差率较低, 这表明所拟合模型精度较高。

## 关键词

进出口外贸总额, 港口货物吞吐量, ARIMA模型, 预测

## Forecast and Analysis of Cargo Throughput and Total Import and Export Volume in Shanghai Port Based on ARIMA Model

Yingshuang Xu\*, Wenwei Hu

Shanghai University of Engineering Science, Shanghai

Received: May 21<sup>st</sup>, 2022; accepted: Jun. 11<sup>th</sup>, 2022; published: Jun. 21<sup>st</sup>, 2022

## Abstract

This paper analyzes the total import and export volume data and port cargo throughput data of  
\*通讯作者。

Shanghai from 1978 to 2019. Firstly, it forecasts the change trend of total import and export volume of Shanghai in the next five years by constructing unitary time series model ARIMA (3,2,0). Secondly, in order to analyze the impact of the development of Shanghai's foreign trade import and export on the port cargo throughput, this paper first carries out the co-integration test of the two, and fits the co-integration dynamic regression model and the logarithmic model of total foreign trade import and export, and then predicts the trend change of Shanghai's port cargo throughput in the next five years according to the predicted value of total import and export. The forecast results show that in the next five years, the total import and export volume and port cargo throughput of Shanghai will be on the rise, and by comparing the deviation rate between the real value and the predicted value, it is found that the error rate of the model prediction is low, which indicates that the precision of the fitted model is high.

## Keywords

Total Import and Export Volume, Port Cargo Throughput, ARIMA Model, Forecast

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

自二十世纪八十年代改革开放以来,我国对外贸易经济得到了飞跃式发展,目前全国外贸进出口总额居世界第二位。而截至 2019 年上海市成为唯一一个进出口超过 3 万亿的城市,以实现进出口总额 34046.8 亿排名第一,在带动我国整个进出口外贸发展浪潮中占据着重要地位。因此,对上海市进出口贸易进行研究并对进出口总额进行预测,这对分析我国未来几年的对外贸易发展起着重要作用。此外,外贸进出口货物的重要运输方式除陆路运输之外,水路运输同样占据着重要地位,因此进出口外贸的快速发展将会对城市港口货物吞吐量起着极大的提升作用,促进了城市港口繁荣发展。

现有研究中,大多只是单独预测外贸进出口总额[1] [2] [3]或港口货物吞吐量[4] [5] [6],而本文通过协整检验验证两者之间的关系,根据所预测的进出口总额,结合所拟合的协整模型来预测未来 5 年上海港货物吞吐量。结果显示,两者之间确实存在协整关系,因而先对未来 5 年上海市进出口总额进行预测,同时根据已有数据计算了预测误差率,结果显示所拟合的 ARIMA(3,2,0)拟合效果较好。因此在此基础上对上海港货物吞吐量进行预测,预测结果显示未来 5 年上海货物吞吐量将随之进出口总额快速发展而呈上升趋势,同时所计算出的预测误差率较小,这表明我们所拟合的模型精准度较高。这有助于我们更好地了解上海市未来进出口总额和港口货物吞吐量发展趋势,并且本文验证了外贸进出口的发展与港口货物吞吐量之间存在的协整关系,这能为政府预测未来港口货物吞吐量发展提供一定参考意见。

## 2. ARIMA 模型介绍

ARIMA 模型最早是由 Box 和 Jenkins 于上世纪 70 年代所提出,该模型也通常被称为求和自回归移动平均模型[7]。ARIMA 模型实质上就是差分运算和 ARMA 模型的组合,即通过差分运算首先将非平稳序列平稳化,再用平稳化序列拟合自回归移动平均模型(ARMA) [8] [9]。而 ARMA 分为 AR(自回归模型)、MA(滑动平均模型)和 ARMA(自回归滑动平均模型)这三类模型[10]。

AR 模型描述的是当前值和历史值之间的关系,其表达式为  $x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t$ ,

其中,  $p$  为自回归模型的阶数,  $\phi$  为待定系数,  $\varepsilon_t$  为误差,  $x_t$  为平稳时间序列。MA 模型是根据所研究的对象过去所产生的干扰值与现在所产生的干扰值之间的一种线性关系组合进行预测[7], 其基本表达式为  $x_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}$ , 其中,  $q$  是为该模型的阶数,  $\theta_q$  为待定系数,  $\varepsilon_{t-q}$  是  $t$  之前  $q$  期的随机干扰值。ARMA 模型 AR 和 MA 模型的组合, 其表达式为

$x_t = \phi_0 + \phi_1x_{t-1} + \phi_2x_{t-2} + \dots + \phi_px_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}$ , 其中  $p$ 、 $q$ 、 $\phi_p$ 、 $\varepsilon_t$ 、 $\theta_q$ 、 $\varepsilon_{t-q}$  与前述的 AR、MA 模型中含义相同。

### 3. 实证分析

#### 3.1. 数据预处理与初步分析

为研究上海市近年来的外贸进出口总额的波动特征与规律, 本文选取了 1978~2019 年的年度数据, 共计 42 个观测值。下列图 1 是上海市年进出口总额的时序图, 我们从中可以较为清晰地发现, 该序列蕴含着比较明显的线性递增的变化趋势, 处于非平稳状态。因此, 我们通过对该序列进行二阶差分运算来从中提取出线性趋势信息, 提取结果如图 2 所示, 可以发现该序列在经过二阶差分处理后呈现出较为平稳的波动趋势特征, 这就表明该二阶差分处理过程比较准确充分地提取出了在原序列中可能蕴含着的某种长期趋势, 使得经过处理后的序列不再继续呈现出确定性趋势。

同时, 本文为研究上海市外贸进出口发展对港口货物吞吐量的影响, 选取了 1978~2019 年间关于上海市港口货物吞吐量的数据, 总计为 42 个观测值。下列图 3 是上海市进出口外贸总额与上海市港口货物吞吐量对数序列时序图。时序图显示, 这两个序列都呈现出显著的线性递增变化的时间趋势, 所以两者都是非平稳的时间序列。但是这两个序列之间又随着时间的逐步增长, 这两序列之间的变化趋势逐渐趋同, 所以为了有效且客观地衡量这两个时间序列之间是否存在长期稳定的相对均衡关系, 本文首先通过构造回归模型检验序列之间是否具有协整关系。

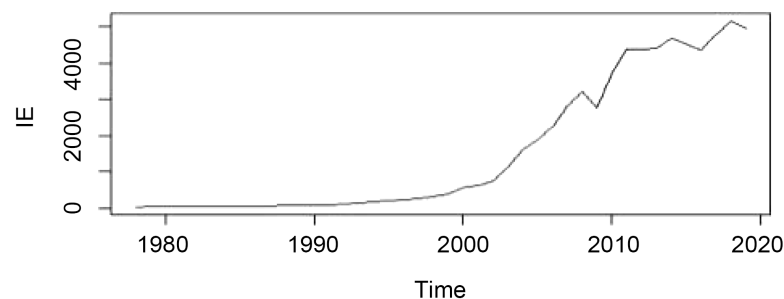


Figure 1. Sequence plot of original data

图 1. 原始数据序列图

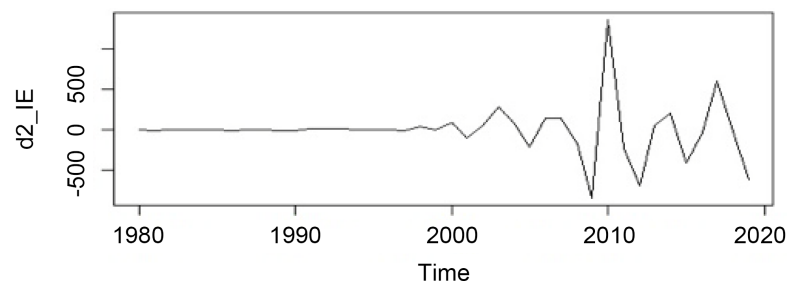
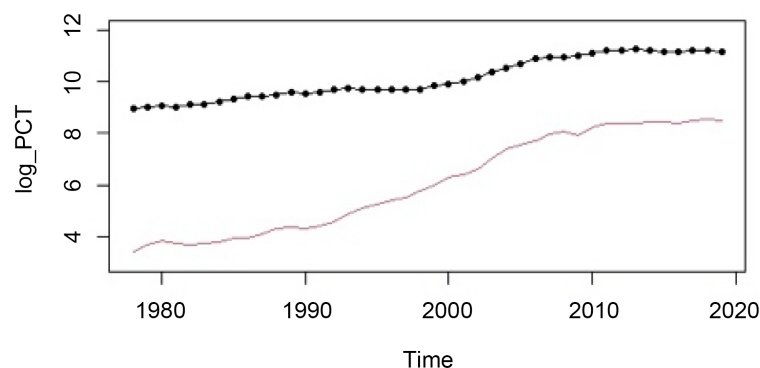


Figure 2. Timing plot after second order difference

图 2. 二阶差分后时序图



**Figure 3.** Timing plot of logarithmic sequence of total import and export volume and port cargo throughput

**图 3.** 进出口总额与港口货物吞吐量对数序列时序图

## 3.2. 序列检验

### 3.2.1. 平稳性检验和白噪声检验

首先通过观察二阶差分后时序图可以初步判断二阶差分后得到的上海市外贸进出口总额序列为平稳, 需要进一步借助 ADF 检验, 判断该序列的平稳性。ADF 检验结果显示: 类型一、类型二和类型三各种模型的  $\tau$  统计量的 P 值均小于显著性水平 ( $\alpha = 0.05$ ), 所以认为该序列显著平稳。

其次, 并对已经进行了二阶差分运算的上海市外贸进出口总额序列进行白噪声检验, 得到如表 1 所示的检验结果。纯随机性检验得到的统计分析结果显示: 在延迟 6 阶和 12 阶下 LB 检验统计量的 P 值都非常小, 低于所给定的显著性水平 0.05, 因而能在 95% 的置信水平下将原有假设拒绝。故可以判断上海市外贸进出口总额序列不属于白噪声序列。

**Table 1.** White noise test results of total foreign trade import and export

**表 1.** 外贸进出口总额序列白噪声检验结果

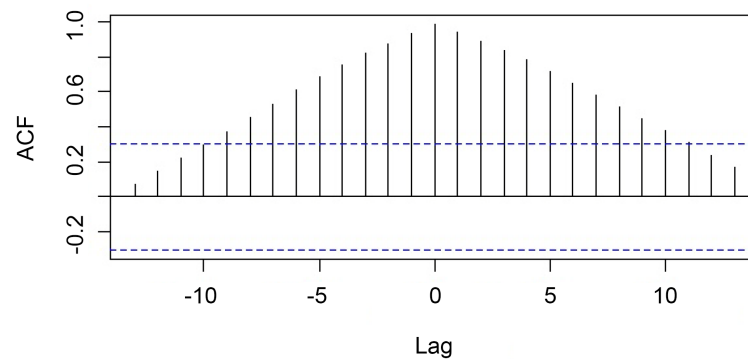
延迟阶数	LB 统计量检验	
	LB 统计量的值	P 值
6	29.824	4.245e-05
12	48.761	2.303e-06

### 3.2.2. 协整检验

为对上海市外贸进出口总额和港口货物吞吐量建立动态回归模型, 先对其进行协整检验。图 4 为上海市外贸进出口对数序列和港口货物吞吐量对数序列的互相关图, 图中显示外贸进出口发展对港口货物吞吐量的影响当期达到最大。因此, 利用最小二乘法, 构建二者之间的回归模型为:

$$\log y_t = 1.5634 \log x_t + \varepsilon_t$$

其中,  $y_t$  表示  $t$  年上海市港口货物吞吐量,  $x_t$  表示  $t$  年上海市外贸进出口总额,  $\varepsilon_t$  表示残差。然后对回归残差序列进行平稳性检验。根据类型一滞后 0 阶的检验结果, 本文有近似 92.5% 的把握认为上海市外贸进出口总额对数序列与港口货物吞吐量对数序列之间存在协整关系。并且我们还同时对回归的残差序列进行了白噪声检验, 检验结果显示, 回归残差序列不是白噪声序列, 需要进一步提取这些回归的残差序列中可能会蕴含着的相关信息。



**Figure 4.** Cross-correlation plot of logarithmic sequence of total import and export value and port cargo throughput

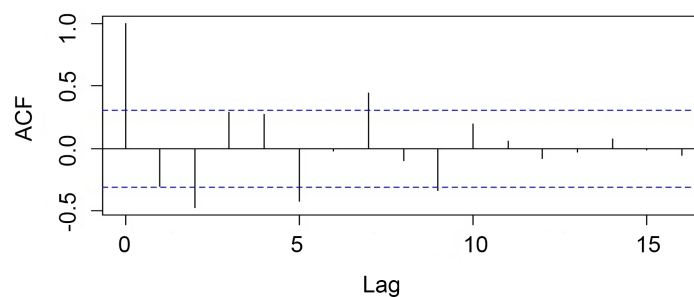
**图 4.** 进出口总额与港口吞吐量对数序列互相关图

### 3.3. 模型选择与显著性检验

#### 3.3.1. 模型识别

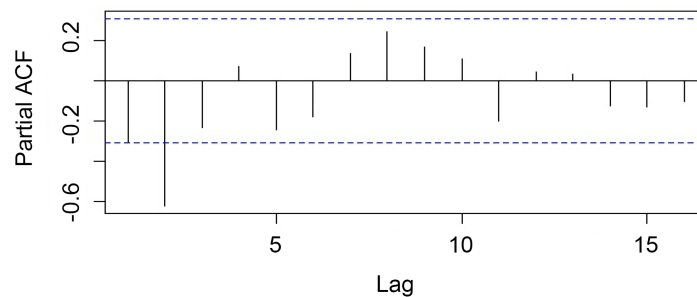
##### 1) 一元时间序列模型

二阶差分后的序列是一个平稳的非白噪声序列, 即  $ARIMA(p,d,q)$  模型中的  $d=2$ 。尚未能确定模型中的  $p$  和  $q$ , 可以通过观察二阶差分后的时序图中自相关系数和偏自相关系数的截尾性来判断[6]。根据图 5 的自相关图所示, 首先可以看见自相关系数表现出较为连续地从顶峰逐渐趋于零的趋势, 这表明自相关系数是拖尾。其次, 根据图 6 的偏自相关系数的变化过程, 可以看出除 1 至 2 阶的偏自相关系数波动范围在 2 倍的标准差之外, 其余所有阶数的波动都在 2 倍的标准差单位之内, 这表明偏自相关系数 2 阶或 3 阶截尾, 所以本文首先将模型定为  $ARIMA(3,2,0)$ , 并对其进行检验。



**Figure 5.** Autocorrelation plot of sequence after second-order difference

**图 5.** 二阶差分后序列自相关图



**Figure 6.** Partial autocorrelation plot of sequence after second-order difference

**图 6.** 二阶差分后序列偏自相关图

## 2) 协整动态回归模型

观察回归残差序列的自相关图(图 7)和偏自相关图(图 8), 可以判断残差序列自回归系数拖尾, 偏自相关系数 1 阶截尾。因此, 本文将残差模型定为 AR(1), 并对其进行检验。

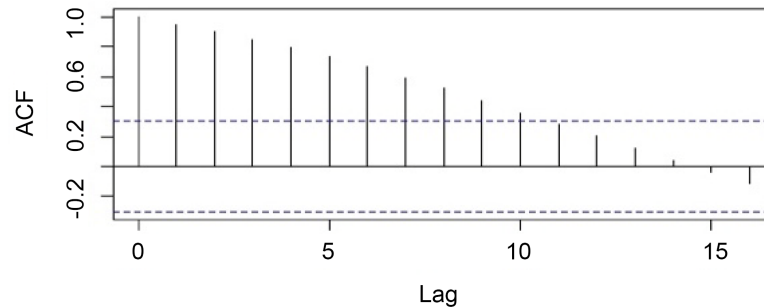


Figure 7. Autocorrelation plot of residual series

图 7. 残差序列自相关图

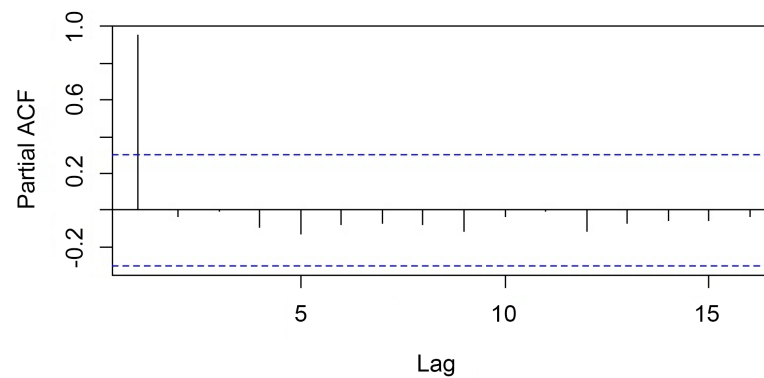


Figure 8. Partial autocorrelation plot of residual series

图 8. 残差序列偏自相关图

### 3.3.2. 模型拟合与显著性检验

#### 1) 一元时间序列模型

经过上述验证, 运用 R 软件拟合模型 ARIMA(3,2,0), 得到如表 2 所示各参数估计值。

Table 2. Automatic identification of the estimated values of each parameter of the model

表 2. 自动识别模型的各项参数估计值

系数	ar1	ar2	ar3
-	-0.7620	-0.8504	-0.3166
s.e.	0.1516	0.1364	0.1519
sigma <sup>2</sup> estimated as 50264: log likelihood = -272.53 AIC = 553.05 AICc = 554.2 BIC = 559.81			

由上表各参数可以得到拟合模型 ARIMA(3,2,0), 且该模型是基于二阶差分后的数据得来的:

$$z_t = -0.762z_{t-1} - 0.8504z_{t-2} - 0.3166z_{t-3} + \varepsilon_t$$

#### ① 残差白噪声检验

对上文所建的模型进行残差白噪声检验, 选用 Box-Ljung 检验统计量, 原假设  $H_0: \rho_1 = \rho_2 = \dots = \rho_m$ ,

$\forall m \geq 1$ , 备择假设  $H1$ : 至少存在某个  $\rho_k \neq 0, k \leq m$ 。如果仍然能够有效拒绝原假设, 就说明在残差序列模型中还可能残存着某些相关信息, 拟合序列的模型尚不显著; 假若检验结果显示不能够拒绝原假设, 就可以认为该拟合的序列模型是显著且有效的。根据图 9 所显示的检验结果来看, 各阶段延迟条件下的白噪声检验统计量中的 P 值都很显著的大于 0.05, 因此不能拒绝原假设。所以, 我们可以认为在这个拟合模型中的残差序列属于白噪声序列, 即表明所拟合的模型显著成立且有效。

如

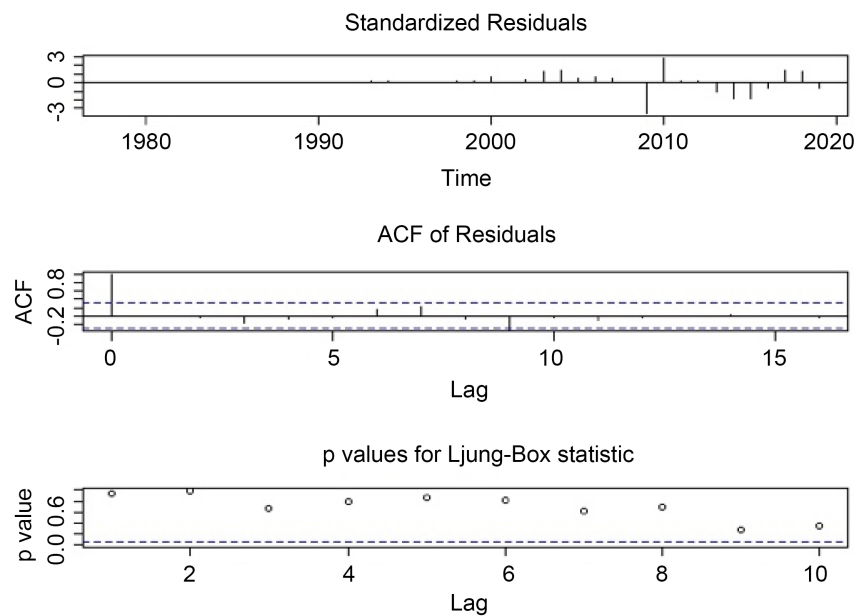


Figure 9. White noise test plot of residual

图 9. 残差白噪声检验图

## ② 参数显著性检验

参数显著性检验主要是通过检验模型中的每一个未知参数是否都显著非零, 从而尽可能使该模型变得精简。如果发现某一个变量参数检验结果是不显著非零, 这表示与该变量参数所对应的那个自变量对模型中的因变量的影响不明显, 二者无显著关系, 所以该自变量也就可以从所拟合的模型参数重一一剔除, 最终模型将由这样一系列参数显著且非零的自变量构成。我们通过对 ARIMA(3,2,0)模型中的三个参数进行了显著性检验, 结果如表 3 所示: ar1、ar2、ar3 这三个参数检验统计量的 P 值都显著小于 0.05, 所以可以判断拟合模型的三参数都显著非零。

Table 3. Parameter significance test

表 3. 参数显著性检验

检验参数	T 值	P 值	结论
ar1	-5.0264	5.761601e-06	显著非 0
ar2	-6.2346	1.222951e-07	显著非 0
ar3	-2.0843	0.02186691	显著非 0

## 2) 协整动态回归模型

经上述验证, 运用 R 软件拟合模型 AR(1), 得到如表 4 所示各参数结果。

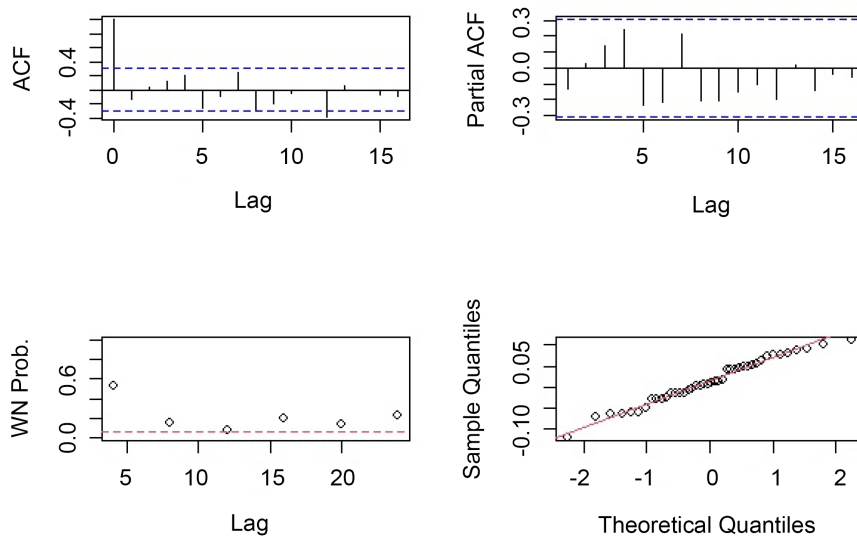
**Table 4.** Estimated values of parameters of the cointegration model  
**表 4.** 协整模型各参数估计值

系数	ar1	logIE
-	0.4212	0.2568
s.e.	0.1896	0.0864
Sigma^2 estimated as 0.003097: log likelihood = 60.16, AIC = -114.32		

由上表参数可以得到拟合的协整动态回归模型结构如下:

$$\log y_t = 0.2568 \log x_t + \frac{\varepsilon_t}{1-0.4212B} \quad \varepsilon_t \sim N(0, 0.0031)$$

这两个参数的估计值均显著大于 2 标准差, 因此两个参数均显著非零。并据此对拟合的回归模型进行显著性检验, 如图 10 显示, 残差序列为白噪声序列, 这表明拟合的协整动态回归模型显著成立。



**Figure 10.** White noise test plot of Residual series  
**图 10.** 残差序列白噪声检验图

为了更加准确地对港口货物吞吐量进行预测, 本文将先对上海市未来 5 年的外贸进出口总额进行预测, 然后将该预测值代入协整模型, 从而得到未来 5 年上海市港口货物吞吐量的预测值。因此, 根据上海市外贸进出口总额对数序列具有的特征, 本文对其拟合了 ARIMA(1,1,0)模型。运用 R 软件拟合模型所得参数如表 5 所示。

**Table 5.** Estimated value of each parameter of import and export logarithmic series fitting model  
**表 5.** 进出口额对数序列拟合模型各参数估计值

系数	ar1	Intercept
-	0.3828	0.1236
s.e.	0.1469	0.0291
Sigma^2 estimated as 0.01363: log likelihood = 29.81, AIC = -53.61		

由上表参数可以得到上海市外贸进出口总额对数拟合模型如下:



$$\nabla \log x_t = 0.1236 + \frac{\varepsilon_t}{1 - 0.3828B}$$

### 3.4. 序列预测

#### 3.4.1. 外贸进出口总额预测

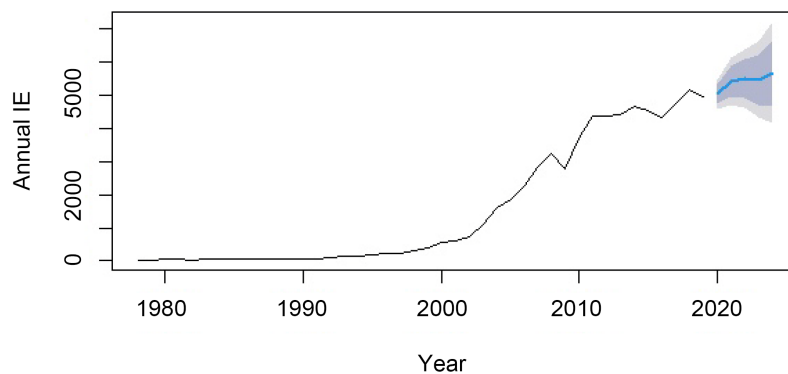
根据前文的分析, 建立 ARIMA(3,2,0)模型, 预测未来 5 年上海市外贸进出口总额, 结果如图 11 所示。根据 ARIMA(3,2,0)模型预测的趋势图我们可以较为清晰看出, 上海市未来 5 年的外贸进出口总额将保持一个上升趋势, 具体的预测分析结果如表 6 所示。此外, 根据可获取的 2020 年进出口总额, 计算出预测误差率。可以发现, 预测误差率仅为-0.23%, 这表明预测精准度相对较高。

表 6 显示了上海市于 2020~2024 年期间的外贸进出口总额的预测数据, 及在 95%以上置信度前提下预测区间的上下限。从预测的结果上来看, 未来五年上海市外贸进出口总额将从 2020 年的 5020 亿美元上升至 2024 年的 5673 亿美元。

**Table 6.** The forecasted volume of foreign trade import and export for next five years

**表 6.** 外贸进出口总额向后 5 期预测值

年份	预测值	下限(95%)	上限(95%)	真实值	误差率
2020	5020	4581	5460	5032	-0.23%
2021	5404	4705	6103	-	-
2022	5497	4619	6374	-	-
2023	5460	4304	6615	-	-
2024	5673	4144	7202	-	-



**Figure 11.** Forecast chart of Shanghai's total foreign trade import and export in the next five years

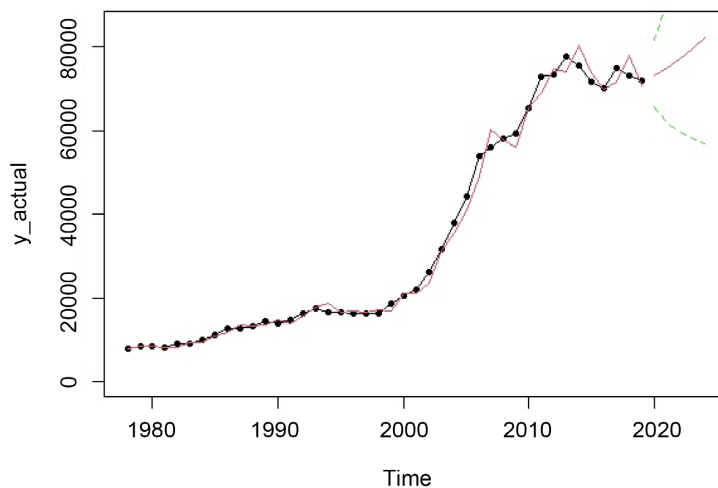
**图 11.** 未来 5 年上海市外贸进出口总额预测图

#### 3.4.2. 港口货物吞吐量预测

根据前文拟合的上海市外贸进出口总额序列模型和协整模型对未来连续 5 年上海市港口货物吞吐量进行粗略地预测, 预测结果大致如表 7 所示。表 7 显示了有关上海市 2020~2024 年的港口货物吞吐量的预测数据, 及在满足 95%以上置信度前提下该预测结果区间数据的上下限。从预测结果上来看, 上海市未来 5 年的港口货物吞吐量将从 2020 年的 73055 万吨上升至 2024 年的 82052 万吨, 从预测图 12 来看呈上升趋势。此外, 通过对比 2020 年和 2021 年的真实值与预测值之间的预测误差率, 可以发现, 预测误差率总体控制在 5%以下, 这表明该模型拟合优度较好。

**Table 7.** The forecasted port cargo throughput for next five years  
**表 7.** 港口货物吞吐量向后预测 5 期

年份	预测值	下限(95%)	上限(95%)	真实值	误差率
2020	73055	65505	81474	71671	1.89%
2021	74900	61967	90533	77635	-3.65%
2022	77113	59604	99767	-	-
2023	79520	57993	109037	-	-
2024	82052	56882	118361	-	-



**Figure 12.** Forecast chart of Shanghai's port cargo throughput in the next five years

**图 12.** 未来 5 年上海市港口货物吞吐量预测图

#### 4. 结论

上海市外贸进出口总额自 1978 年以来一直保持着长期逐年上升趋势,说明正由于国家改革开放之后要求中国加强了对外经济贸易交流及贸易洽谈活动,而使得上海市作为国际金融中心城市之一,在对外开放经济中发挥的辐射作用越来越重要,对外贸易发展迅速,进出口总额增长迅猛。因此该序列处于非平稳状态,由此通过二阶差分后使其平稳化,并建立短期预测模型。通过对上海市未来 5 年外贸进出口总额的预测发现,依旧呈现上涨趋势,这符合上海市未来发展势态。

而我们通过协整检验则发现,上海市外贸进出口总额的对数序列和港口货物吞吐量的对数序列之间存在着协整关系,这表明外贸进出口的快速发展,能够带动城市港口运输的繁荣。检验分析表明,上海市外贸进出口的长期上升发展带动了港口货物吞吐量同样呈现上升发展趋势。

#### 参考文献

- [1] 王玉荣. ARIMA 模型在我国出口贸易预测中的应用[J]. 统计与决策, 2004(4): 33-34.
- [2] 沈汉溪, 林坚. 基于 ARIMA 模型的中国外贸进出口预测: 2006-2010[J]. 国际贸易问题, 2007(6): 24-26.
- [3] 陈蔚. 基于线性 ARIMA 与非线性 BP 神经网络组合模型的进出口贸易预测[J]. 统计与决策, 2015(22): 47-49.
- [4] 张家善, 林晓群. 基于马尔科夫链的港口吞吐量区间预测模型研究[J]. 数学的实践与认识, 2016, 46(15): 159-164.
- [5] 鲁渤, 杨显飞, 汪寿阳. 基于情境变动的港口吞吐量预测模型[J]. 管理评论, 2018, 30(1): 195-201.

- 
- [6] 黄跃华, 陈小龙, 王亚辉. 基于正弦和的 GM(1,1) 模型在港口吞吐量预测中的应用[J]. 上海海事大学学报, 2019, 40(3): 69-73.
- [7] Box, G. and Jenkins, G. (1976) *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco, 1-11.
- [8] 张莹, 谭艳春, 彭发定, 廖杏杰, 余昱昕. 基于 EEMD 和 ARIMA 的海温预测模型研究[J]. 海洋学研究, 2019, 37(1): 9-14.
- [9] 易丹辉. 时间序列分析: 方法与应用[M]. 北京: 中国人民大学出版社, 2011: 48-103.
- [10] Dilling, S. and Macvicar, B.J. (2017) Cleaning High-Frequency Velocity Profile Data with Autoregressive Moving Average (ARMA) Models. *Flow Measurement & Instrumentation*, **54**, 68-81.  
<https://doi.org/10.1016/j.flowmeasinst.2016.12.005>