

论文题目 全国快递业务量的时间序列分析与预测

# 目录

一、引言.....	3
二、基于 SAS 软件的全国快递业务量时间分析 .....	3
1.时间序列预处理.....	4
(1) 建立数据集.....	4
(2) 绘制时序图.....	5
(3) 平稳化处理——差分法.....	5
2.模型识别.....	7
(1) 平稳性检验.....	7
(2) 白噪声检验.....	7
3.模型定阶.....	8
4.模型的参数估计及检验.....	8
(1) 参数估计.....	8
(2) 模型检验.....	9
5.模型的预测.....	10
三、结论.....	11
四、模型的不足.....	11
附录 (SAS 程序) .....	11

# 全国快递业务量的时间序列分析与预测

——基于SAS软件

## 一、引言

近年来，随着“六一八”、“双十一”、“双十二”活动的火爆，我国快递业务量增长十分迅速，不断刷新历史记录，早已成为物流大国。相比于世界其他国家，中国的网购市场更加发达，巨大的网络零售市场意味着巨大的快递消费市场。

现如今，网购早已成为人们生活中不可分割的一部分，甚至连农村都通了物流、设了快递点。随着网购的继续渗透，人们对于快递和物流的需求只会越来越大。目前，我国人均快递量虽然早已超过世界水平，但主要是东部地区的人均快递量较高，中西部地区人均快递量水平相对较低，从人均快递量的结构来看，我国的快递市场仍然有成长的空间。

本文基于时间序列理论，以我国2010年11月至2020年11月的每月全国快递当期业务量数据作为时间序列，并进行作图分析、模型识别与模型拟合，在此基础上根据拟合得到的模型做出未来五个月我国快递业务量的预测，为我国快递行业未来的发展提供依据。

## 二、基于 SAS 软件的全国快递业务量时间分析

下面，取2010年11月至2020年11月的每月全国快递当期业务量数据作为时间序列，对其进行分析：

月	快递业务量	月	快递业务量	月	快递业务量	月	快递业务量
2010-11	23220	2013-06	72154.13	2016-01	215505.3	2018-08	410136
2010-12	26366.2	2013-07	74802.13	2016-02	124618.3	2018-09	447629.1
2011-01	25301.63	2013-08	75266.69	2016-03	236922.2	2018-10	469187.5
2011-02	17177.4	2013-09	83637.13	2016-04	237303.2	2018-11	586357
2011-03	28645.4	2013-10	85303.54	2016-05	253108.3	2018-12	541822.2
2011-04	28015	2013-11	108703.4	2016-06	257633.1	2019-01	452289.5
2011-05	29113.79	2013-12	106945.3	2016-07	249640.6	2019-02	275950.6
2011-06	30078.41	2014-01	88950	2016-08	252296.1	2019-03	486392.8
2011-07	29599.01	2014-02	68886.17	2016-09	282685.9	2019-04	491910.9
2011-08	31235.55	2014-03	102528.2	2016-10	302614.2	2019-05	523276.4
2011-09	32462.6	2014-04	105417.6	2016-11	376447.9	2019-06	546077.6

2011-10	33414.7	2014-05	111634.7	2016-12	339540	2019-07	524847.6
2011-11	39393.81	2014-06	112926.3	2017-01	221113.6	2019-08	530243.1
2011-12	42873.81	2014-07	111245.1	2017-02	234430.1	2019-09	559722.6
2012-01	27319.4	2014-08	114050.7	2017-03	303493.7	2019-10	575730.1
2012-02	35164.15	2014-09	122992	2017-04	298237	2019-11	712424.1
2012-03	41850.92	2014-10	128997.1	2017-05	334061.8	2019-12	673425.7
2012-04	40373.75	2014-11	164616.5	2017-06	340453.2	2020-01	378096.8
2012-05	48993.91	2014-12	163681.1	2017-07	320498.9	2020-02	276521.4
2012-06	45447.64	2015-01	144570.4	2017-08	326398.6	2020-03	598423.2
2012-07	46244.93	2015-02	81779.49	2017-09	360605.6	2020-04	649726.4
2012-08	46561.8	2015-03	142537.8	2017-10	374771.5	2020-05	738292.5
2012-09	51345.25	2015-04	151483.9	2017-11	471490.3	2020-06	746967.5
2012-10	52421.3	2015-05	161002.5	2017-12	420037.7	2020-07	693601.4
2012-11	67126.63	2015-06	164486.7	2018-01	398613.3	2020-08	723622.1
2012-12	65698.42	2015-07	163988.3	2018-02	198867.2	2020-09	809158.2
2013-01	68917.73	2015-08	169020.3	2018-03	394330.2	2020-10	823388
2013-02	35473.1	2015-09	191336.9	2018-04	375275.5	2020-11	972623.3
2013-03	67043.79	2015-10	194064.8	2018-05	417842.2		
2013-04	66249.37	2015-11	260537.8	2018-06	422881		
2013-05	74178.66	2015-12	241828	2018-07	408083.3		

（来源：中经网全国宏观月度数据）

# 1.时间序列预处理

## （1）建立数据集

将上述文本格式数据导入SAS的DIR数据库。导入后结果如下图所示（由于数据量较大，仅展示部分结果）：

表1 原始数据

Obs	month	express	time
1	2010-11	23220.00	1
2	2010-12	26366.20	2
3	2011-01	25301.63	3
4	2011-02	17177.40	4
5	2011-03	28645.40	5
6	2011-04	28015.00	6
7	2011-05	29113.79	7
8	2011-06	30078.41	8
9	2011-07	29599.01	9
10	2011-08	31235.55	10
11	2011-09	32462.60	11
12	2011-10	33414.70	12

## (2) 绘制时序图

首先绘制全国快递业务量的原始数据的时间序列图，通过输入代码实现：

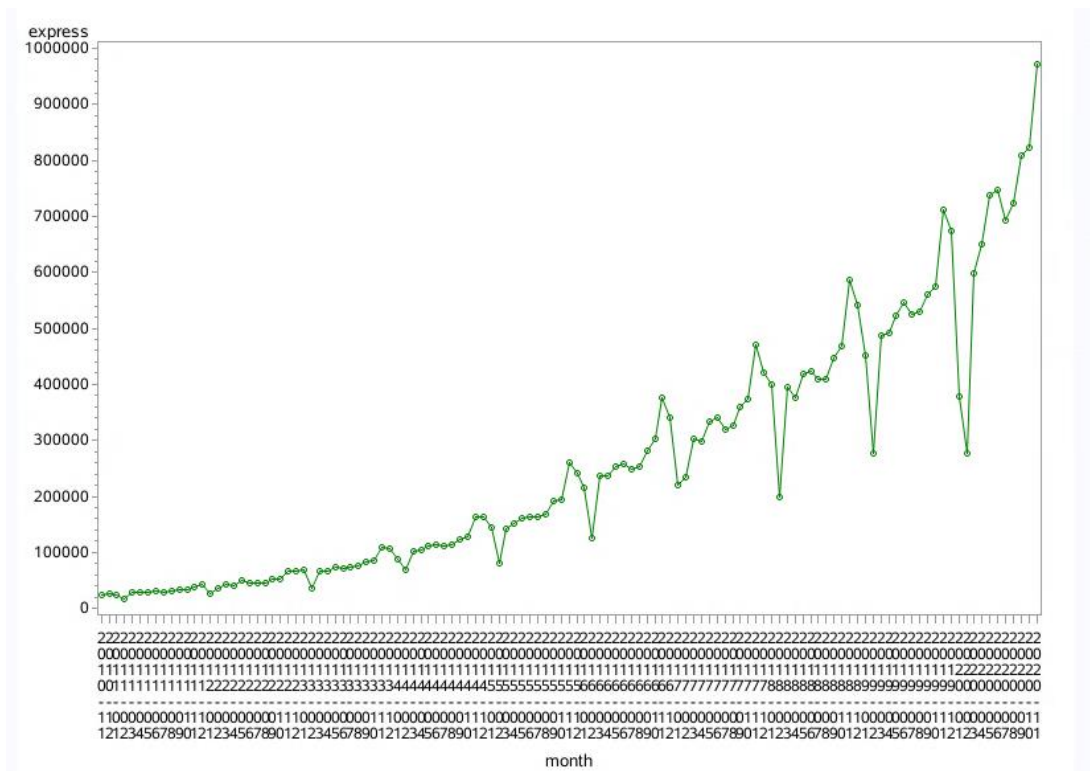


图1 原始数据时序图

从上图可以看出，全国快递业务量总体呈逐年上升趋势，但每年中的月度快递业务量都会有较大波动，且随着年度快递业务量总体的上升，月度之间的波动也越来越大。总体来看，原始序列并非平稳的时间序列，因此我们需要对其进行时间序列的平稳化处理。

### (3) 平稳化处理——差分法

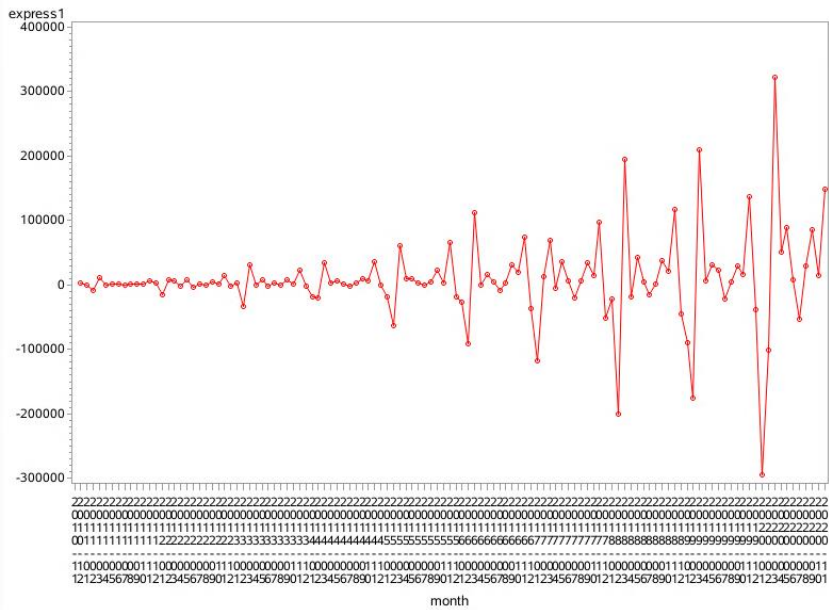
由于该时间序列并非呈指数分布,所以选择使用差分法对序列进行平稳化处理。由差分的选择可知,序列蕴藏着曲线趋势,并且通常低阶差分就能够提取出曲线趋势的影响,故下面对原始数据进行一阶、二阶差分。部分差分结果如下:

表2 差分数据

Obs	month	express	time	express1	express2
1	2010-11	23220.00	1	.	.
2	2010-12	26366.20	2	3146.20	.
3	2011-01	25301.63	3	-1064.57	-4210.77
4	2011-02	17177.40	4	-8124.23	-7059.66
5	2011-03	28645.40	5	11468.00	19592.23
6	2011-04	28015.00	6	-630.40	-12098.40
7	2011-05	29113.79	7	1098.79	1729.19
8	2011-06	30078.41	8	964.62	-134.17
9	2011-07	29599.01	9	-479.40	-1444.02
10	2011-08	31235.55	10	1636.54	2115.94
11	2011-09	32462.60	11	1227.05	-409.49
12	2011-10	33414.70	12	952.10	-274.95

下面，通过SAS输出一阶差分的时序图，并在此基础上进一步进行单位根检验和自相关图检验。

图2 一阶差分时序图



通过一阶差分的时序图，大致可以看出一阶差分的结果较为平稳。

## 2.模型识别

### （1）平稳性检验

进一步通过单位根检验判断其平稳性。

表3 单位根检验

增广 Dickey-Fuller 单位根检验							
类型	滞后	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
零均值	0	-127.254	0.0001	-11.40	<.0001		
	1	-199.950	0.0001	-9.67	<.0001		
单均值	0	-128.833	0.0001	-11.52	<.0001	66.41	0.0010
	1	-209.845	0.0001	-9.90	<.0001	48.98	0.0010
趋势	0	-129.874	0.0001	-11.61	<.0001	67.40	0.0010
	1	-216.535	0.0001	-10.05	<.0001	50.51	0.0010

上图单位根检验结果表明，Tau统计量的P值显著小于0.05，因此，认为一阶差分后的序列式平稳序列。所以ARIMA模型的差分阶数d=1。

### （2）白噪声检验

为了判断序列是否有分析价值，在进行平稳化处理之后，我们还需要对其进行白噪声检验。如果得到平稳的白噪声序列，则序列无分析价值；如果得到的是平稳的非白噪声序列，则可以进行模型拟合。

表4 白噪声检验

白噪声的自相关检查									
至滞后	卡方	自由度	Pr > 卡方	自相关					
6	26.01	6	0.0002	-0.079	-0.225	-0.270	0.003	0.038	0.273
12	86.79	12	<.0001	-0.010	0.056	-0.232	-0.203	0.007	0.594
18	102.01	18	<.0001	-0.094	-0.127	-0.220	0.012	0.038	0.184
24	139.90	24	<.0001	-0.017	0.042	-0.198	-0.112	-0.001	0.443

一阶差分后的序列的白噪声检验结果如上图所示。从图中可以看出，LB统计量的P值小于0.05，所以我们可以认为一阶差分序列属于非白噪声序列。结合前面的平稳性检验结果，可以看出该序列为平稳的非白噪声序列，可以进行模型拟合。

### 3.模型定阶

对模型进行定阶，通过minic命令做相对最优定阶。

表5 最小信息准则

最小信息准则						
滞后	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	22.20312	22.18782	22.1936	22.2159	22.25341	22.29102
AR 1	22.23619	22.15688	22.18502	22.22416	22.24701	22.28606
AR 2	22.217	22.1484	22.14902	22.18275	22.21909	22.22425
AR 3	22.12547	22.11868	22.12827	22.08572	22.10249	21.79847
AR 4	22.12341	22.15124	22.15792	22.1239	22.06844	21.75523
AR 5	22.11169	22.15047	21.98762	21.93762	21.6996	21.73934

从图中可以看出，BIC信息量相对最小的是ARMA（5，4）模型。

### 4.模型的参数估计及检验

#### （1）参数估计

通过SAS软件对ARMA（5，4）模型进行参数估计。



表6 条件最小二乘估计

条件最小二乘估计					
参数	估计	标准误差	t 值	近似 Pr >  t	滞后
MU	8892.8	5633.5	1.58	0.1173	0
MA1,1	1.05023	0.13402	7.84	<.0001	1
MA1,2	-0.39185	0.10041	-3.90	0.0002	2
MA1,3	1.14206	0.09470	12.06	<.0001	3
MA1,4	-0.88596	0.13053	-6.79	<.0001	4
AR1,1	0.83539	0.13979	5.98	<.0001	1
AR1,2	-0.49127	0.12117	-4.05	<.0001	2
AR1,3	0.97194	0.12044	8.07	<.0001	3
AR1,4	-0.67587	0.14593	-4.63	<.0001	4
AR1,5	0.27439	0.11756	2.33	0.0214	5

(2) 模型检验

拟合模型之后，需要对残差序列进行检验。检验结果如下图所示：

表7 残差的自相关检验

残差的自相关检查									
至滞后	卡方	自由度	Pr > 卡方	自相关					
6	3.86	2	0.1450	-0.047	0.133	0.126	0.085	0.048	0.001
12	6.01	8	0.6467	0.012	-0.040	0.068	-0.012	0.092	0.087
18	9.05	14	0.8276	0.069	0.095	0.091	0.072	0.053	-0.014
24	9.49	20	0.9765	-0.059	-0.013	-0.019	0.007	0.003	0.001

从图中可以看出，p值大于0.05，残差通过了自相关检验，该拟合模型成立。

得到的拟合结果为：

自回归因子	
因子 1:	1 - 0.83539 B**(1) + 0.49127 B**(2) - 0.97194 B**(3) + 0.67587 B**(4) - 0.27439 B**(5)

移动平均因子	
因子 1:	1 - 1.05023 B**(1) + 0.39185 B**(2) - 1.14206 B**(3) + 0.88596 B**(4)

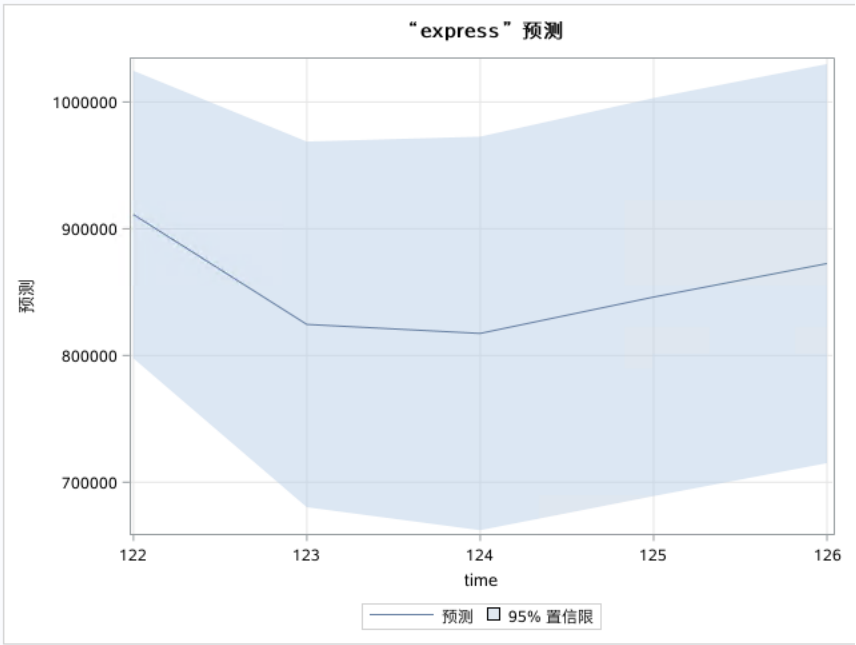
上述输出结果可表示为：

$$(1 - 0.83539B + 0.49127B^2 - 0.97194B^3 + 0.67587B^4 - 0.27439B^5)x_t$$
$$= (1 - 1.05023B + 0.39185B^2 - 1.14206B^3 + 0.88596B^4)\varepsilon_t$$

5.模型的预测

利用此模型对全国未来五个月的当期快递业务量进行预测，结果如下。

变量“express”的预测				
观测	预测	标准误差	95% 置信限	
122	911244.8536	57873.175	797815.5148	1024674.192
123	824580.5084	73580.356	680365.6609	968795.3560
124	817454.3233	79198.980	662227.1749	972681.4718
125	846141.3635	80114.533	689119.7645	1003162.962
126	872566.5515	80316.649	715148.8116	1029984.291



整理得到：

月份	2020-12	2021-01	2021-02	2021-03	2021-04
全国快递业务量 (万件)	911244.8536	824580.5084	817454.3233	846141.3635	872566.5515

### 三、结论

本文根据2010年11月至2020年11月的每月全国快递当期业务量数据作为时间序列，对其进行分析。针对原始数据非平稳数列特征，通过一阶差分将其转变为平稳时间序列，建立全国快递业务量时间序列的ARIMA模型，并在此基础上用于未来五个月全国快递业务量的预测分析。结果表明，该模型能够很好地解决全国快递业务量的估计和预测问题，预测精度较为准确。

### 四、模型的不足

ARIMA模型是变量自身对自身的过去加干扰组合而成的模型，模型中只有一个变量，因此最终得到的模型在进行预测时知识很机械地依据自身固定规律加入干扰，所以模型是刚性的。而现实生活中，全国快递业务量可能会受到各种各样不确定的人为因素的影响，这是模型所无法预测的。例如2020年暴发的新冠疫情对物流行业就产生了长达几个月的颠覆式的影响。因此模型预测的结果仅供参考，难以克服现实生活中不确定因素造成的影响。

### 附录（SAS 程序）

SAS程序代码如下：

```
libname dir "c:\";

data dir.express;
input month $ express;
time=_n_;
cards;
2010-11    23220
2010-12    26366.2
2011-01    25301.63
2011-02    17177.4
2011-03    28645.4
2011-04    28015
2011-05    29113.79
2011-06    30078.41
2011-07    29599.01
2011-08    31235.55
2011-09    32462.6
2011-10    33414.7
```

2011-11	39393.81
2011-12	42873.81
2012-01	27319.4
2012-02	35164.15
2012-03	41850.92
2012-04	40373.75
2012-05	48993.91
2012-06	45447.64
2012-07	46244.93
2012-08	46561.8
2012-09	51345.25
2012-10	52421.3
2012-11	67126.63
2012-12	65698.42
2013-01	68917.73
2013-02	35473.1
2013-03	67043.79
2013-04	66249.37
2013-05	74178.66
2013-06	72154.13
2013-07	74802.13
2013-08	75266.69
2013-09	83637.13
2013-10	85303.54
2013-11	108703.37
2013-12	106945.27
2014-01	88950
2014-02	68886.17
2014-03	102528.22
2014-04	105417.55
2014-05	111634.7
2014-06	112926.28
2014-07	111245.13
2014-08	114050.68
2014-09	122991.97
2014-10	128997.1
2014-11	164616.45
2014-12	163681.07
2015-01	144570.39
2015-02	81779.49
2015-03	142537.83
2015-04	151483.91
2015-05	161002.5
2015-06	164486.72
2015-07	163988.31
2015-08	169020.28
2015-09	191336.92
2015-10	194064.78
2015-11	260537.75
2015-12	241827.96
2016-01	215505.33
2016-02	124618.28
2016-03	236922.22
2016-04	237303.2
2016-05	253108.25
2016-06	257633.1
2016-07	249640.62

2016-08	252296.12
2016-09	282685.88
2016-10	302614.2
2016-11	376447.9
2016-12	339539.99
2017-01	221113.57
2017-02	234430.09
2017-03	303493.69
2017-04	298237
2017-05	334061.75
2017-06	340453.2
2017-07	320498.9
2017-08	326398.6
2017-09	360605.6
2017-10	374771.54
2017-11	471490.26
2017-12	420037.73
2018-01	398613.3
2018-02	198867.2
2018-03	394330.19
2018-04	375275.54
2018-05	417842.24
2018-06	422881.02
2018-07	408083.25
2018-08	410135.98
2018-09	447629.1
2018-10	469187.47
2018-11	586357
2018-12	541822.2
2019-01	452289.51
2019-02	275950.6
2019-03	486392.83
2019-04	491910.89
2019-05	523276.39
2019-06	546077.64
2019-07	524847.57
2019-08	530243.05
2019-09	559722.64
2019-10	575730.08
2019-11	712424.06
2019-12	673425.68
2020-01	378096.8
2020-02	276521.36
2020-03	598423.15
2020-04	649726.38
2020-05	738292.51
2020-06	746967.52
2020-07	693601.43
2020-08	723622.1
2020-09	809158.2
2020-10	823388
2020-11	972623.3

;

run;

proc print;

run;

```

/*绘制散点图*/
proc gplot data = dir.express;
    plot express*month=1;
    symbol1 c=green v=circle i=spline;
run;

/*平稳化处理——差分法*/
data dir.express1 (keep = month time express express1 express2);
set dir.express;
express1=dif(express);
express2=dif(dif(express));
run;
proc print;
run;

/*一二阶差分散点图*/
proc gplot data = dir.express1;
    plot express1*month=2 express2*month=3;
    symbol2 c=red v=circle i=join;
    symbol3 c=black v=circle i=join;
run;

/*一阶差分的自相关图、偏自相关图、白噪声检验、单位根检验，一阶差分平稳*/
proc arima data=dir.express1;
    identify var=express1 stationarity=(adf=1);
run;

/*相对最优定阶，相对最小是（5，4）*/
proc arima data=dir.express1;
    identify var=express1 nlag=18 minic p=(0:5) q=(0:5);
run;

/*参数估计*/
proc arima data=dir.express1;
    identify var=express(1) nlag=8 stationarity=(adf) minic p=(0:5)
q=(0:5);
    estimate p=5 q=4;
    forecast lead=5 id=time out=results;
run;

```