

时间序列分析 期中报告

学号：19020182203614 姓名：吕明倩

一、数据来源及分析背景

本文数据来源于中国民用航空局官网 (<http://www.caac.gov.cn/index.html>)，采集 2006 年 2 月至 2021 年 2 月中国民航国内航线客运量（单位：万人）作为原始数据，通过平稳性分析、季节性分析、模型识别、参数估计、模型诊断等时间序列分析，对未来中国民航国内航线客运量（单位：万人）进行合理预测；同时考虑到 2020 年 2 月新冠疫情对国内民航客运量带来的冲击，结合干扰分析以及异常值检验，对新冠疫情这一干扰因素进行评估。

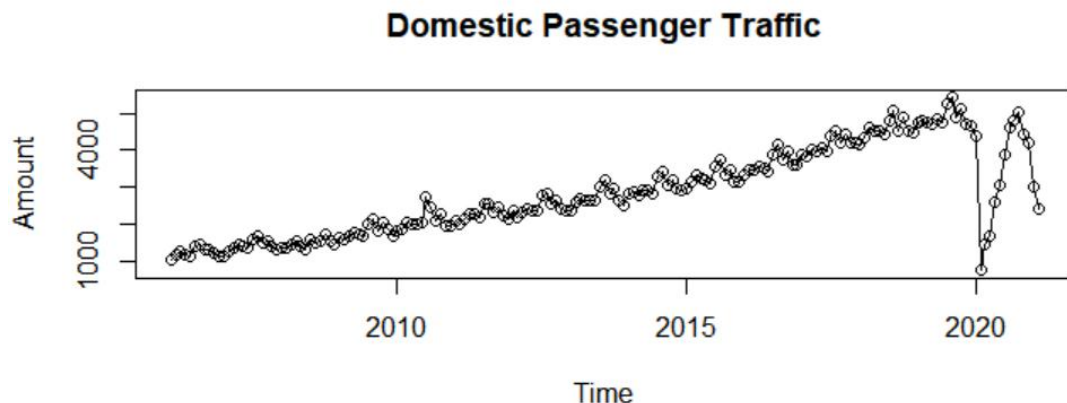
二、时间序列生成

首先导入样本数据 data_inland.csv，该数据集存放 2006 年 2 月至 2021 年 2 月共计 181 个数据（单位：万人）。

```
> head(data1)
Domestic Passenger Traffic
1      1018.0
2      1150.0
3      1234.6
4      1180.1
5      1129.6
6      1382.8
```

生成时间序列 series1，存放 2006 年 2 月至 2021 年 2 月中国民航国内航线客运量（单位：万人）的月度数据。

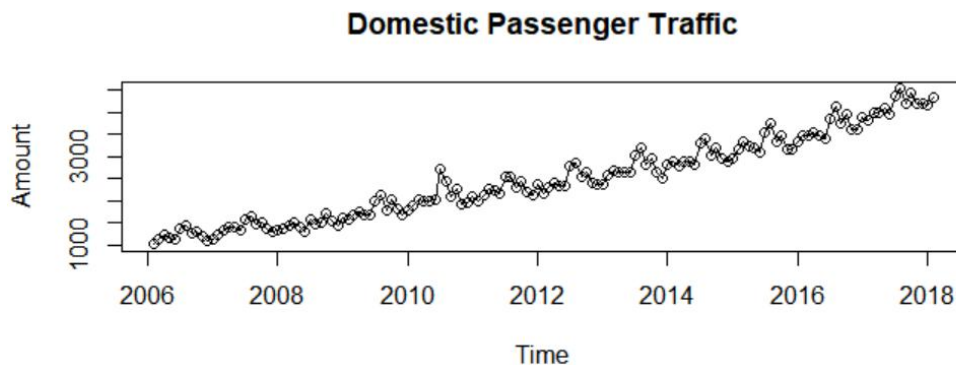
```
> t<-time(series1)
> t[which.min(series1)]
[1] 2020.083
> series1[which.min(series1)]
[1] 728.8
> tsp(series1)
[1] 2006.083 2021.083 12.000
```



通过完整样本数据的时间序列图，我们不难看出该时间序列样本存在显著的异常值，该异常值位于 2020 年 1 月，该月中国民航国内航线客运量为 728.8 万人，显著低于正常水平，并且此后的客运量也受到该异常值的影响。为了得到合理的拟合模型，我们首先截取未受到疫情影响的 2006 年 2 月至 2018 年 2 月的数据（series2）进行模型拟合。

三、平稳性分析

2006 年 2 月至 2018 年 2 月的样本数据（series2）时间序列图如下所示：



通过样本的时间序列图，我们不难发现该时间序列不平稳，具有**线性上升趋势**，且存在**异方差现象**，随着时间的增加方差逐渐增大。

我们可以对其进行**单位根检验**。

```
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-0.135098 -0.020358 -0.000783  0.025700  0.187878

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.14391    0.09335   1.542 0.125856
z.lag.1       -0.01284    0.01183  -1.085 0.279950
z.diff.lag1   -0.40659    0.08507  -4.779 5.11e-06 ***
z.diff.lag2   -0.47414    0.09051  -5.238 7.17e-07 ***
z.diff.lag3   -0.28798    0.09252  -3.113 0.002327 **
z.diff.lag4   -0.46851    0.09152  -5.119 1.21e-06 ***
z.diff.lag5   -0.32296    0.09150  -3.529 0.000595 ***
z.diff.lag6   -0.37539    0.09214  -4.074 8.40e-05 ***
z.diff.lag7   -0.28545    0.09136  -3.124 0.002243 **
z.diff.lag8   -0.46146    0.08912  -5.178 9.35e-07 ***
z.diff.lag9   -0.30307    0.08945  -3.388 0.000957 ***
z.diff.lag10  -0.39629    0.09003  -4.402 2.37e-05 ***
z.diff.lag11  -0.25332    0.08638  -2.933 0.004040 **
z.diff.lag12   0.32972    0.08230   4.006 0.000108 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04597 on 118 degrees of freedom
Multiple R-squared:  0.707,    Adjusted R-squared:  0.6747
F-statistic: 21.9 on 13 and 118 DF, p-value: < 2.2e-16

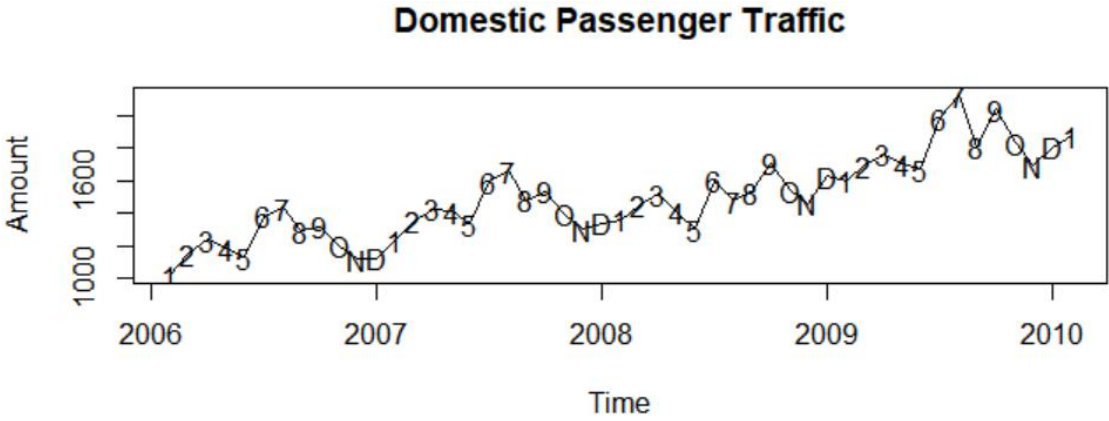
Value of test-statistic is: -1.0854 12.7489

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phil  6.52  4.63  3.81
```

由于 $\tau = -1.0854 > -2.57$ ，即使在 90% 的置信水平下，我们仍无法拒绝原假设，因此我们有充足的理由相信模型是非平稳的。

四、季节性分析

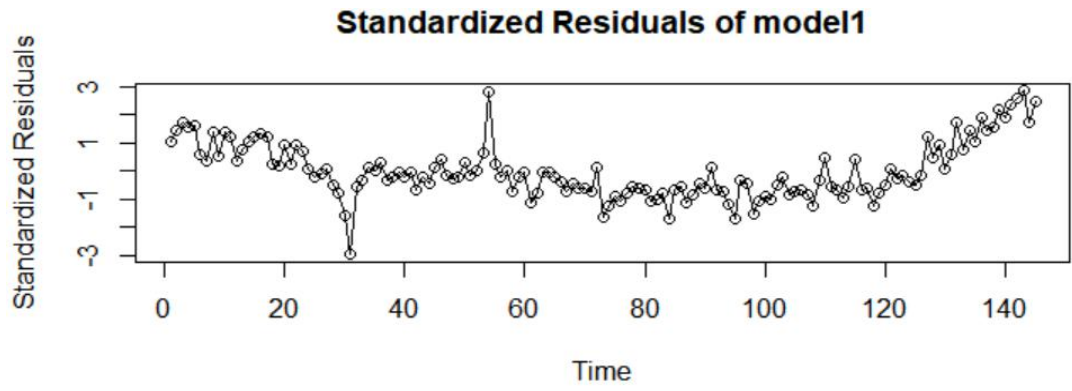
通过截取 2006 年 2 月至 2010 年 2 月的样本数据进行分析，根据时间序列图可以看出样本序列具有显著的季节性波动，中国民航国内航线客运量在 6，7 月份数值较大，在 11，12 月数值较小，数值与季节相关。



为了消除季节性，首先可以考虑样本序列是否具有确定性趋势，利用带有线性趋势的季节均值模型对样本数据进行拟合（Model1），模型参数拟合结果以及残差序列图如下：

```
Call:
lm(formula = series2 ~ time(series2) + month.)

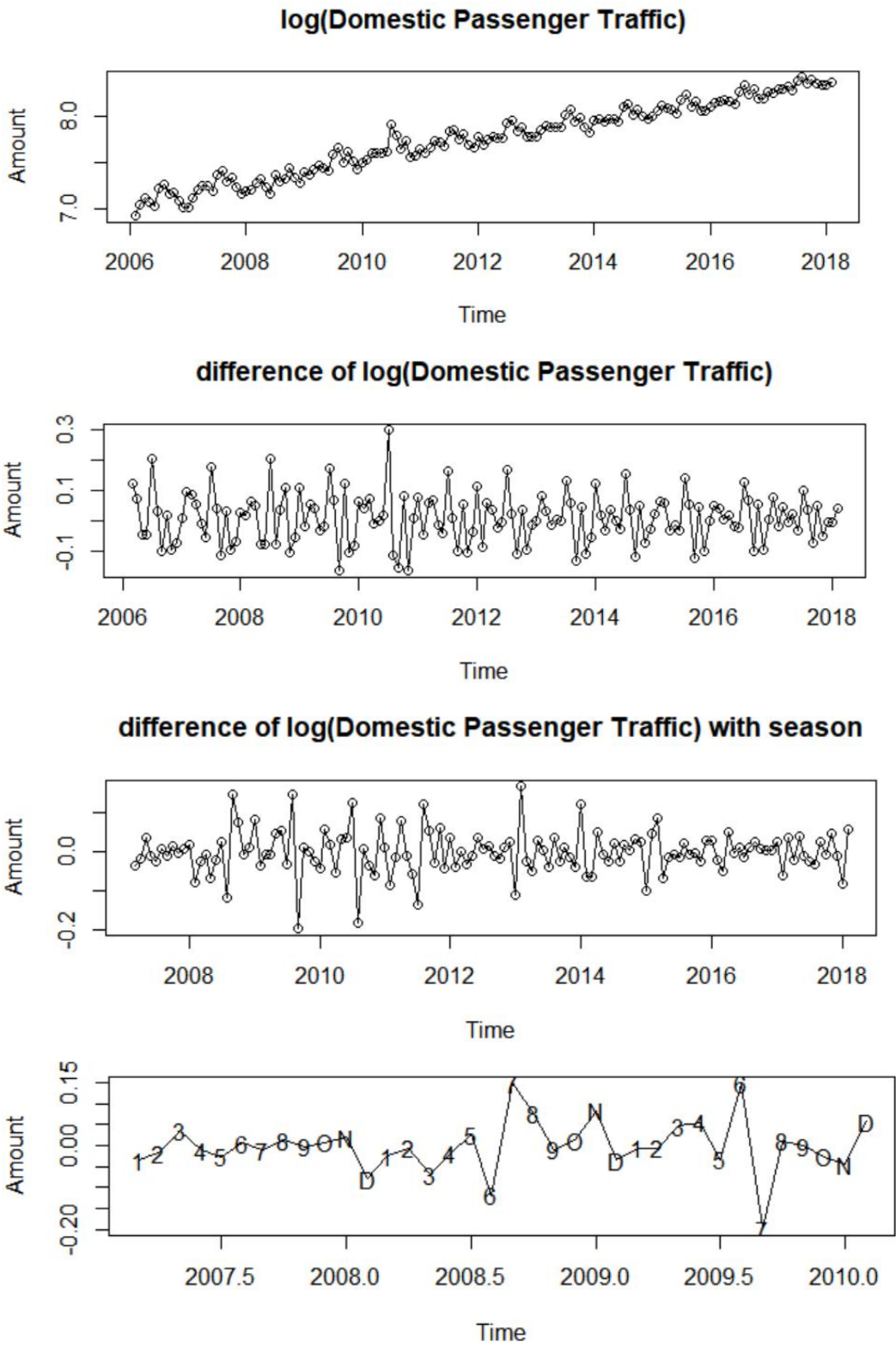
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.171e+05  7.216e+03 -71.659  < 2e-16 ***
time(series2)  2.582e+02  3.586e+00  72.005  < 2e-16 ***
month.February  3.757e+01  6.015e+01   0.625  0.533355
month.March     8.755e+01  6.140e+01   1.426  0.156252
month.April     1.107e+02  6.138e+01   1.804  0.073521 .
month.May       6.363e+01  6.137e+01   1.037  0.301698
month.June      -1.757e+01  6.136e+01  -0.286  0.775115
month.July      3.610e+02  6.135e+01   5.885  3.11e-08 ***
month.August    4.117e+02  6.134e+01   6.711  5.18e-10 ***
month.September 1.091e+02  6.134e+01   1.779  0.077464 .
month.October   2.353e+02  6.133e+01   3.836  0.000193 ***
month.November  -2.723e+01  6.133e+01  -0.444  0.657795
month.December  -1.072e+02  6.133e+01  -1.748  0.082705 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



由于只有 6 月、7 月、10 月的季节变量是显著的，且模型的残差序列不满足零均值，同方差假设，与白噪声存在显著差异，因此带有线性趋势的季节均值模型拟合效果不好，所以考虑采用随机季节模型拟合样本数据的季节趋势。

五、非平稳序列转化

对于非平稳时间序列 series2，我们将采用对数变换、差分变换以及季节差分变换得到较为平稳的数据，过程时间序列图如下所示：



可以看到经过对数变换，样本序列更接近同方差假定，一次差分后，样本序列线性增长趋势消失，经过季节差分，通过部分样本时序图可以看到样本季节性消失，平稳性增强。

再次进行 **ADF 扩展单位根检验**，原假设 $H_0: \rho = 0$ ，在 99% 的置信水平下， $\tau = -4.3567 < -3.46$ ，我们可以拒绝原假设，即认为经过对数变换以及差分变换后的时间序列数据是平稳的，同时由于 $\phi = 9.498 > 6.52$ ，我们可以有 99% 的把握认为样本序列的均值不为零。

Value of test-statistic is: -4.3567 9.498

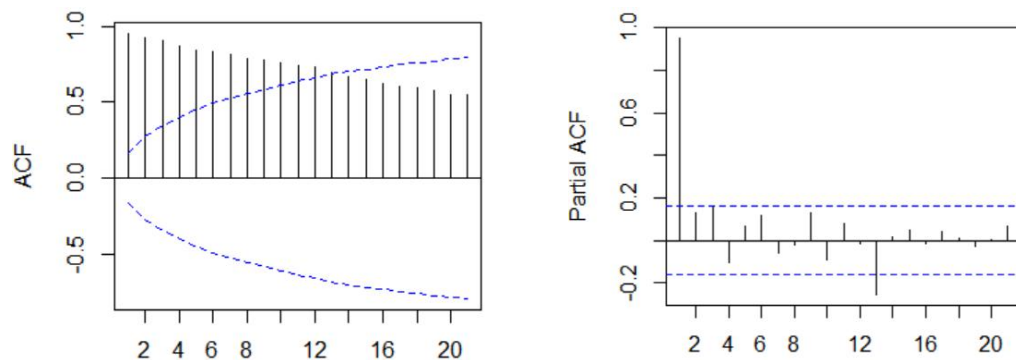
Critical values for test statistics:

| | 1pct | 5pct | 10pct |
|------|-------|-------|-------|
| tau2 | -3.46 | -2.88 | -2.57 |
| phi1 | 6.52 | 4.63 | 3.81 |

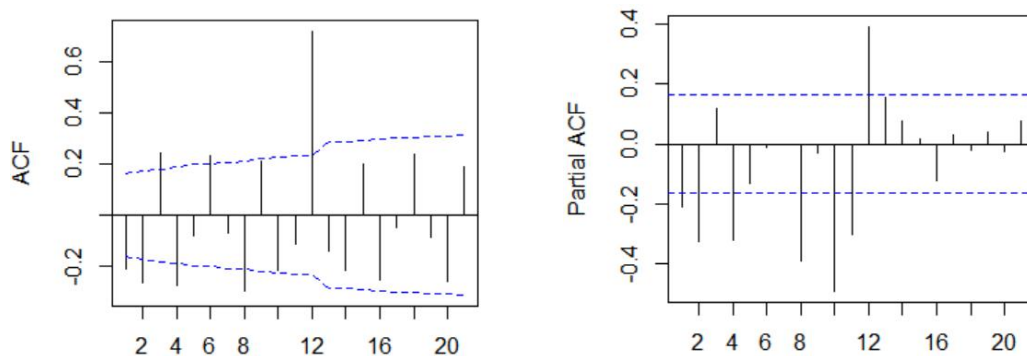
六、模型识别

首先对样本依次添加**对数变换、一次差分变换、季节差分变换**，通过观察**样本自相关图 ACF** 和**偏自相关图 PACF** 的变化情况，识别最终平稳序列的 ARMA 模型。

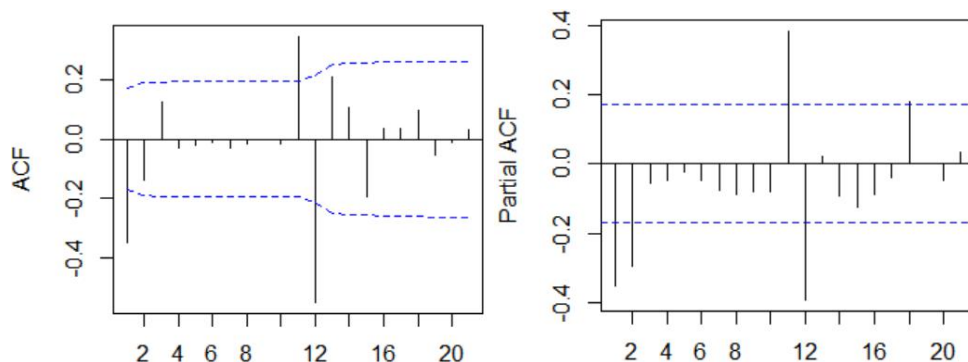
原始样本序列 ACF 和 PACF



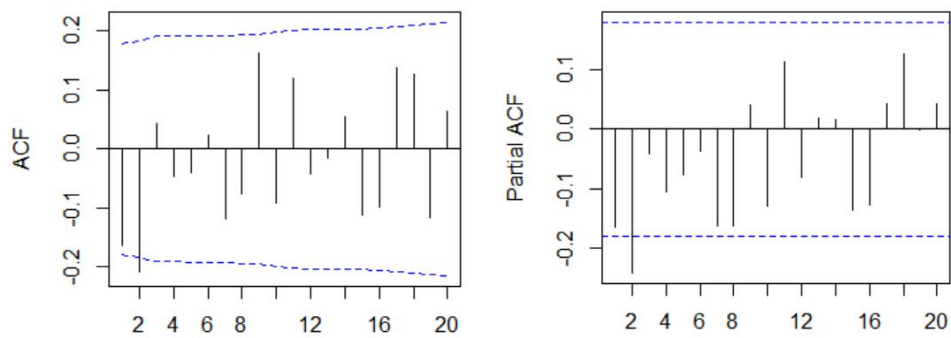
添加对数变换后样本序列 ACF 和 PACF



添加一次差分变换后样本序列 ACF 和 PACF



添加季节差分变换后样本序列 ACF 和 PACF



根据五非平稳序列转化我们知道经过对数变换、一次差分变换、季节差分变换后的时间序列是平稳的，由于其 ACF 是二阶截尾的，可以识别出 MA(2)模型，PACF 也是二阶截尾的，可以识别出 AR(2)模型。

其次根据样本扩展自相关函数 EACF，经过对数变换、一次差分变换、周期为 24 的季节差分变换后，可以识别出 MA(2)、ARMA(1,2)、ARMA(2,2)模型。

```
> eacf(diff(diff(log(series2)),lag=24))
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 0 x o o o o o o o o o o o o
1 x x o o o o o o o o o o o o
2 x x o o o o o o o o o o o o
3 x x x o o o o o o o o o o o
4 x o o x o o o o o o o o o o
5 x o x o o o o o o o o o o o
6 x x x o o x o o o o o o o o
7 x x x o o x x x o o o o o o
```

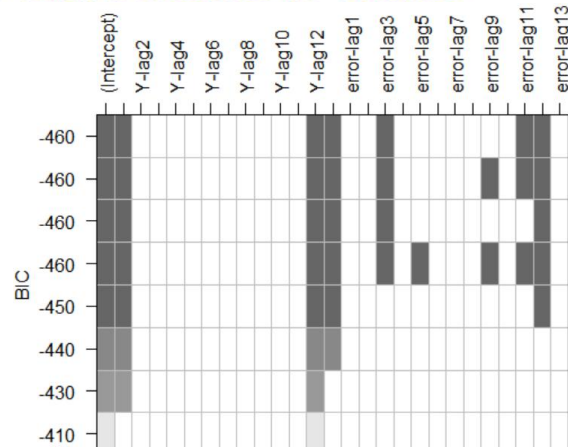
最后采用子集 ARMA(p,q)方法进行模型识别，确定滞后阶为 13，在 BIC 准则下，一次差

分后样本数据拟合模型 $Y_t = \phi Y_{t-1} + \Phi Y_{t-12} + e_t - \theta_3 e_{t-3} - \theta_{11} e_{t-11} - \theta_{12} e_{t-12}$

```
> ar(diff(log(series2)))
Call:
ar(x = diff(log(series2)))

Coefficients:
      1      2      3      4      5      6      7      8
-0.3625 -0.3864 -0.1163 -0.3350 -0.1688 -0.2605 -0.1322 -0.3494
      9     10     11     12     13
-0.1602 -0.3004 -0.0748  0.4370  0.1553

Order selected 13  sigma^2 estimated as  0.002634
```



因此，还原已进行的差分变换，我们可以识别出 7 个模型： $ARIMA(2,1,0)*(0,1,0)_{24}$ ， $ARIMA(0,1,2)*(0,1,0)_{24}$ ， $ARIMA(1,1,2)*(0,1,0)_{24}$ ， $ARIMA(2,1,2)*(0,1,0)_{24}$ ， $ARIMA(2,1,1)*(1,1,1)_{12}$ ， $ARIMA(1,1,1)*(0,1,1)_{12}$ ， $ARIMA(0,1,1)*(0,1,1)_{12}$ ，依次记为 Model2 至 Model8。

七、参数估计

我们采用最大似然方法（MLE）对 Model2 至 Model8 参数进行估计，各模型具体估计结果如“附录一”所示，我们将各样本参数估计值以及 AIC 值汇总如下。

| ARIMA | $(p, d, q) * (P, D, Q)_s$ | ϕ_1 | ϕ_2 | θ_1 | θ_2 | Φ_1 | Θ_1 | AIC |
|--------|---------------------------|----------|----------|------------|------------|----------|------------|---------|
| Model2 | $(2,1,0) * (0,1,0)_{24}$ | -0.205 | -0.244 | 0 | 0 | 0 | 0 | -369.83 |
| Model3 | $(0,1,2) * (0,1,0)_{24}$ | 0 | 0 | -0.279 | -0.322 | 0 | 0 | -373.93 |
| Model4 | $(1,1,2) * (0,1,0)_{24}$ | 0.473 | 0 | -0.745 | -0.113 | 0 | 0 | -373.16 |
| Model5 | $(2,1,2) * (0,1,0)_{24}$ | -0.247 | 0.541 | 0.001 | -0.877 | 0 | 0 | -373.31 |
| Model6 | $(2,1,1) * (1,1,1)_{12}$ | 0.444 | -0.010 | -0.127 | 0 | -0.126 | -0.659 | -461.39 |
| Model7 | $(1,1,1) * (0,1,1)_{12}$ | 0.426 | 0 | -0.838 | 0 | 0 | -0.735 | -464.44 |
| Model8 | $(0,1,1) * (0,1,1)_{12}$ | 0 | 0 | -0.482 | 0 | 0 | -0.712 | -460.83 |

通过模型参数估计，我们可以得到如下结论：

首先，对于 Model4、Model5 和 Model6，在采用最大似然准则估计模型参数时，参数的估计值均小于两倍的标准误，因此参数是不显著的，所以 Model4、Model5 和 Model6 不是适合的模型。

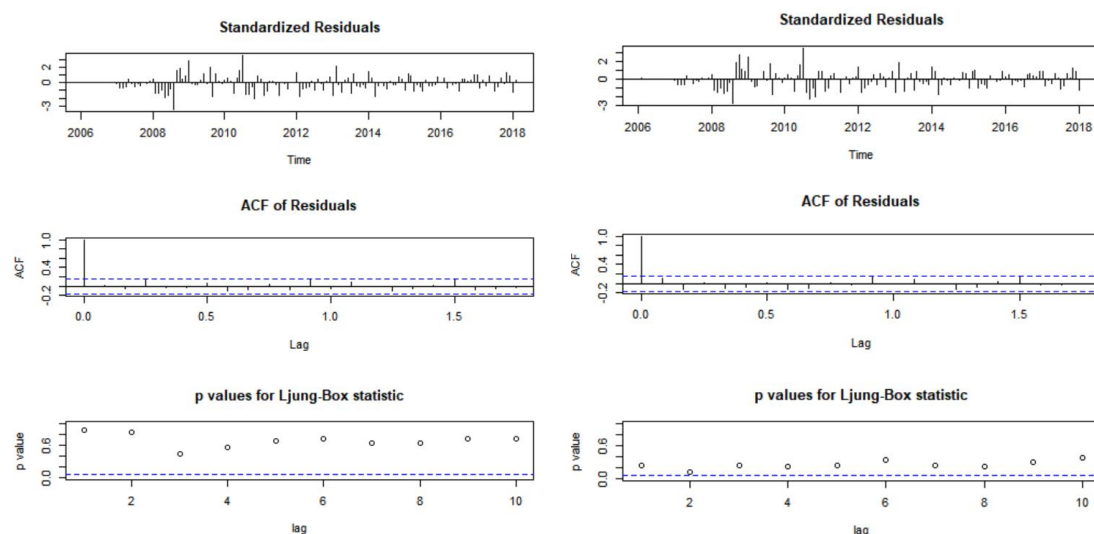
其次，在 AIC 准则下，采用周期为 12 的季节差分得到的模型，其 AIC 值明显较小，因此周期为 12 的季节差分模型拟合效果优于周期为 24 的季节差分，所以 Model2 和 Model3 不是较为适合的模型。

最后，对于 Model7 和 Model8，二者各参数估计值均是显著的，并且二者 AIC 值相差不大，我们将进一步进行残差分析以及正态性检验来做出抉择。

八、模型诊断

(i) 残差分析

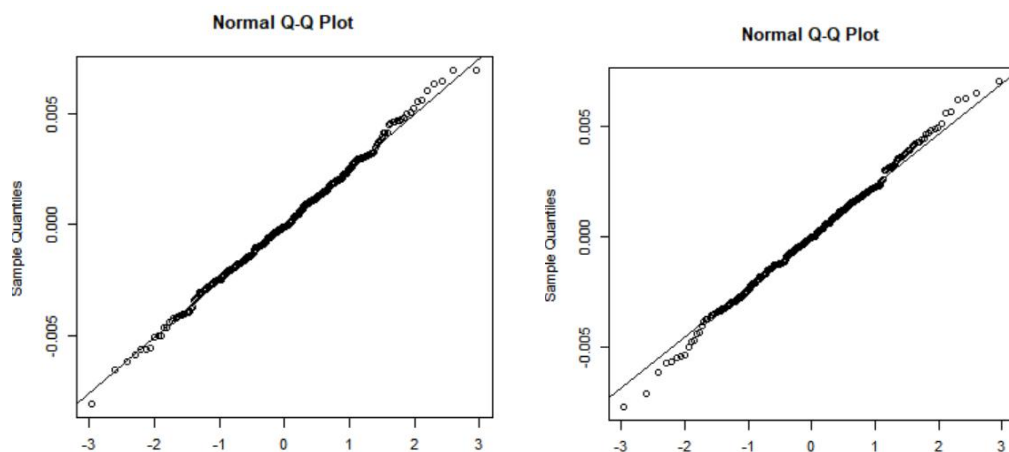
对 Model7 和 Model8 分别进行 `tsdiag()` 检验，结果如下图所示：



通过样本残差分析我们可以看出，Model7 和 Model8 的残差在 0 附近均匀随机分布，不存在残差显著异常的值点，并且残差之间几乎不存在自相关性。在 **Ljung-Box 检验**中，虽然 Model7 和 Model8 的 p-value 均比 0 大，但相较于 Model8，Model7 的 p-value 与 0 的差异性更大，我们更偏向于 Model7 的样本自相关性更弱，模型拟合效果更好。

(ii) 正态性检验

QQ 图



shapiro 检验

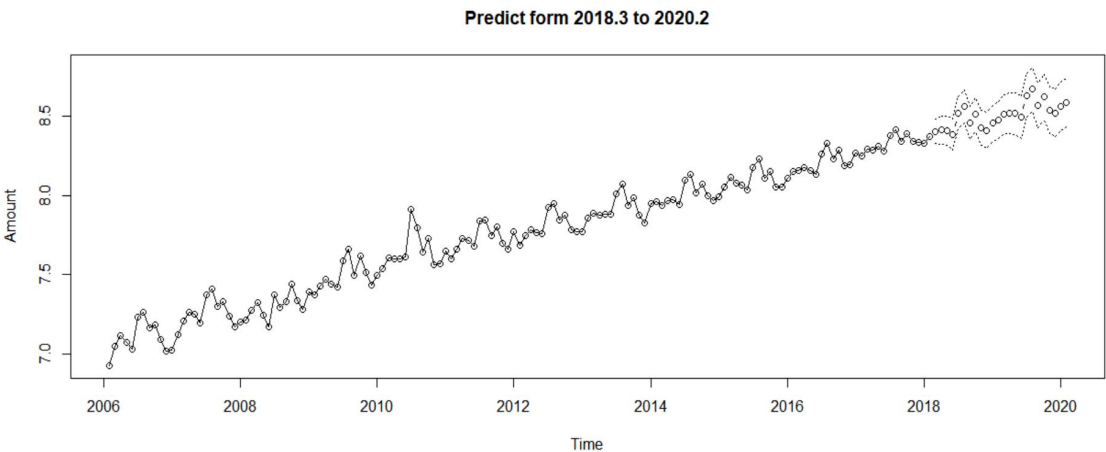
```
Shapiro-Wilk normality test      Shapiro-Wilk normality test
data: residuals(model3)          data: residuals(model4)
W = 0.99758, p-value = 0.9194    W = 0.99689, p-value = 0.791
```

通过正态性检验我们可以看出，Model7 和 Model8 均不拒绝正态性假设，但在 **shapiro-wilk 检验**中 Model7 的 p-value=0.9194 比 Model8 的 p-value=0.791 更大，所以 Model7 的正态性更好，拟合残差更接近白噪声的正态性假定。

九、模型预测

根据对于 2006 年 2 月至 2018 年 2 月中国国内民航客运量的时间序列分析，我们以下将采用 $ARIMA(1,1,1) \times (0,1,1)_{12}$ 模型对 2018 年 3 月至 2020 年 2 月（疫情之前）的中国国内民航客运量进行预测，并与已知数据进行比较，判断预测的有效性。

预测曲线以及预测区间如下所示：



```
> cbind(real,ci.down,predict,ci.up)
```

| | real | ci.down | predict | ci.up |
|----------|--------|----------|----------|----------|
| Mar 2018 | 4611.0 | 4132.789 | 4462.933 | 4819.451 |
| Apr 2018 | 4540.5 | 4119.240 | 4503.224 | 4923.001 |
| May 2018 | 4504.8 | 4081.885 | 4486.757 | 4931.788 |
| Jun 2018 | 4415.9 | 3963.876 | 4372.228 | 4822.647 |
| Jul 2018 | 4811.7 | 4541.569 | 5023.295 | 5556.119 |
| Aug 2018 | 5062.7 | 4708.266 | 5220.496 | 5788.454 |
| Sep 2018 | 4526.1 | 4243.448 | 4715.985 | 5241.142 |
| Oct 2018 | 4883.3 | 4465.551 | 4973.857 | 5540.023 |
| Nov 2018 | 4502.3 | 4089.476 | 4564.827 | 5095.432 |
| Dec 2018 | 4458.3 | 4020.860 | 4497.729 | 5031.153 |
| Jan 2019 | 4733.4 | 4192.940 | 4699.944 | 5268.255 |
| Feb 2019 | 4785.4 | 4277.883 | 4804.923 | 5396.895 |
| Mar 2019 | 4745.8 | 4395.897 | 4973.927 | 5627.964 |
| Apr 2019 | 4703.8 | 4418.180 | 5021.607 | 5707.448 |
| May 2019 | 4847.8 | 4387.809 | 5004.422 | 5707.688 |
| Jun 2019 | 4733.3 | 4263.210 | 4877.167 | 5579.543 |
| Jul 2019 | 5271.4 | 4884.312 | 5603.664 | 6428.961 |
| Aug 2019 | 5422.7 | 5062.272 | 5823.754 | 6699.781 |
| Sep 2019 | 4869.4 | 4560.927 | 5260.984 | 6068.494 |
| Oct 2019 | 5093.8 | 4797.824 | 5548.676 | 6417.034 |
| Nov 2019 | 4715.4 | 4392.041 | 5092.382 | 5904.397 |
| Dec 2019 | 4643.6 | 4316.635 | 5017.532 | 5832.234 |
| Jan 2020 | 4401.1 | 4499.585 | 5243.119 | 6109.518 |
| Feb 2020 | 728.8 | 4588.919 | 5360.231 | 6261.186 |

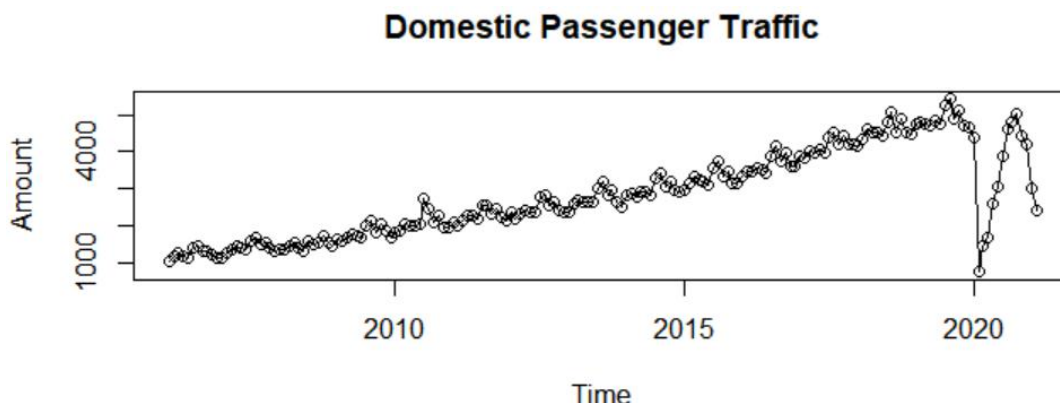
根据预测结果和真实值相比较，除 2020 年 2 月的真实数据之外，其余时间的真实值均落在 95% 的预测区间内，所以该模型的预测效果很好，因此未受到干扰（疫情之前）的时间序列最终可以确定为 $ARIMA(1,1,1) \times (0,1,1)_{12}$ 模型。

十、异常值分析及影响

最后我们将 2020 年 2 月之后的数据纳入模型中进行考量，对采集到的完整数据进行模型拟合，并预测未来中国国内实际民航客运量的变化情况。

由于 2020 年 1 月底新冠疫情的爆发，我国迅速采取积极防疫措施，鼓励就地过年，减少人口的流动，所以在 2020 年 1 月我国国内民航客运量总额明显降低，并且后续一段时间内国内民航客运量也受到影响，我们将进行干扰分析以及异常值检验，在原模型的基础上增加对异常值的考量。

完整时间序列（series1）的样本时间序列图如下所示：



(i) 干扰分析

对于模型 $Y_t = m_t + N_t$ ，其中 N_t 为无干扰自然过程，经过前序分析可以采用

$ARIMA(1,1,1) \times (0,1,1)_{12}$ 模型进行拟合，对于干扰均值效应 m_t ，考虑到疫情发生当下即刻产生脉冲响应，且随着时间的增加，疫情逐渐好转，在较长一段时间的控制以及调整后，该干扰对中国国内民航客运量的影响最终将消失，所以我们设定干扰均值效应为

$$m_t = w_0 P_t^{(T)} + \frac{w_1}{1 - \delta B} P_t^{(T)}, \text{ 进行如下干扰分析。}$$

```
> model9

Call:
arimax(x = log(series1), order = c(1, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12), method = "ML", xtransf = data.frame(p1 = 1 * (seq(series1) == 169), p2 = 1 * (seq(series1) == 169)), transfer = list(c(0, 0), c(1, 0)))

Coefficients:
      ar1      mal      smal  p1-MA0  p2-AR1  p2-MA0
 0.0255 -0.0037 -0.4848 -0.2635  0.7146 -1.4508
s.e.      NaN      NaN  0.1032  0.1066  0.0378  0.1160

sigma^2 estimated as 0.003328:  log likelihood = 239.27,  aic = -466.54
```

可以看到参数 w_0 、 w_1 、 δ 值均在两倍标准误内，因此新引入的干扰均值效应是显著的。

(ii) 异常值检验

我们在上述进行干扰分析后所得模型（Model9）的基础上，继续进行异常值检验。每次仅处理最显著的异常点，直至消除所有异常点，具体过程如“附录二”所示。

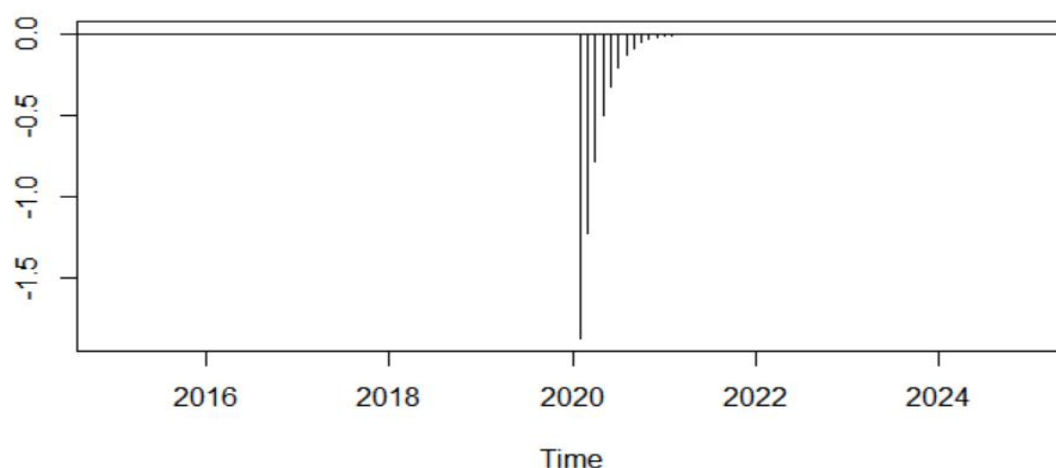
根据异常值检验可以检测到的异常值有：可加异常值(AO)：181(2021年2月)；新息异常值(IO)：180(2021年1月)，171(2020年4月)以及54(2010年7月)。在该过程中我们建立模型 m1 至 m4，依次为将异常值点 180、181、171、54 添加到模型后的拟合模型，最终得到的 m4 模型不再含有异常值，可以作为本次模型拟合的最终模型从而对未来客运量做出合理预测。

(iii) 影响时限

最后，我们对干扰点的影响时限做出简要分析。

由于 2020 年 1 月新冠疫情爆发这一时刻的外部干扰，除了带来当下猛烈冲击外，理论上其影响会随着时间逐渐消失，通过滤波可以预测大约一年，也就是 2021 年 3 月开始，中国国内民航客运量的变化情况不再收到新冠疫情冲击的影响。

但上述考量未考虑到新冠疫情的持续性，即从 2020 年 2 月开始的一段时间内，该样本时间序列数据持续受到新冠疫情的干扰，结合现实情况我们不难发现，实际影响时限将远远大于一年。



总结

本文从 2006 年 2 月至 2021 年 2 月中国民航国内客运量这一原始数据出发，首先对未受到干扰前的自然序列进行模型拟合。截取 2006 年 2 月至 2018 年 2 月的时间序列数据作为训练集，通过样本时间序列图以及单位根分析，从直观上识别并从理论上证明模型的平稳性，通过分别拟合带有线性趋势的季节均值模型和随机差分模型对样本的季节性规律进行识别，从而通过对数差分得到平稳数据，再进行 ARMA 模型识别、参数估计、模型诊断，综合残差分析、自相关分析以及 AIC 值等多重指标得到最合适的 $ARIMA(1,1,1)*(0,1,1)_{12}$ 模型，进而可以预测未受到新冠疫情影响时，中国民航国内客运量的发展趋势。

紧接着，结合新冠疫情这一重大干扰点的分析，识别出模型的异常点，对 $ARIMA(1,1,1)*(0,1,1)_{12}$ 进行异常值检验，通过逐个消除异常点对模型进行修正，并预测影响时限。在经过一段时间的恢复后，干扰会逐渐消失，我们仍然可以用遭受疫情之前的 $ARIMA(1,1,1)*(0,1,1)_{12}$ 模型，对未来发展情况进行合理预测。

附录一：Model2 至 Model8 参数估计

```
> model2

Call:
arima(x = log(series2), order = c(2, 1, 0), seasonal = list(order = c(0, 1,
0), period = 24))

Coefficients:
          ar1          ar2
      -0.2051   -0.2435
s.e.   0.0889   0.0888

sigma^2 estimated as 0.002595:  log likelihood = 186.91,  aic = -369.83
> model3

Call:
arima(x = log(series2), order = c(0, 1, 2), seasonal = list(order = c(0, 1,
0), period = 24))

Coefficients:
          ma1          ma2
      -0.279   -0.3220
s.e.   0.090   0.0989

sigma^2 estimated as 0.002502:  log likelihood = 188.97,  aic = -373.93
> model4

Call:
arima(x = log(series2), order = c(1, 1, 2), seasonal = list(order = c(0, 1,
0), period = 24))

Coefficients:
          ar1          ma1          ma2
      0.4727   -0.7452   -0.1127
s.e.   0.4998   0.5321   0.3036

sigma^2 estimated as 0.002473:  log likelihood = 189.58,  aic = -373.16
> model5

Call:
arima(x = log(series2), order = c(2, 1, 2), seasonal = list(order = c(0, 1,
0), period = 24))

Coefficients:
          ar1          ar2          ma1          ma2
      -0.2466   0.5405   0.0011   -0.8766
s.e.   0.3495   0.3964   0.2557   0.2875

sigma^2 estimated as 0.002421:  log likelihood = 190.66,  aic = -373.31
> model6

Call:
arima(x = log(series2), order = c(2, 1, 1), seasonal = list(order = c(1, 1,
1), period = 12))

Coefficients:
          ar1          ar2          ma1          sar1          smal
      0.4435   -0.0100   -0.8386   -0.1261   -0.6592
s.e.   0.1502   0.1176   0.1255   0.1265   0.1159

sigma^2 estimated as 0.00153:  log likelihood = 235.7,  aic = -461.39
```



```

> model17

Call:
arima(x = log(series2), order = c(1, 1, 1), seasonal = list(order = c(0, 1,
1), period = 12))

Coefficients:
      ar1      mal      smal
    0.4255 -0.8383 -0.7354
s.e.  0.1459  0.1006  0.0788

sigma^2 estimated as 0.001537:  log likelihood = 235.22,  aic = -464.44

> model18

Call:
arima(x = log(series2), order = c(0, 1, 1), seasonal = list(order = c(0, 1,
1), period = 12))

Coefficients:
      mal      smal
   -0.4817  -0.7157
s.e.   0.0997  0.0810

sigma^2 estimated as 0.001618:  log likelihood = 232.42,  aic = -460.83

```

附录二：Model9、m1 至 m4 异常值检验

```

> detectAO(model9)
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
ind    168.000000 169.00000 171.000000 172.000000 180.000000 181.000000
lambda2 -5.391263 -4.03587 -3.698535  4.855669 -7.833283 -4.550102

> detectIO(model9)
      [,1]      [,2]      [,3]
ind    172.000000 180.000000 181.000000
lambda1  4.877713 -7.934122 -4.550102

> detectAO(m1)
      [,1]      [,2]      [,3]      [,4]
ind    169.00 171.000000 172.000000 181.000000
lambda2 -4.55 -3.719223  5.435228 -5.571565

> detectIO(m1)
      [,1]      [,2]
ind    172.000000 181.000000
lambda1  5.454713 -5.571565

> detectAO(m2)
[1] "No AO detected"

> detectIO(m2)
      [,1]      [,2]      [,3]
ind    31.000000 54.000000 171.000000
lambda1 -3.986852  3.937864 -5.265715

> detectAO(m3)
[1] "No AO detected"

> detectIO(m3)
      [,1]      [,2]
ind    31.000000 54.000000
lambda1 -4.107215  4.091975

> detectAO(m4)
[1] "No AO detected"

> detectIO(m4)
[1] "No IO detected"

```


附录三：R 语言代码

```
library(TSA)
library(fUnitRoots)
library(urca)
library(forecast)

rm(list=ls())
getwd()
setwd("E:/2021 春季学期/时间序列分析/作业/大作业")
data1=read.csv("data_inland.csv",header=F)
colnames(data1)=c("Domestic Passenger Traffic")
head(data1)
series1=ts(data1,start=c(2006,2),freq=12)
tsp(series1)
t<-time(series1)
t[which.min(series1)]
series1[which.min(series1)]

win.graph(width=7,height=3)
plot(series1,type='o',main="Domestic Passenger Traffic",ylab="Amount")
#ADF 检验
summary(ur.df(log(series2),type="drift",lags =12,selectlags = "AIC") )
series2=window(series1,end=c(2018,2))
plot(series2,type='o',main="Domestic Passenger Traffic",ylab="Amount")
win.graph(width=7,height=3)
Month=c("1","2","3","4","5","6","7","8","9","O","N","D")
plot(window(series1,end=c(2010,2)),main="Domestic Passenger Traffic",ylab="Amount")
points(window(series1,end=c(2010,2)),pch=Month)

#确定性趋势
month.=season(series2)
model1=lm(series2~time(series2)+month.)
summary(model1)
plot(rstudent(model1),xlab='Time',ylab='Standardized Residuals',
      main='Standardized Residuals of model1',type='o')

#得到平稳数据
plot(log(series2),type='o',main="log(Domestic Passenger Traffic)",ylab="Amount")
plot(diff(log(series2)),type='o',
      main="difference of log(Domestic Passenger Traffic)",ylab="Amount")
plot(diff(diff(log(series2)),lag=12),type='o',
      main="difference of log(Domestic Passenger Traffic) with season",ylab="Amount")
plot(diff(diff(log(window(series1,end=c(2010,2))))),lag=12),
      main="difference of log(Domestic Passenger Traffic) with seasond",ylab="Amount")
```

```
points(diff(diff(log(window(series1,end=c(2010,2))))),lag=12),pch=Month)
summary(ur.df(diff(diff(log(series2))),lag=12,type="drift",lags =12,selectlags = "AIC") )
```

#模型识别

```
win.graph(width=4,height=4)
acf(as.numeric(log(series2)),ci.type='ma',main='log(series2)',xaxp=c(0,20,10))
acf(as.numeric(diff(log(series2))),ci.type='ma',
    main='difference of log(series2)',xaxp=c(0,20,10))
acf(as.numeric(diff(diff(log(series2))),lag=12)),ci.type='ma',
    main='difference of log(series2) with season',xaxp=c(0,20,10))
acf(as.numeric(diff(diff(log(series2))),lag=24)),ci.type='ma',
    main='difference of log(series2) with season',xaxp=c(0,20,10))

pacf(as.numeric(log(series2)),main='log(series2)',xaxp=c(0,20,10))
pacf(as.numeric(diff(log(series2))),main='difference of log(series2)',xaxp=c(0,20,10))
pacf(as.numeric(diff(diff(log(series2))),lag=12)),main='difference of log(series2) with
season',xaxp=c(0,20,10))
pacf(as.numeric(diff(diff(log(series2))),lag=24)),main='difference of log(series2) with
season',xaxp=c(0,20,10))

eacf(diff(diff(log(series2))),lag=24))
ar(diff(diff(log(series2))),lag=24))

ar(diff(diff(log(series2))),lag=24))
res=armasubsets(series2,nar=13,nma=13)
win.graph(width=6,height=5)
plot(res)
```

#参数估计

```
model2=arima(log(series2),order=c(2,1,0),seasonal=list(order=c(0,1,0),period=24))
model2
model3=arima(log(series2),order=c(0,1,2),seasonal=list(order=c(0,1,0),period=24))
model3
model4=arima(log(series2),order=c(1,1,2),seasonal=list(order=c(0,1,0),period=24))
model4
model5=arima(log(series2),order=c(2,1,2),seasonal=list(order=c(0,1,0),period=24))
model5
model6=arima(log(series2),order=c(2,1,1),seasonal=list(order=c(1,1,1),period=12))
model6
model7=arima(log(series2),order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12))
model7
model8=arima(log(series2),order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12))
model8
```

#模型诊断

```
win.graph(width=7,height=7)
tsdiag(model7)
tsdiag(model8)
win.graph(width=4,height=4)
qqnorm(residuals(model7))
qqline(residuals(model7))
shapiro.test(residuals(model7))
qqnorm(residuals(model8))
qqline(residuals(model8))
shapiro.test(residuals(model8))
```

#预测

```
win.graph(width=7,height=4)
result=plot(model7,n.ahead=24,type='b',xlab='Time',ylab='Amount',main='Predict form 2018.3 to
2020.2')
abline(h=coef(model7)[names(coef(model7))=='intercept'])
series3=window(series1,start=c(2018,3),end=c(2020,2))
```

```
forecast=result$pred
real=series3
predict=exp(forecast)
ci.up=exp(result$upi)
ci.down=exp(result$ipi)
cbind(real,ci.down,predict,ci.up)
```

#干扰分析

```
model9=arimax(log(series1),order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12),
  xtransf=data.frame(p1=1*(seq(series1)==169),p2=1*(seq(series1)==169)),
  transfer=list(c(0,0),c(1,0)),method='ML')
model9
```

#异常值检验

```
detectAO(model9)
detectIO(model9)
m1=arimax(log(series1),order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12),
  xtransf=data.frame(p1=1*(seq(series1)==169),p2=1*(seq(series1)==169)),
  transfer=list(c(0,0),c(1,0)),method='ML',
  io=c(180))
detectAO(m1)
detectIO(m1)
m2=arimax(log(series1),order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12),
  xtransf=data.frame(p1=1*(seq(series1)==169),p2=1*(seq(series1)==169)),
```

```

transfer=list(c(0,0),c(1,0)),method='ML',
io=c(180),xreg=data.frame(AO=seq(series1)==181))
detectAO(m2)
detectIO(m2)
m3=arimax(log(series1),order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12),
xtransf=data.frame(p1=1*(seq(series1)==169),p2=1*(seq(series1)==169)),
transfer=list(c(0,0),c(1,0)),method='ML',
io=c(180,171),xreg=data.frame(AO=seq(series1)==181))
detectAO(m3)
detectIO(m3)
m4=arimax(log(series1),order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12),
xtransf=data.frame(p1=1*(seq(series1)==169),p2=1*(seq(series1)==169)),
transfer=list(c(0,0),c(1,0)),method='ML',
io=c(180,171,31),xreg=data.frame(AO=seq(series1)==181))
detectAO(m4)
detectIO(m4)

#影响何时会消失
win.graph(width=7,height=4)
p=1*(seq(series1)==169)
plot(ts(p*(0.0366)+ filter(p,filter=0.6419,method='recursive',side=1)*(-1.9037),
frequency=12,start=c(2006,2)),xlim=c(2015,2025),type='h')
abline(h=0)

```