

时间序列大作业

2022 年 5 月 4 日

班 级:	1907031	学 号:	19070300002
姓 名:	徐文韬	指导老师:	王红军

摘要

本文尝对俄乌战争中的防空炮损坏数量进行建模分析。

1. 首先,我们观测时间序列,这是一个明显不平稳的时间序列。通过二次差分来使得时间序列平稳化。
2. 然后,我们尝试对时间序列进行 ARIMA 模型建模。我们通过观察 ACF,PACF 以及 EACF 图来确定模型的阶数,并借助 AUTO.ARIMA 函数来快速计算各个阶数组合下的 AIC 值来辅助我们的阶数确定。最后,我们将模型确定为 ARIMA(0,2,1) 模型。
3. 接着,我们对模型进行模型诊断。我们通过观察残差的 ACF 和计算 LJUNG-BOX 统计量,认为残差是符合独立性假设的。接着我们观察残差的 QQ 图,发现发现大部分点都符合正态分布,但是有四个异常点发生了显著的偏移现象。
4. 最后,我们单独分析了这四个点 (2/28, 3/4, 3/16, 4/12),发现在这四天左右发生了大规模空军的军事活动。这些特别的军事战术决策导致的异常现象是我们无法从防空炮自身的时间序列中挖掘到的。因此,我们提出了从飞机损失数量等时序中提取数据构造干预函数,从而对 ARIMA 模型进行修正的意见。

Keywords: ARIMA; 差分; LJUNG-BOX; 模型诊断; 干预函数

目录

1 模型准备	2
2 ARIMA 模型建立	2
2.1 平稳化	2
2.2 模型阶数选择	3
2.3 参数估计与拟合	5
3 模型诊断	5
4 问题分析与模型改进	7
5 总结	8

1 模型准备

让我们首先来仔细观察俄乌战争中，防空炮损失数量的时间序列图。

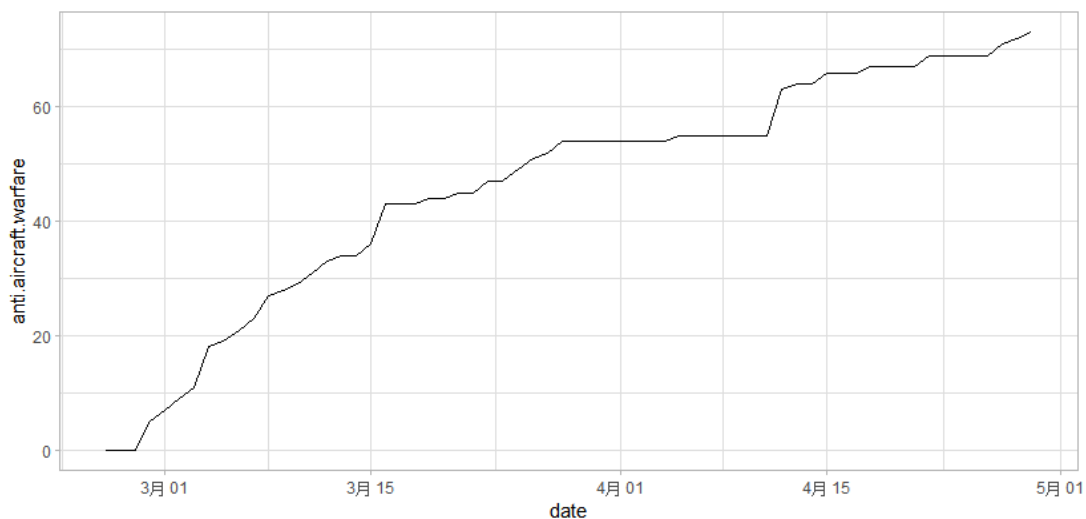


图 1: 高射炮损失数量时间序列图

显然的，这个时间序列明显呈现上升趋势，所以它的均值一定不符合平稳性的要求。其次，时间序列在诸如 3.16, 4.11 日等时间节点就陡然增加的情况，这是我们在后面的模型建立与诊断需要注意的。(数据来自<https://www.kaggle.com/datasets/piterfm/2022-ukraine-russian-war>)

2 ARIMA 模型建立

2.1 平稳化

首先，正如模型准备中提到的，这个时间序列是不平稳的。它呈现上升趋势，因此我们尝试通过差分的手段将它平稳化：

值得注意的是，一次差分后得到的时间数列会比原序列减少一个数据。

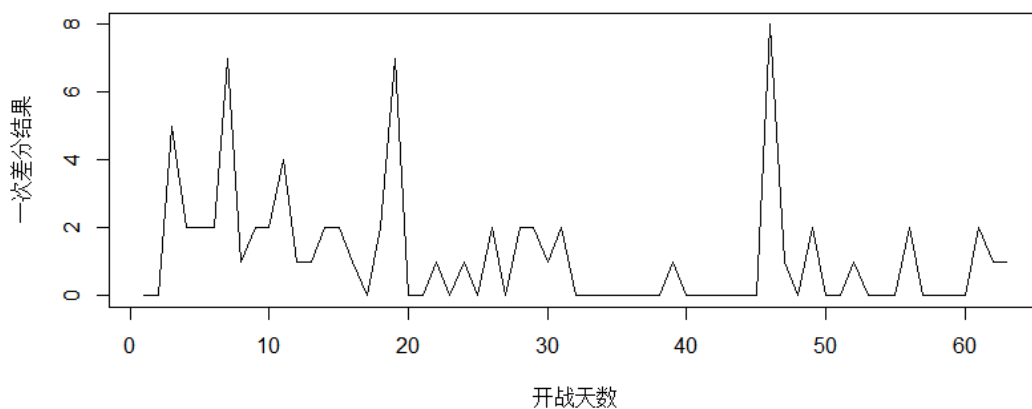


图 2: 一次差分序列

我们看到经过一次差分后的时序图仍然呈现非平稳的状态：数据集中在 0 的上方浮动。因此，我们对时间序列进行二次差分。

$$\nabla^2 Y_t = \nabla Y_t - \nabla Y_{t-1} \quad (1)$$

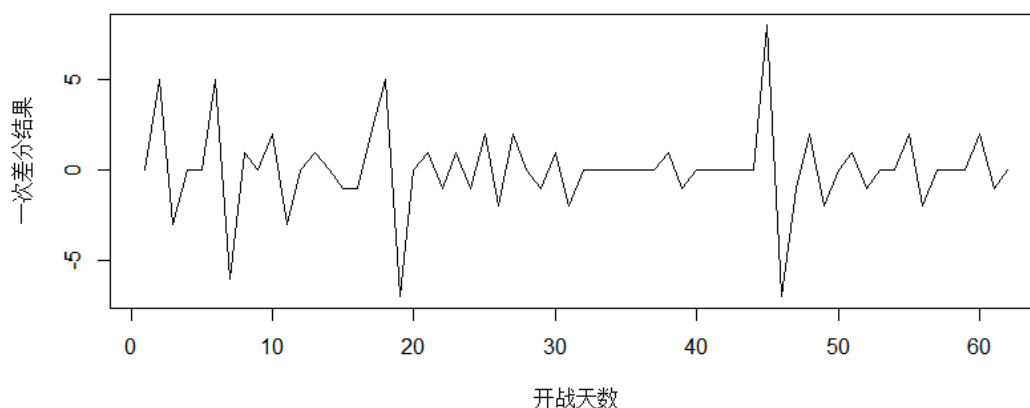


图 3: 二次差分序列

经过二次差分后，我们可以看到此时的时间序列图已经呈现在 0 附近上下波动的状态。此时，我们认为平稳模型是恰当的。

2.2 模型介数选择

接下来，对于 ARMA 模型：

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \cdots - \theta_q e_{t-q} \quad (2)$$

我们尝试确定它的阶数 p , q ，我们可以根据 ACF, PACF 和 EACF 图来进行抉择。我们首先来看 ACF 图。

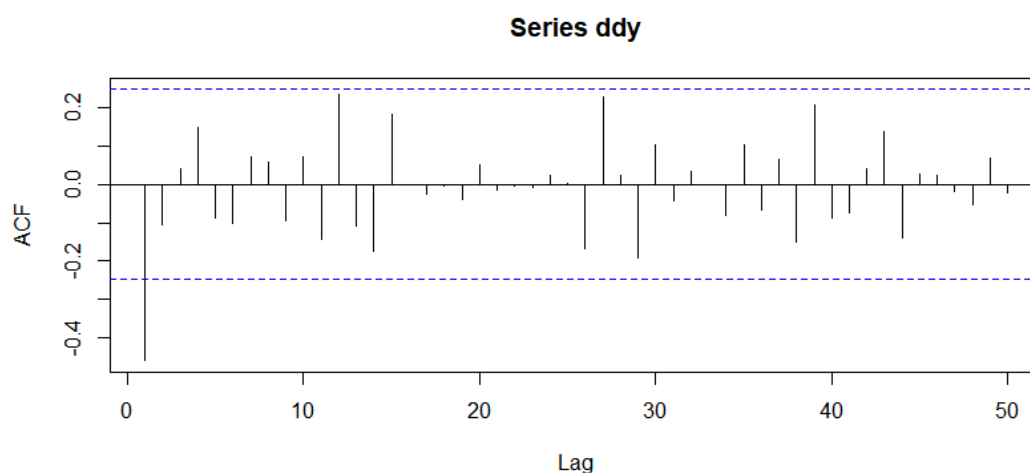


图 4: ACF 图

当我们使用简单标准误差作为临界值时，从 ACF 图上我们可以看到，经过两次差分后的时间序列的自相关系数值，从滞后一节开始都都在临界范围内，可以认为是 0。

当然，他们在一定程度上也呈现阻尼震荡的趋势。然而一阶处的自相关系数显著的不为 0，在整体上呈现出与 MA(1) 模型相似的一阶截尾的状态。

接着，我们来看它的 PACF 图。

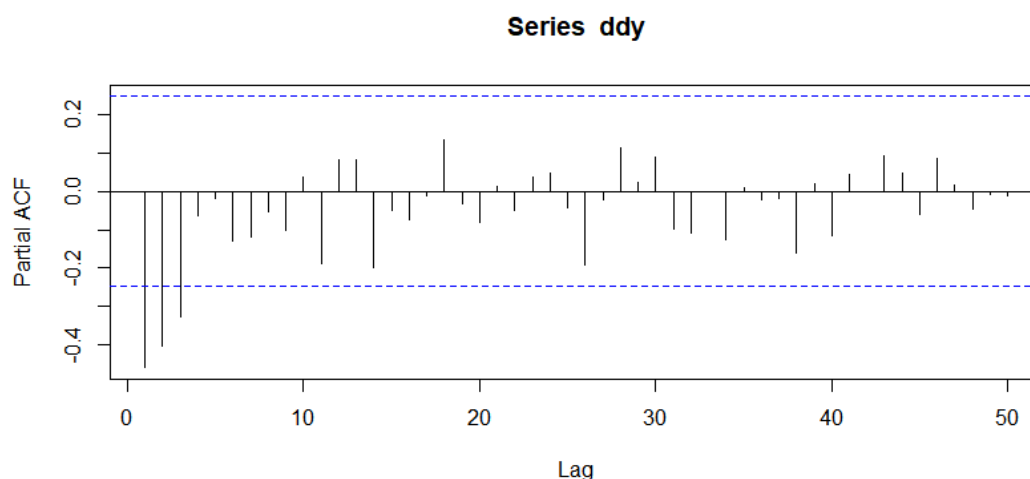


图 5: PACF 图

从 PACF 图上，我们可以看出 PACF 值呈现递减趋势，随着滞后介数的增加，快速衰减至 0(可以看到，在三阶滞后的 PACF 值已经处于临界范围内)

这也是符合 MA(1) 模型的特点。

接下来，我们通过参考 EACF 图来辅助判断。

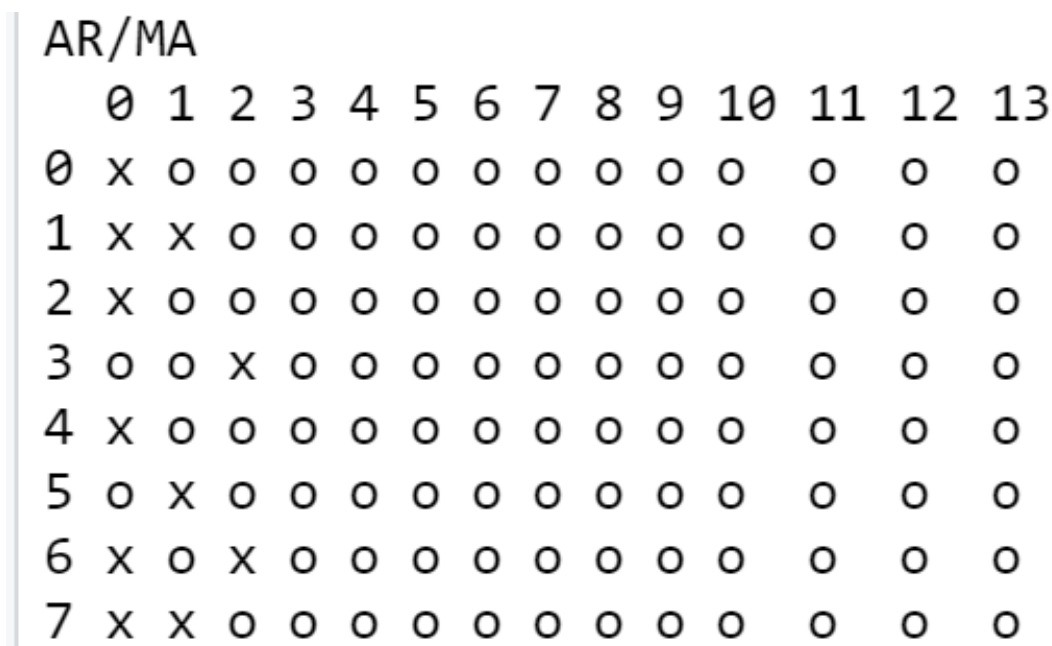


图 6: EACF 图

从图上我们可以明显的看出，EACF 图也建议我们对两次差分后的时间序列建立一个 MA(1) 模

型。

当然，我们也可以通过 TSA 包中的 `auto.arima` 模型对 `arima(p,d,q)` 的各种阶数的组合的情况进行汇总。

```
> auto.arima(y, trace = T)

ARIMA(2,2,2) : 256.0271
ARIMA(0,2,0) : 285.5015
ARIMA(1,2,0) : 273.0214
ARIMA(0,2,1) : 250.4061
ARIMA(1,2,1) : 252.5736
ARIMA(0,2,2) : 252.5568
ARIMA(1,2,2) : 254.8476

Best model: ARIMA(0,2,1)
```

图 7: 各种阶数组合

可以看到, 以 AIC 值为标准, 在众多阶数组合中, MA(1) 模型具有最小的 AIC 值。同时, MA(1) 模型也符合用尽量简洁的模型描述数据这一原则。

2.3 参数估计与拟合

由上文所述, 我们欲建立一个 MA(1) 模型。那么对于 MA(1) 模型的参数估计, 矩估计是不合适。因此我们尝试使用最小二乘估计和极大似然估计。我们使用 TSA 包中的 ARIMA 函数来计算这个过程。

表 1: MA(1) 模型参数估计

	条件 SS 估计	无条件 SS 估计	极大似然估计
估计值	-0.8453	-0.9080	-0.9080
s.e.	0.0773	0.0531	0.0531

我们可以看到, 三种估计之间相差都不大。我们不妨选择极大似然估计作为估计值。因此我们最后建立的是参数 θ 为: -0.908 的 ARIMA(0,1,1) 模型。

3 模型诊断

现在我们来进行模型诊断。首先, 我们来计算残差:

$$\text{残差} = \text{实际值} - \text{预测值} \quad (3)$$

按照我们所做的假设, 残差应该近似于白噪声的性质。

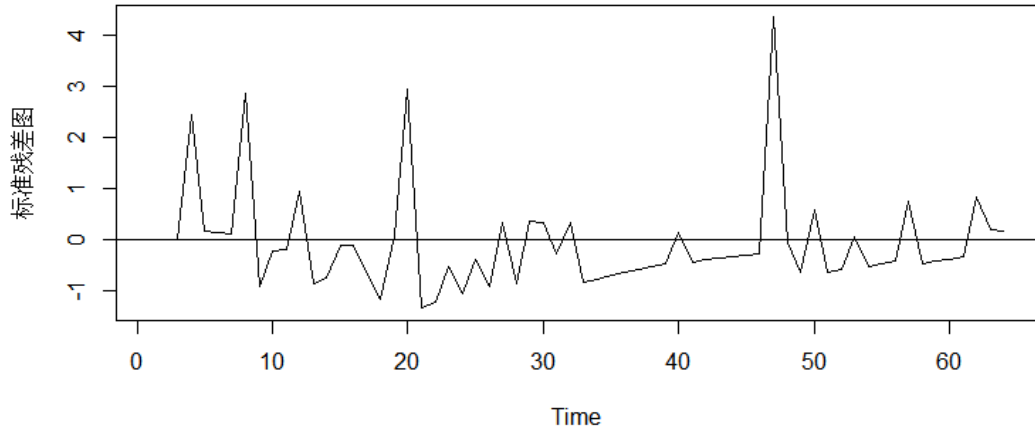


图 8: 标准残差图

从图中我们可以看出，整个残差序列有些奇怪。在四个点（开战后第 4，8，20，47 天）明显的接近或超过 3，这在标准正态分布中是不正常的。

接着我们来残差的自相关图：

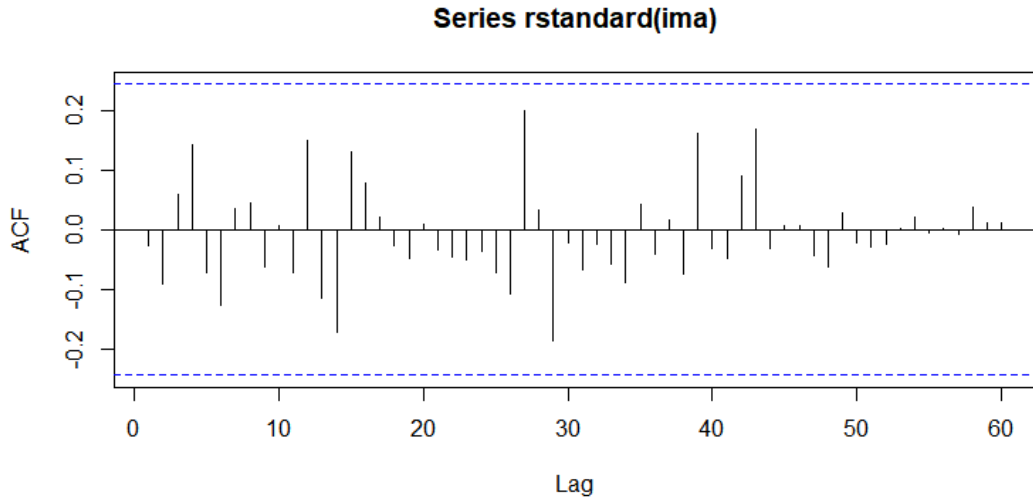


图 9: 残差的 ACF 图

在 ACF 图中我们可以明显的看出滞后各个阶数的残差都明显的小于临界值。于是我们有理由相信这是独立的。

当然，除了观察各个单独的滞后处的残差的相关系数，将这些相关系数的值作为一个组来进行检验也是有用的。我们使用 LJUNG-BOX 统计量：

$$Q_* = n(n+1) \left(\frac{\hat{r}_1^2}{n-1} + \cdots + \frac{\hat{r}_K^2}{n-K} \right) \quad (4)$$

在原假设为独立分布下，有限样本的 Q_* 的分布可以参阅 LJUNG 和 BOX(1978) 以及 DAVIES, TRIGGS 和 NEWBOLD(1977)。我们使 TSA 中的 `tsdiag` 函数来计算各个 K 值时的 p 值。

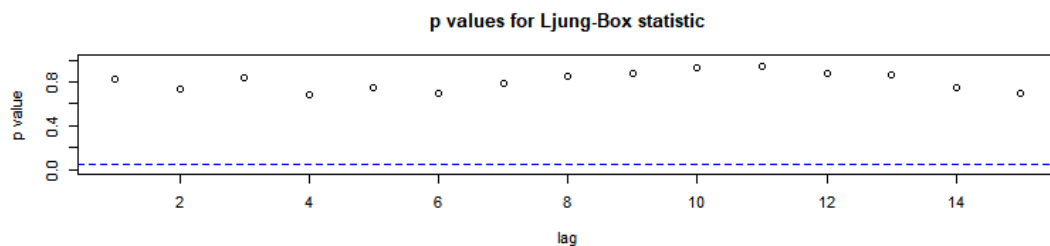


图 10: LJUNG-BOX 统计量的 P 值

显然，我们无法拒绝误差项的独立性。

但是，我们不能忽略在图 8 中发现的四个异常值，我们画出残差的 QQ 图：

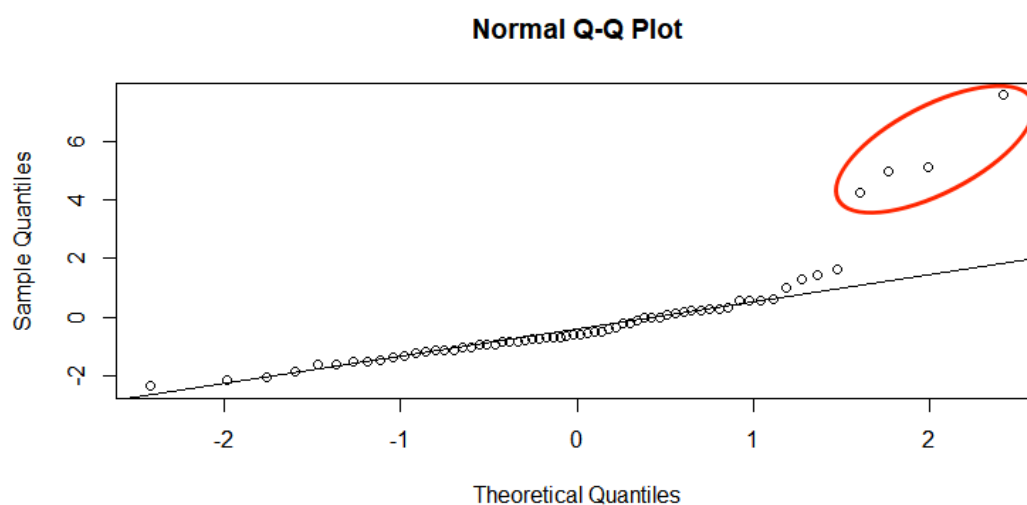


图 11: 残差的 QQ 图

我们可以看出这四个点发生了明显的偏离现象。

因此，我们的时间序列模型虽然在大部分点都取得了不错的效果，但是在某些个别点上仍然有一些值得进一步思考的地方。

4 问题分析与模型改进

正如我们在上文中提到的，有四个点明显不符合我们残差为白噪声的假设：他们分别发生在开战后第 4, 8, 20, 47 天，即 (2/28, 3/4, 3/16, 4/12) 我们来观察俄乌战争中的飞机（不包含无人机）损失数：

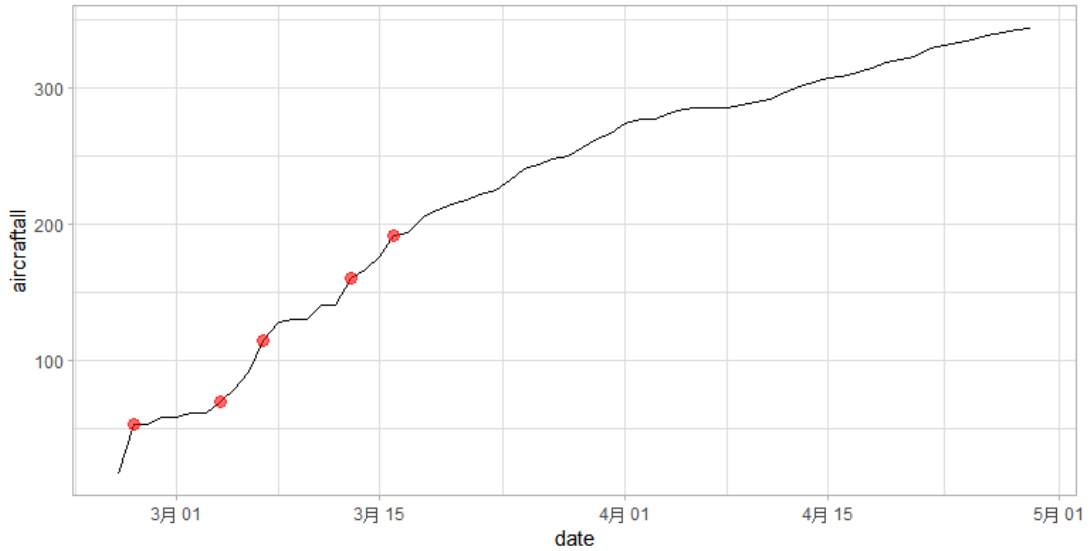


图 12: 飞机损失数

在图 12 中，我们可以看到在 2/28, 3/4, 3/16 这几天左右，出现了飞机损失数陡增的现象。而在 4/12 号的战报上，我们可以看到俄罗斯国防部发言人伊戈尔在记者会上说：俄军使用“口径”海基高精度导弹摧毁了 4 套 S-300 防空导弹系统，在第聂伯罗郊区摧毁了欧洲提供的 S-300 防空导弹系统。

所以，我们有理由猜测在这几天发生的残差值异常是由于军事冲突的突然升级，一些空军战术决策导致的。

而这种因素，我们是很难从防空炮损失数量自身的时间序列中挖掘到的，我们需要通过对其他的时间序列，如飞机损失数，整个冲突激烈程度等时间序列中提取相关因素，来建立起一个干预函数 m_t

从而对我们的时间序列做出修正为：

$$Y_t = m_t + N_t \quad (5)$$

其中， N_t 为一般的 ARIMA 过程。

5 总结

本文中，我们使用了二次差分来获得一个平稳的时间序列，然后对其建立 AARIMA(0,2,1) 模型，并对模型做出诊断。整个模型对时间序列的拟合得到了比较好的效果，但是在个别点发成了残差异常大的现象。我们通过分析飞机损害数量的时序图和一些战场简报，认为这是由于一些战场决策导致的。并提出了从飞机损失数量等时序图中提取信息来构造干预函数来修正时间序列模型的改进意见。

参考文献

- [1] Jonathan D. Cryer, Kung-Sik Chan. Time Series Analysis with Applications in R