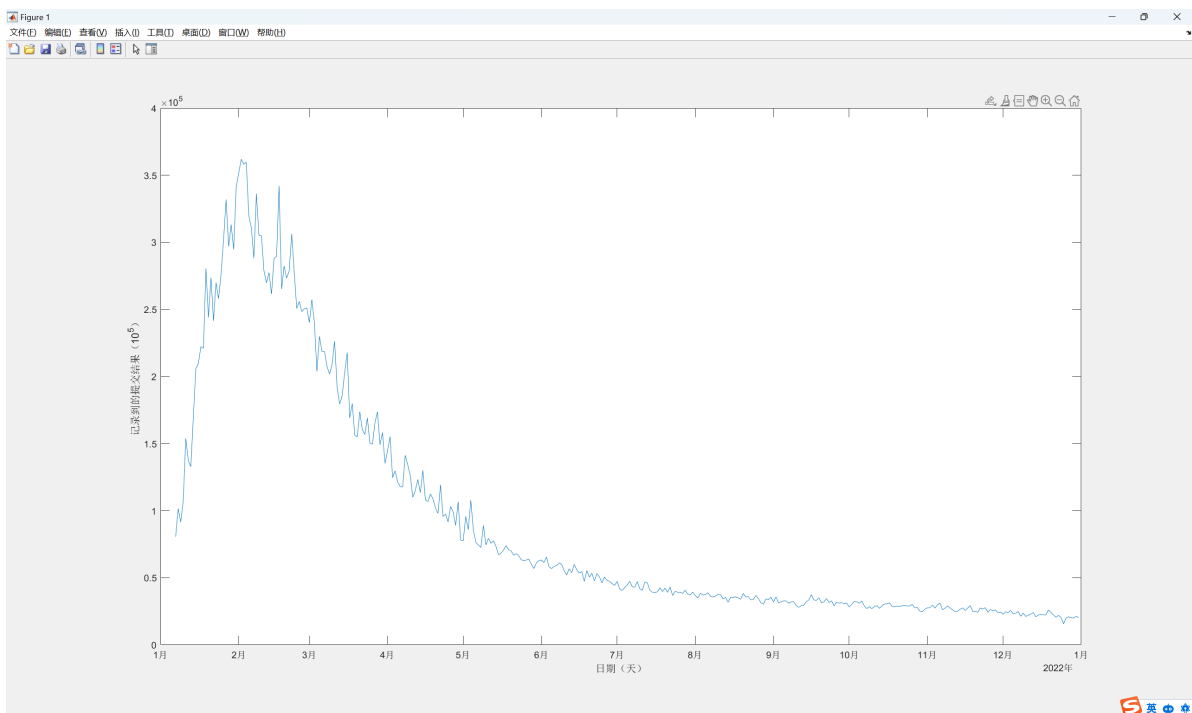
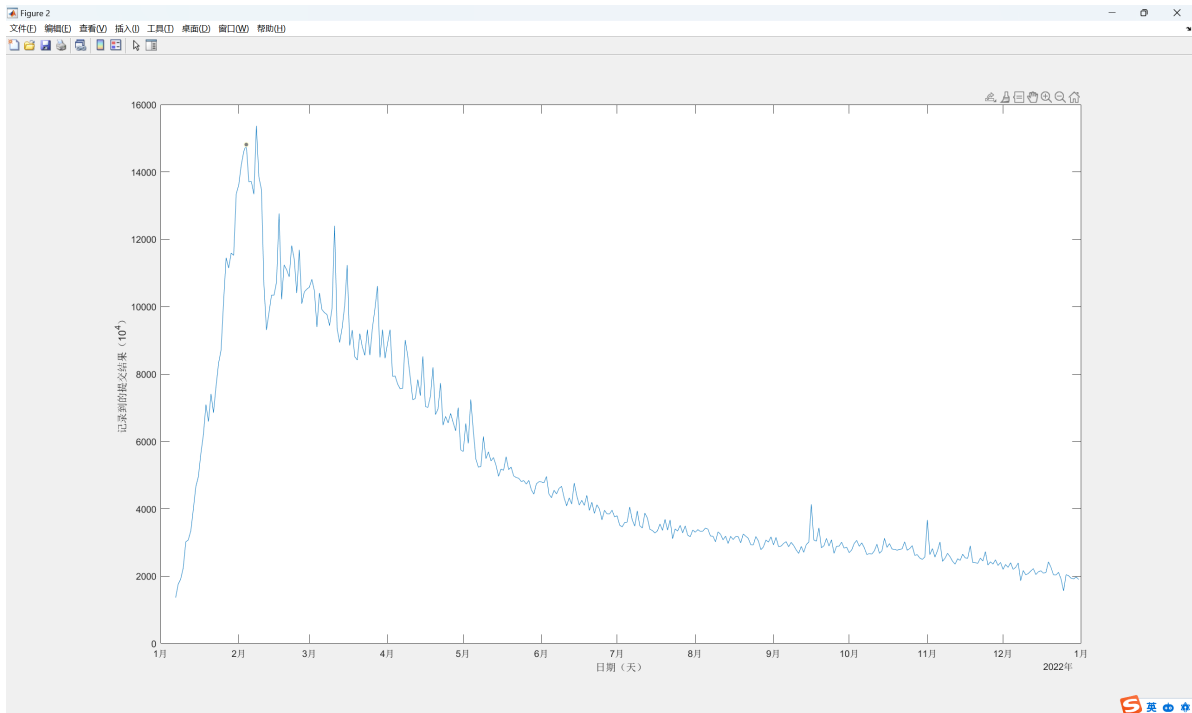


第一问模型判断：

首先清洗了数据，在总的提交次数和困难模式提交次数中各发现一个明显的异常值（总的在2022-11-30，困难的在2022-2-13），采用线性插值法替换了异常值



第一张是从2022-1-7到2022-12-31的困难模式提交次数为y轴，日期为x轴绘制的折线图，而第二张是从2022-1-7到2022-12-31的总的提交次数为y轴，日期为x轴绘制的折线图

首先，可以发现无论是困难提交还是总的提交，都呈现迅速增长——迅速下跌——缓慢下跌的趋势；而我们需要估计的2023年3月1日的数据显然处于“缓慢下跌”的时期，因此在建立模型时需要舍去前半部分的数据。

通过对图像的估计，选择了从5-22和7-1开始的两种数据

下面是5-22开始的数据的处理，这种数据不断下降，下降趋势总体变慢的模型看起来更可能是泊松模型或指数模型，先检验这两种模型是否可能符合其分布

首先考虑的是泊松模型来模拟，为此以SPSS为工具，采用单样本柯尔莫哥洛夫——斯米诺夫检验

NPar 检验

单样本柯尔莫戈洛夫-斯米诺夫检验			
		总的真实值	困难模式真实值
个案数		185	185
泊松参数 ^{a,b}	平均值	31174.62	2854.44
最极端差值	绝对	.503	.407
	正	.503	.365
	负	-.434	-.407
柯尔莫戈洛夫-斯米诺夫 Z		6.837	5.534
渐近显著性（双尾）		.000	.000
a. 检验分布为泊松分布。			
b. 根据数据计算。			

检验结果如下，渐进显著性为零，实际上不符合泊松模型

然后是指数模型：

► NPar 检验

单样本柯尔莫戈洛夫-斯米诺夫检验

		总的真实值	困难模式真实值
个案数		185	185
指数参数。 a,b	平均值	31174.62	2854.44
最极端差值	绝对	.468	.475
	正	.219	.243
	负	-.468	-.475
柯尔莫戈洛夫-斯米诺夫 Z		6.367	6.461
渐近显著性（双尾）		.000	.000

a. 检验分布为指数分布。

b. 根据数据计算。

渐进显著性依然为零，不符合指数模型

考虑到提交次数的数据记录的时间段为一年内，其中工作日和休息日可能对提交次数带来较大的影响，而且在总体下降的基础上数值的浮动也符合一定的周期性规律；因此猜测带有季节性因素的时间回归模型，即温特斯加法或乘法模型进行拟合的效果可能比较好

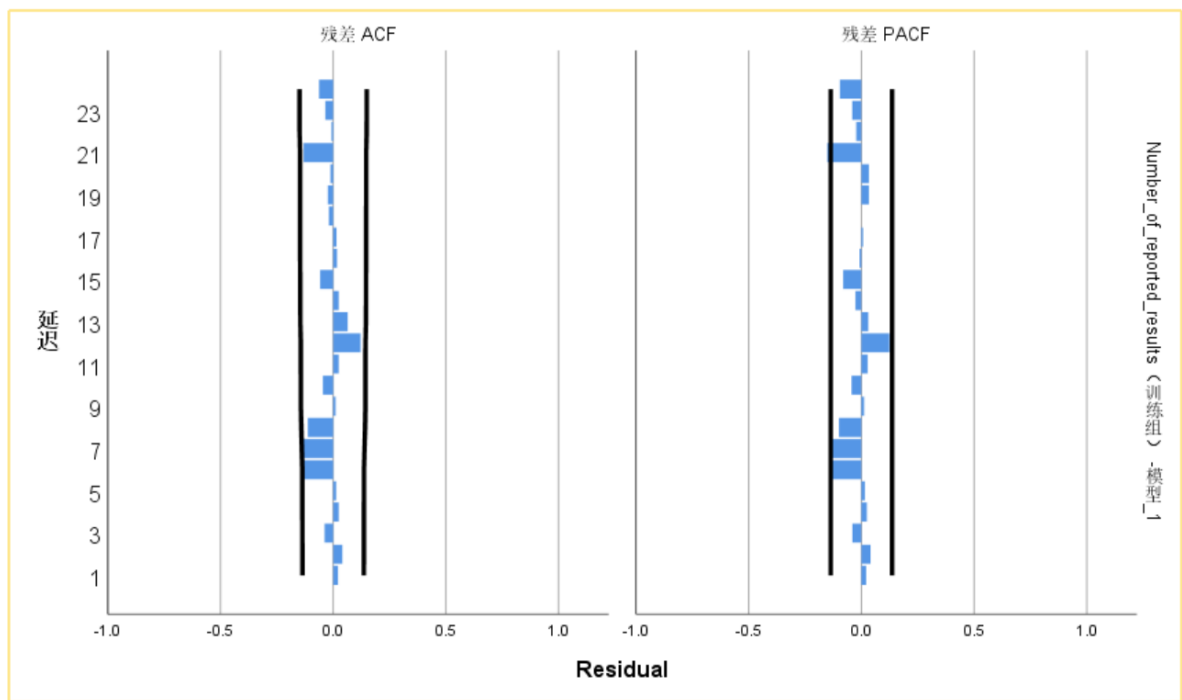
采用SPSS中“专家建模器”提供相对最为准确的模型，将12月2日以后的数据作为测试组，5-22到12-1的数据作为训练组得到的结果如下：

模型描述			
			模型类型
模型 ID	Number_of_reported_results (训练组)	模型_1	温特斯乘性

模型摘要

模型拟合度											
拟合统计	平均值	标准误差	最小值	最大值	百分位数						
					5	10	25	50	75	90	95
平稳 R 方	.676	.	.676	.676	.676	.676	.676	.676	.676	.676	.676
R 方	.973	.	.973	.973	.973	.973	.973	.973	.973	.973	.973
RMSE	2326.199	.	2326.199	2326.199	2326.199	2326.199	2326.199	2326.199	2326.199	2326.199	2326.199
MAPE	4.530	.	4.530	4.530	4.530	4.530	4.530	4.530	4.530	4.530	4.530
MaxAPE	17.667	.	17.667	17.667	17.667	17.667	17.667	17.667	17.667	17.667	17.667
MAE	1764.258	.	1764.258	1764.258	1764.258	1764.258	1764.258	1764.258	1764.258	1764.258	1764.258
MaxAE	8810.292	.	8810.292	8810.292	8810.292	8810.292	8810.292	8810.292	8810.292	8810.292	8810.292
正态化 BIC	15.581	.	15.581	15.581	15.581	15.581	15.581	15.581	15.581	15.581	15.581

模型统计						
模型	预测变量数	模型拟合度统计		杨-博克斯 Q(18)		
		平稳 R 方	统计	DF	显著性	离群值数
Number_of_reported_results (训练组) - 模型_1	0	.676	17.385	15	.296	0



可以看到训练组得到的模型正是温特斯乘法模型，残差进行q检验的显著值大于0.05，通过白噪声检验；所有滞后阶数的自相关系数和偏自相关系数均和0没有显著差异。



此图是将测试组和上述模型得到的预测值比较的结果，可以看到预测的数据在保留了原始序列一定程度的季节效应的同时具有向上的线性趋势，因此可以认为温特斯乘法模型符合总提交次数的估计

接下来利用5-22至12-31的数据来估计2023-3-1的总提交次数，利用SPSS建立温特斯乘法模型，结果如下：

→ 时间序列建模器

模型描述

			模型类型
模型 ID	Number_of_reported_results	模型_1	温特斯乘法

模型摘要

模型拟合度

拟合统计	平均值	标准误差	最小值	最大值	百分位数						
					5	10	25	50	75	90	95
平稳 R 方	.670	.	.670	.670	.670	.670	.670	.670	.670	.670	.670
R 方	.976	.	.976	.976	.976	.976	.976	.976	.976	.976	.976
RMSE	2256.445	.	2256.445	2256.445	2256.445	2256.445	2256.445	2256.445	2256.445	2256.445	2256.445
MAPE	4.682	.	4.682	4.682	4.682	4.682	4.682	4.682	4.682	4.682	4.682
MaxAPE	34.646	.	34.646	34.646	34.646	34.646	34.646	34.646	34.646	34.646	34.646
MAE	1694.675	.	1694.675	1694.675	1694.675	1694.675	1694.675	1694.675	1694.675	1694.675	1694.675
MaxAE	8980.842	.	8980.842	8980.842	8980.842	8980.842	8980.842	8980.842	8980.842	8980.842	8980.842
正态化 BIC	15.512	.	15.512	15.512	15.512	15.512	15.512	15.512	15.512	15.512	15.512

模型统计

		模型拟合度统计	杨-博克斯 Q(18)			
模型	预测变量数	平稳 R 方	统计	DF	显著性	离群值数
Number_of_reported_results-模型_1	0	.670	17.446	15	.293	0

指数平滑法模型参数

模型		估算	标准误差	t	显著性	
Number_of_reported_res ults-模型_1	不转换	Alpha (水平)	.223	.038	5.864	.000
		Gamma (趋势)	.058	.021	2.796	.006
		Delta (季节)	.025	.017	1.441	.151

$$\left\{ \begin{array}{l} l_t = \alpha \frac{x_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (\text{水平平滑方程}) \\ b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (\text{趋势平滑方程}) \\ s_t = \gamma \frac{x_t}{l_{t-1} + b_{t-1}} + (1 - \gamma)s_{t-m} \quad (\text{季节平滑方程}) \\ \hat{x}_{t+h} = (l_t + hb_t)s_{t+h-m(k+1)}, k = [\frac{h-1}{m}] \quad (\text{预测方程}) \end{array} \right.$$

m : 周期长度（月度数据取12，季度数据取4）

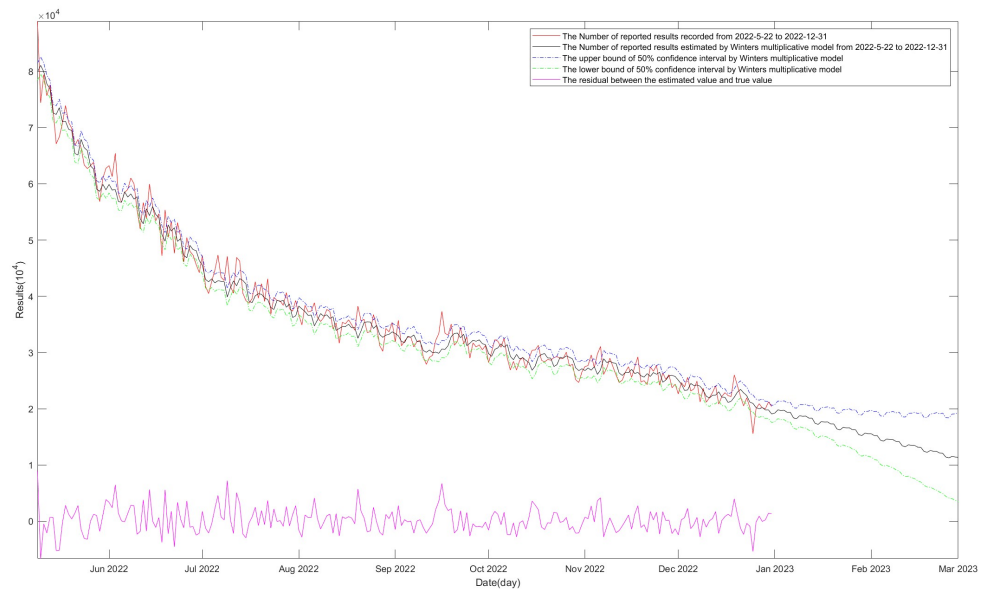
α : 水平的平滑参数

β : 趋势的平滑参数

γ : 季节的平滑参数

\hat{x}_{t+h} : 第 h 期的预测值

模型的拟合优度看“模型摘要”中的 R^2 ，在置信区间为50%时能够得到相对合理的上下界（置信区间较大时下界会出现负数）；答案就是最后一个表格中的阿尔法对应公式阿尔丁，Gamma对应公式贝塔，Delta对应公式的伽马



3月1日的总的提交数置信区间上界为19280.1，下界为3457.5

5-22开始的困难：

单样本柯尔莫戈洛夫-斯米诺夫检验

		总的真实值	困难模式真实值
个案数		185	185
泊松参数 ^{a,b}	平均值	31174.62	2854.44
最极端差值	绝对	.503	.407
	正	.503	.365
	负	-.434	-.407
柯尔莫戈洛夫-斯米诺夫 Z		6.837	5.534
渐近显著性（双尾）		.000	.000

a. 检验分布为泊松分布。

b. 根据数据计算。

NPAR TESTS

/K-S(EXPONENTIAL)=总的真实值 困难模式真实值
/MISSING ANALYSIS.

NPar 检验

单样本柯尔莫戈洛夫-斯米诺夫检验

		总的真实值	困难模式真实值
个案数		185	185
指数参数。 ^{a,b}	平均值	31174.62	2854.44
最极端差值	绝对	.468	.475
	正	.219	.243
	负	-.468	-.475
柯尔莫戈洛夫-斯米诺夫 Z		6.367	6.461
渐近显著性（双尾）		.000	.000

a. 检验分布为指数分布。

b. 根据数据计算。

7-1

单样本柯尔莫戈洛夫-斯米诺夫检验

		Number_of_reported_results	Number_in_hard_mode
个案数		185	185
泊松参数 ^{a,b}	平均值	31174.62	2854.44
最极端差值	绝对	.503	.407
	正	.503	.365
	负	-.434	-.407
柯尔莫戈洛夫-斯米诺夫 Z		6.837	5.534
渐近显著性（双尾）		.000	.000

a. 检验分布为泊松分布。

b. 根据数据计算。

单样本柯尔莫戈洛夫-斯米诺夫检验 2

		Number_of_reported_results	Number_in_hard_mode
个案数		185	185
指数参数。 ^{a,b}	平均值	31174.62	2854.44
最极端差值	绝对	.468	.475
	正	.219	.243
	负	-.468	-.475
柯尔莫戈洛夫-斯米诺夫 Z		6.367	6.461
渐近显著性（双尾）		.000	.000

a. 检验分布为指数分布。

b. 根据数据计算。

这个是SPSS的科尔莫戈洛夫-斯米诺夫检验结果，可见无论是总的还是困难，都不符合科尔莫哥洛夫——斯米诺夫检验

这个温特斯乘法的模型我自己都不信，个人还用7-1之后的数据跑，结果跑出来温特斯加法的，可信度好很多)