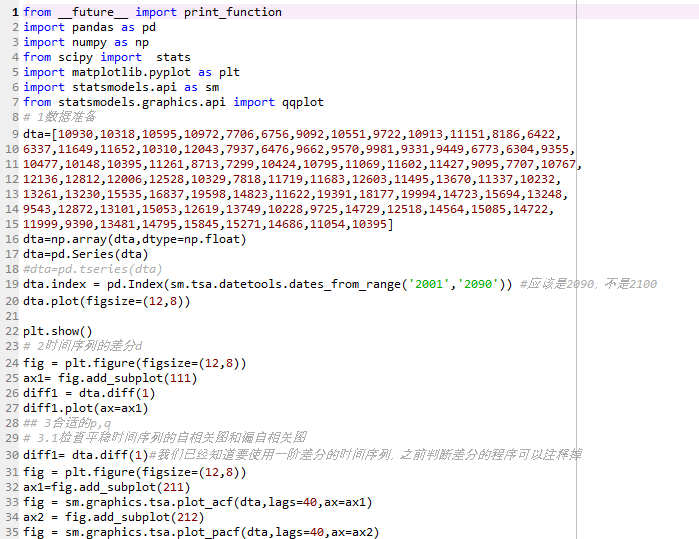
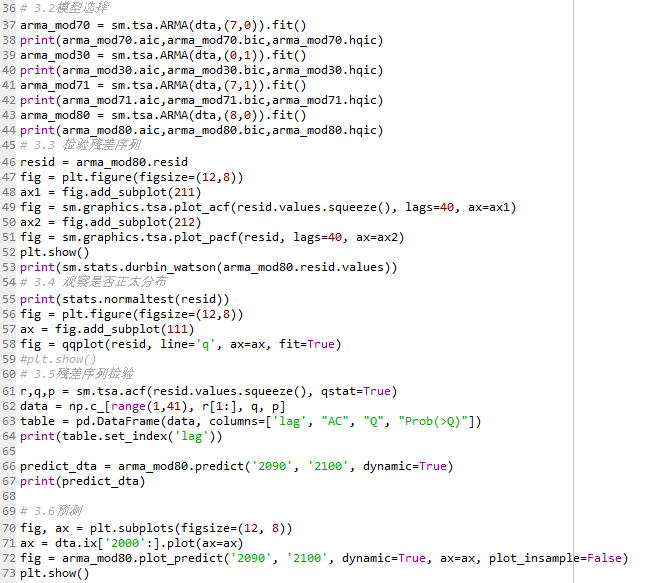
通过下图所示代码，我们希望能通过ARIMA即时间序列分析模型解决预测问题，这是一个极其经典的通过差分、自回归、滑动平均问题从而解决预测问题的机器学习或数据挖掘算法。

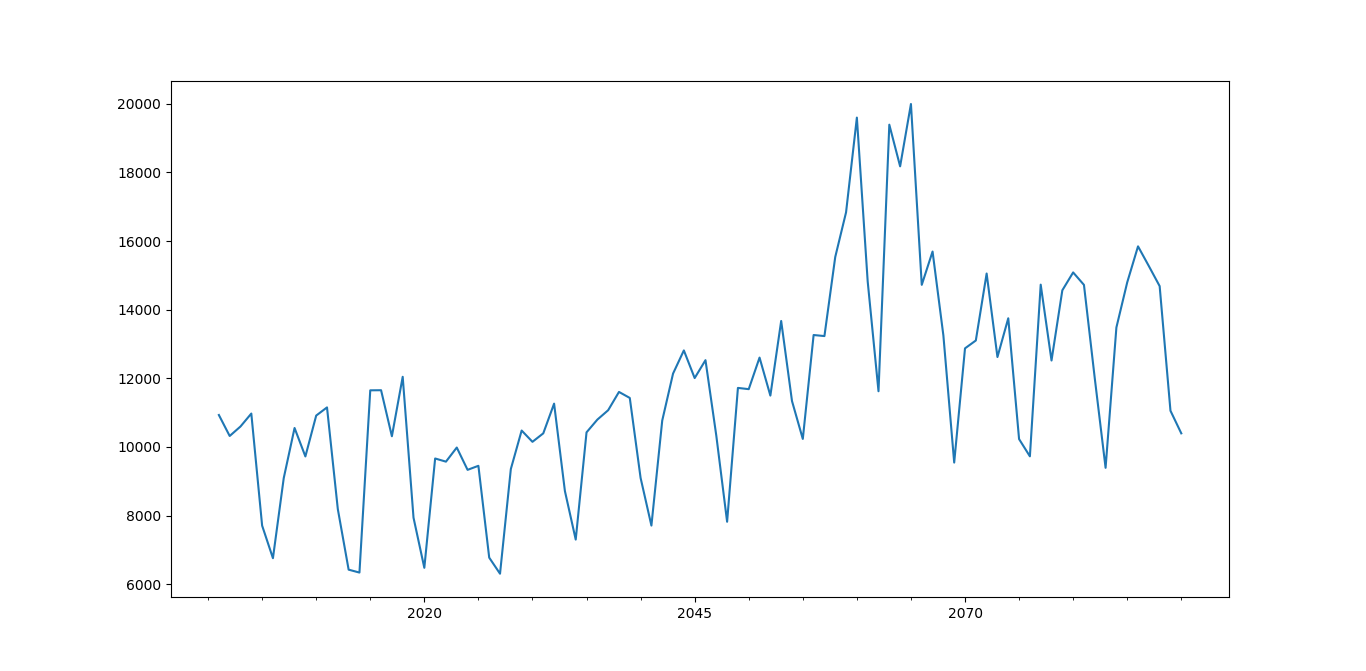
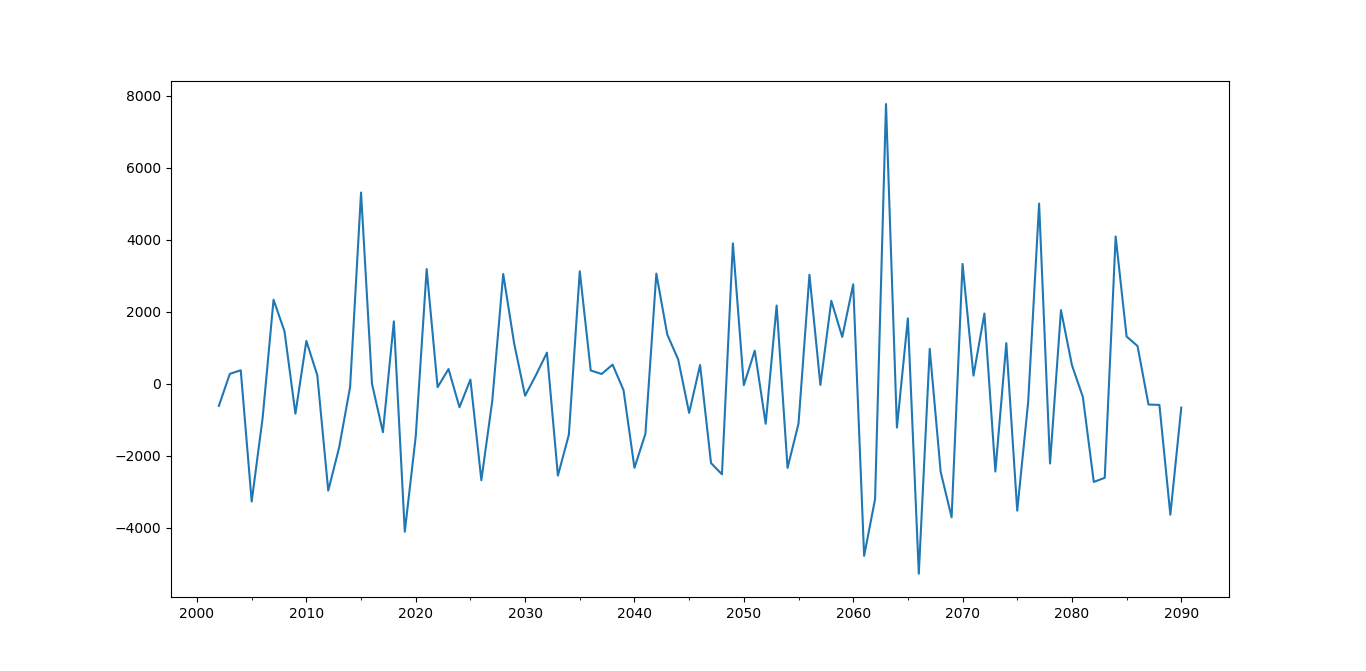
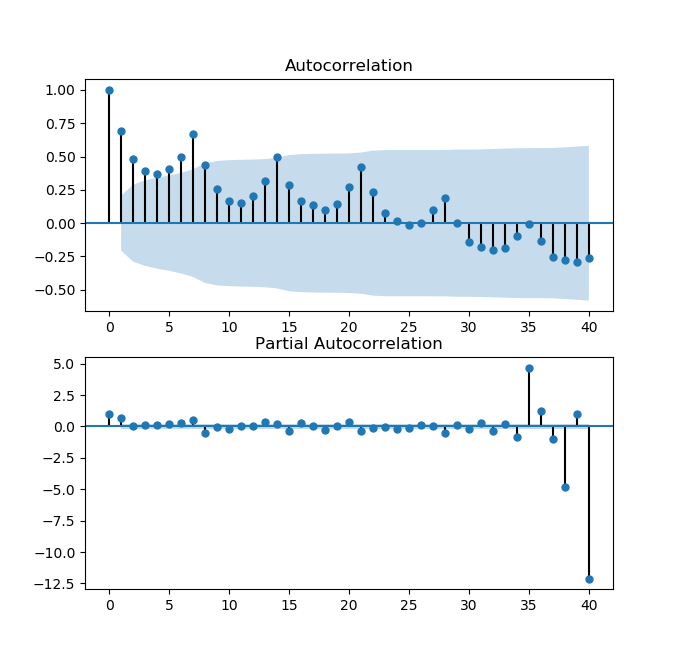
C:\Users\Administrator\Desktop\homework\ARIMA\0.jpg

这是ARIMA的算法公式，即可表示为ARIMA(p,d,q)，AR是"自回归"，p为自回归项数；MA为"滑动平均"，q为滑动平均项数，d为使之成为平稳序列所做的差分次数（阶数）。

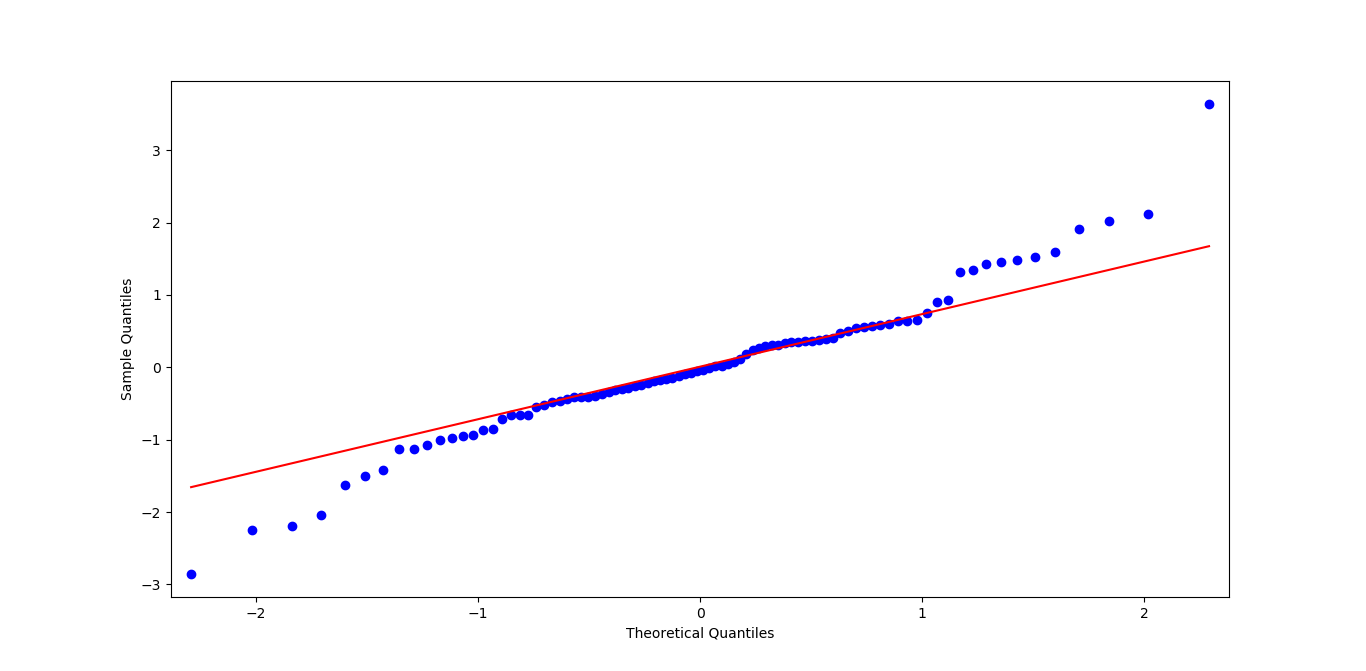
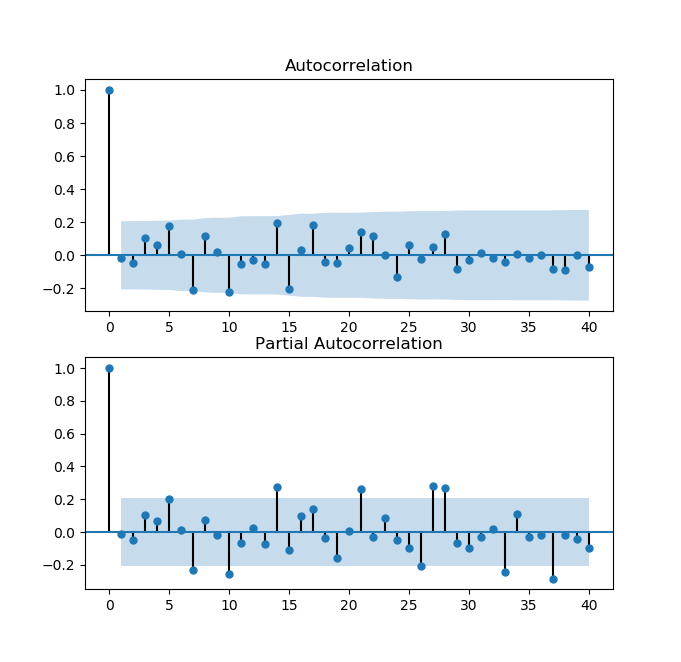




代码步骤如下：

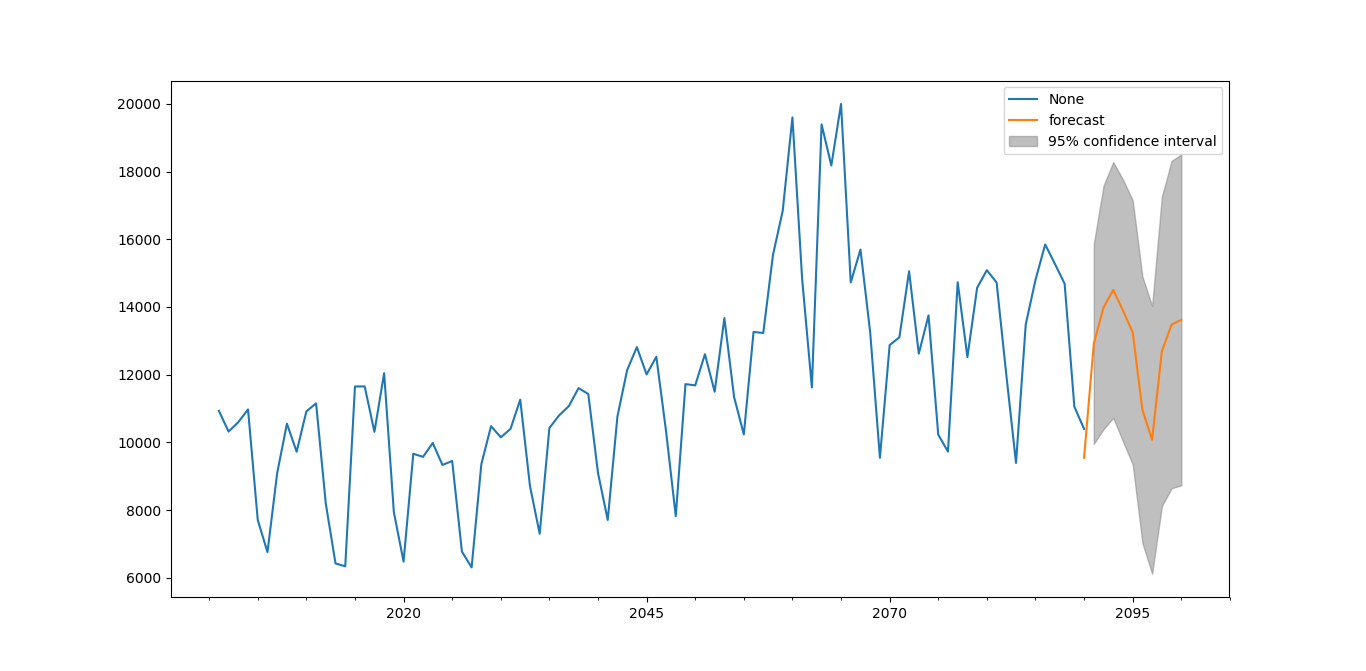
1. 根据[时间序列](https://baike.baidu.com/item/%E6%97%B6%E9%97%B4%E5%BA%8F%E5%88%97" \t "_blank)的散点图、自相关函数和偏自相关函数图识别其平稳性。我们这里是通过matplotlib.pyplot将这些数据通过12\*8的折线图展示出来，当然数据都是需要我们进行了预处理出来的。
2. 对非平稳的时间序列数据进行平稳化处理。直到处理后的[自相关函数](https://baike.baidu.com/item/%E8%87%AA%E7%9B%B8%E5%85%B3%E5%87%BD%E6%95%B0" \t "_blank)和[偏自相关函数](https://baike.baidu.com/item/%E5%81%8F%E8%87%AA%E7%9B%B8%E5%85%B3%E5%87%BD%E6%95%B0)的数值非显著非零。这里我们给数据进行了一阶差分的转化，即diff1 = dta.diff(1)，然后通过matplotlib得到了一个一重差分的平稳时间序列。
3. 根据所识别出来的特征建立相应的时间序列模型。平稳化处理后，若[偏自相关函数](https://baike.baidu.com/item/%E5%81%8F%E8%87%AA%E7%9B%B8%E5%85%B3%E5%87%BD%E6%95%B0)是[截尾](https://baike.baidu.com/item/%E6%88%AA%E5%B0%BE)的，而[自相关函数](https://baike.baidu.com/item/%E8%87%AA%E7%9B%B8%E5%85%B3%E5%87%BD%E6%95%B0)是[拖尾](https://baike.baidu.com/item/%E6%8B%96%E5%B0%BE)的，则建立[AR模型](https://baike.baidu.com/item/AR%E6%A8%A1%E5%9E%8B)；若[偏自相关函数](https://baike.baidu.com/item/%E5%81%8F%E8%87%AA%E7%9B%B8%E5%85%B3%E5%87%BD%E6%95%B0)是[拖尾](https://baike.baidu.com/item/%E6%8B%96%E5%B0%BE)的，而[自相关函数](https://baike.baidu.com/item/%E8%87%AA%E7%9B%B8%E5%85%B3%E5%87%BD%E6%95%B0)是[截尾](https://baike.baidu.com/item/%E6%88%AA%E5%B0%BE)的，则建立[MA模型](https://baike.baidu.com/item/MA%E6%A8%A1%E5%9E%8B)；若[偏自相关函数](https://baike.baidu.com/item/%E5%81%8F%E8%87%AA%E7%9B%B8%E5%85%B3%E5%87%BD%E6%95%B0)和[自相关函数](https://baike.baidu.com/item/%E8%87%AA%E7%9B%B8%E5%85%B3%E5%87%BD%E6%95%B0)均是[拖尾](https://baike.baidu.com/item/%E6%8B%96%E5%B0%BE)的，则序列适合[ARMA模型](https://baike.baidu.com/item/ARMA%E6%A8%A1%E5%9E%8B)。

这里我们没有进行分析，因为我们搜寻到的数据是极其符合ARMA模型的。

1. 参数估计，检验是否具有统计意义。通过我们对模型分析，这里选择了对数据分布进行正态分布检验，而我们看到的图形，点大部分分布在了线上，基本符合正态分布的分布规律。
2. 假设检验，判断（诊断）残差序列是否为[白噪声序列](https://baike.baidu.com/item/%E7%99%BD%E5%99%AA%E5%A3%B0%E5%BA%8F%E5%88%97" \t "_blank)。

通过残差prob值均大于0.05，所以残差序列不存在自相关性。

1. 利用已通过检验的模型进行预测。



阴影部分为我们对2091-2100年代的预测，验证了我们对ARIMA的建模是成功的。统计到预测数据如下：

2090-12-31     9541.934170

2091-12-31    12906.497016

2092-12-31    13979.448587

2093-12-31    14499.448647

2094-12-31    13892.217206

2095-12-31    13247.336277

2096-12-31    10958.837240

2097-12-31    10070.285452

2098-12-31    12680.448582

2099-12-31    13472.910370

2100-12-31    13611.974609

此为我们通过statsmodels（经常作为计量经济学和统计学的建模工具）包中带有的数据集分析，然后我们又找到了两组经典的数据集，然后尝试着解决新环境问题。

以下是我们通过对模型了解之后总结的思路：

(1) 获取被观测系统时间序列数据；  
(2) 对数据绘图，观测是否为平稳时间序列；对于非平稳时间序列要先进行d阶差分运算，化为平稳时间序列；  
(3) 经过第二步处理，已经得到平稳时间序列。要对平稳时间序列分别求得其自相关系数ACF 和偏自相关系数PACF，通过对自相关图和偏自相关图的分析，得到最佳的阶层 p 和阶数 q

(4) 由以上得到的d、q、p，得到ARIMA模型。然后开始对得到的模型进行模型检验。

首先我们找到了，此为经典的波士顿时间序列问题；其次我们又找到了国家统计年鉴问题的数据集。同理上述模型，然后我们将新的两组数据进行录入，就得到了我们对这两个新的实际问题的分析并建模从而得到的预测结果。