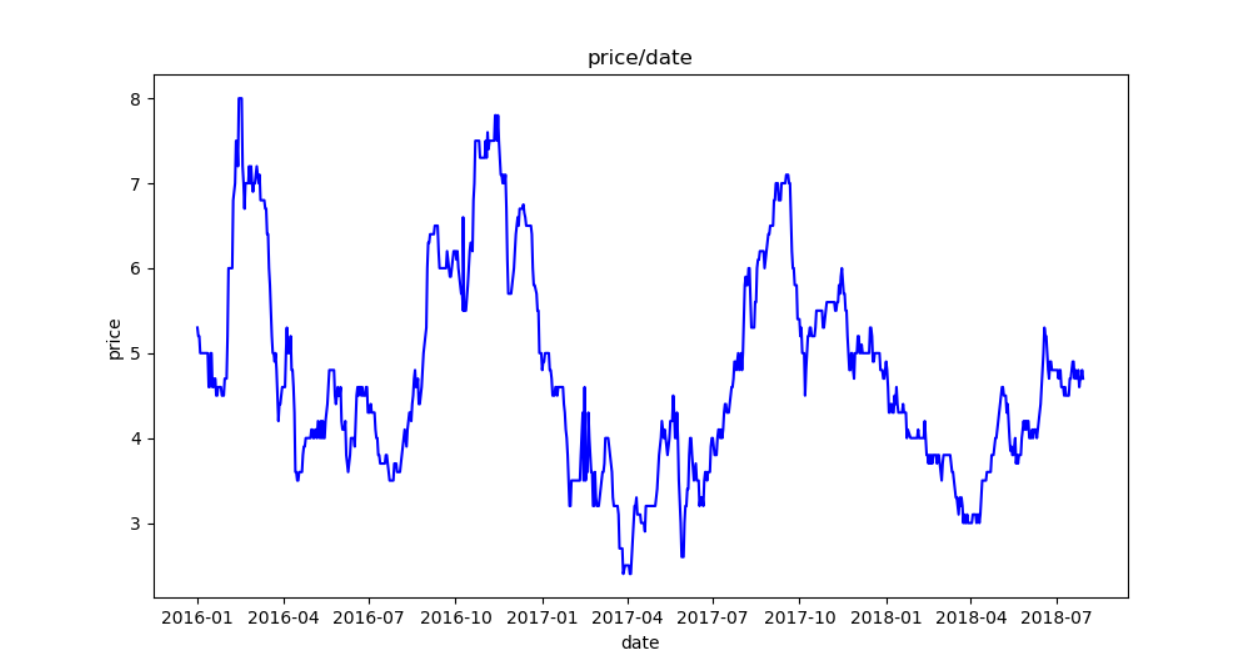
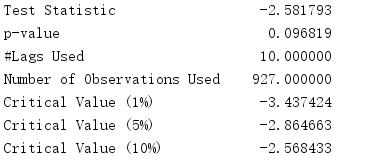
以大蒜价格为例，进行ARIMA时间序列的初步测试。

下面是没经过处理的价格曲线

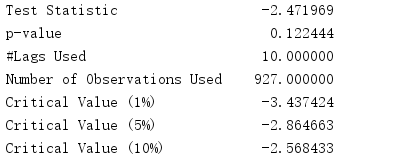


可以看出还是有一定趋势的，比如整年价格最高的时候集中在一年的十月左右，最低时是一年的4月左右。之后进行平稳性检验，采用ADF单位根检验方法，结果如下：

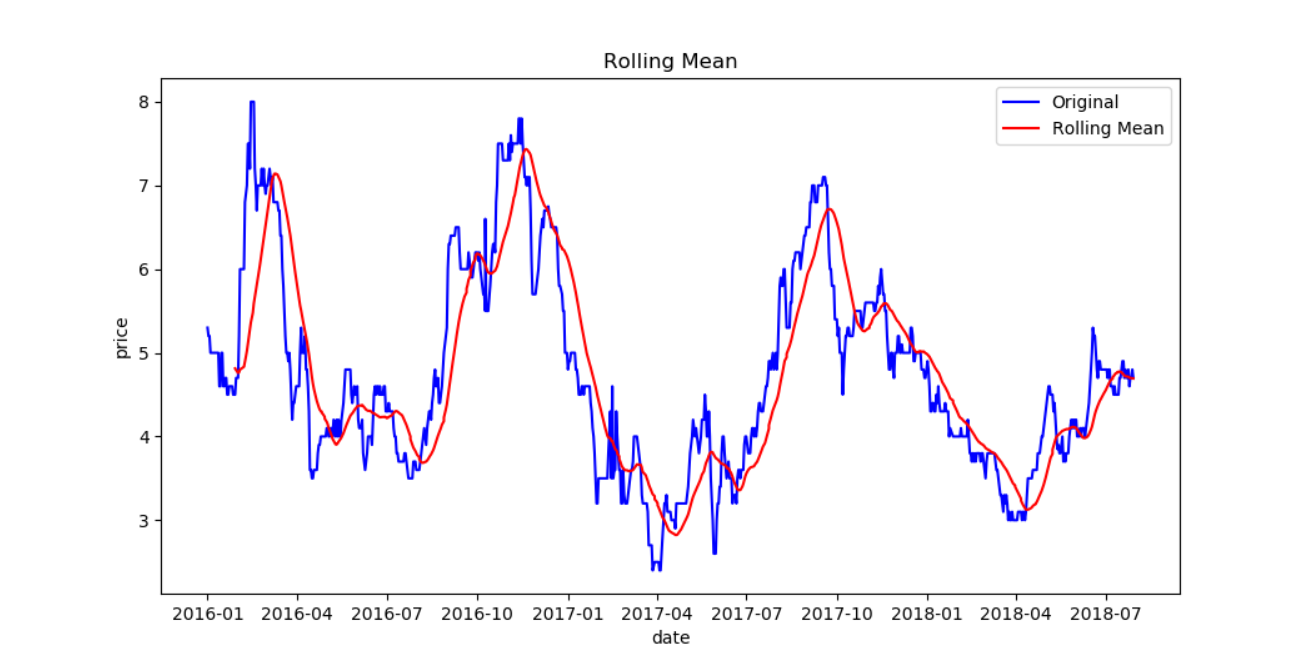


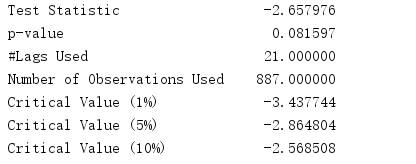
p值大于0.05，所以说明不能拒绝原假设，而原假设是序列具有单位根，即非平稳，故接下来需要做平稳性处理。

一开始用对数变化处理，结果不尽人意，如下：

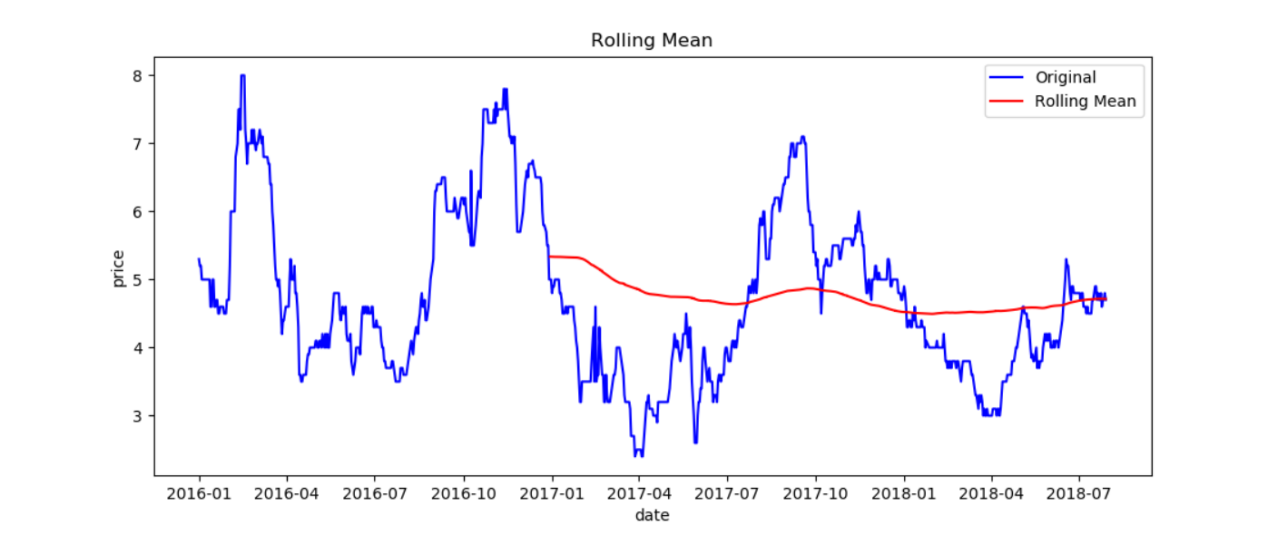


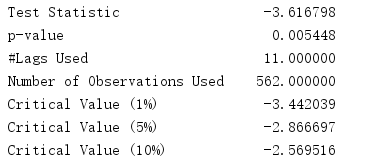
一开始认为可能具有月周期因素，以窗口为30的移动平均处理曲线如下：



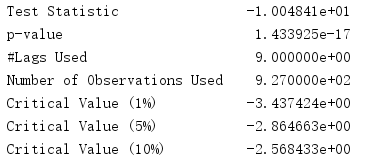


改为窗口为365的移动平均：



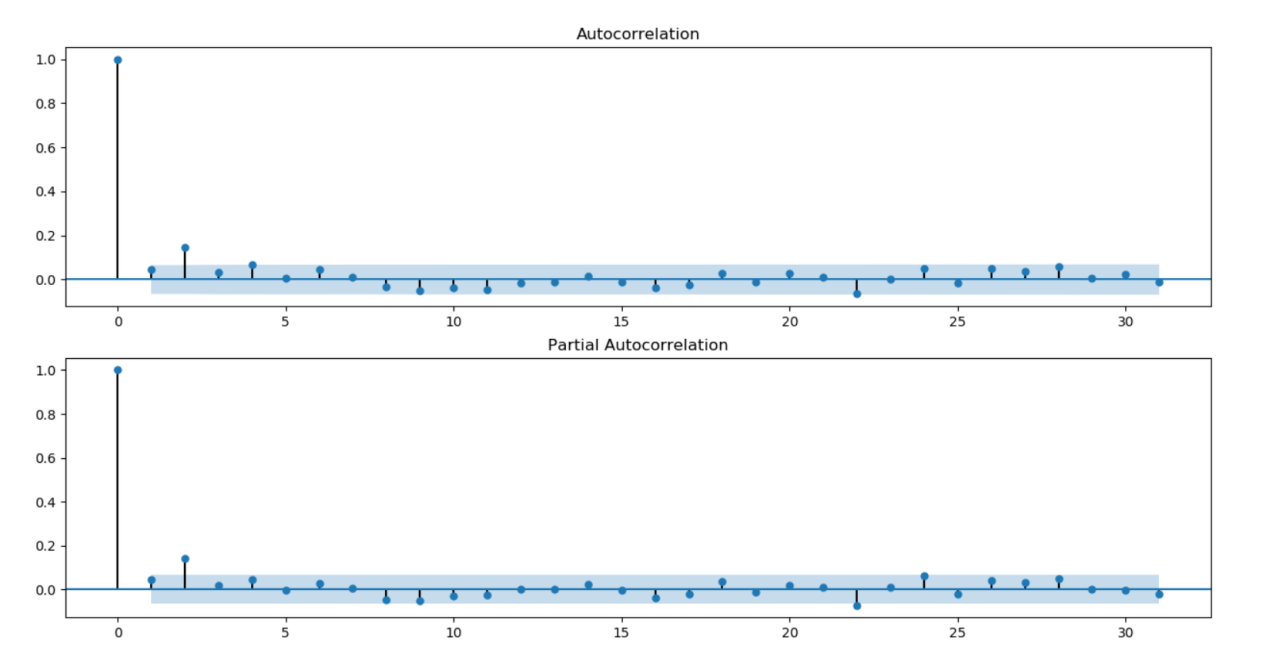


曲线变的十分平滑，故而，价格曲线数据可能是存在年周期性因素的，这里由于数据量原因，便暂且不考虑年周期性因素，以一阶差分对原数据进行处理，结果如下：



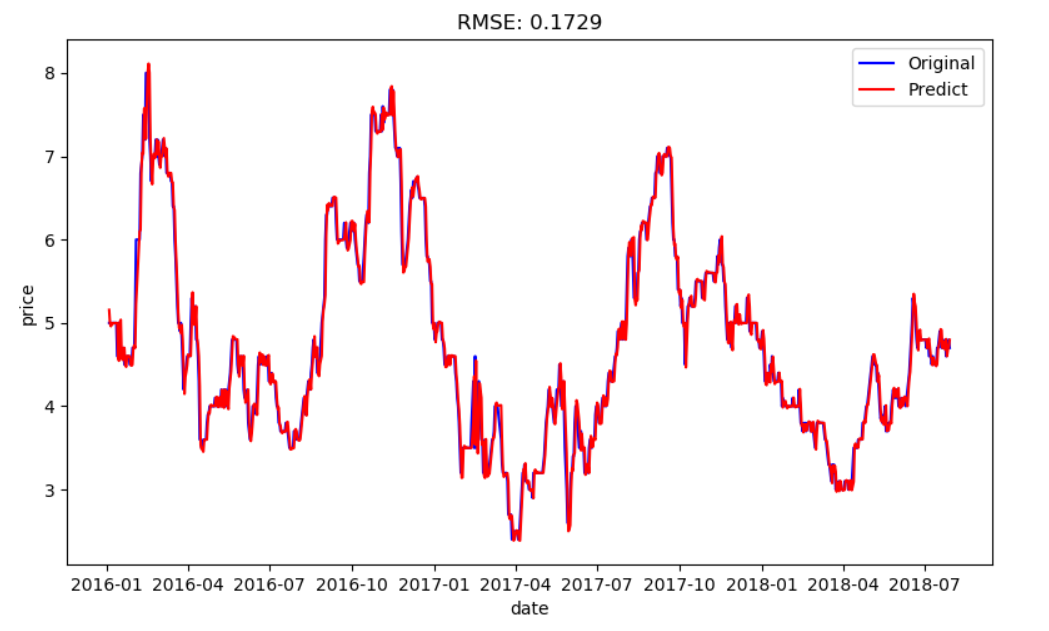
已符合平稳性条件。而这差分也正是ARIMA模型比ARMA模型多出的一个操作。

符合平稳性条件后，需要确定需要对模型定阶，即确定p、q的阶数，观察一阶差分后的数据的自相关和偏相关图如下：



很明显的自相关和偏自相关都是拖尾，且均为2阶拖尾（n从2开始缩至置信区间），故取p，q均为2。

可使用均方根误差（RMSE）来评估模型样本内拟合的好坏。利用该准则进行判别时，需要剔除“非预测”数据的影响。

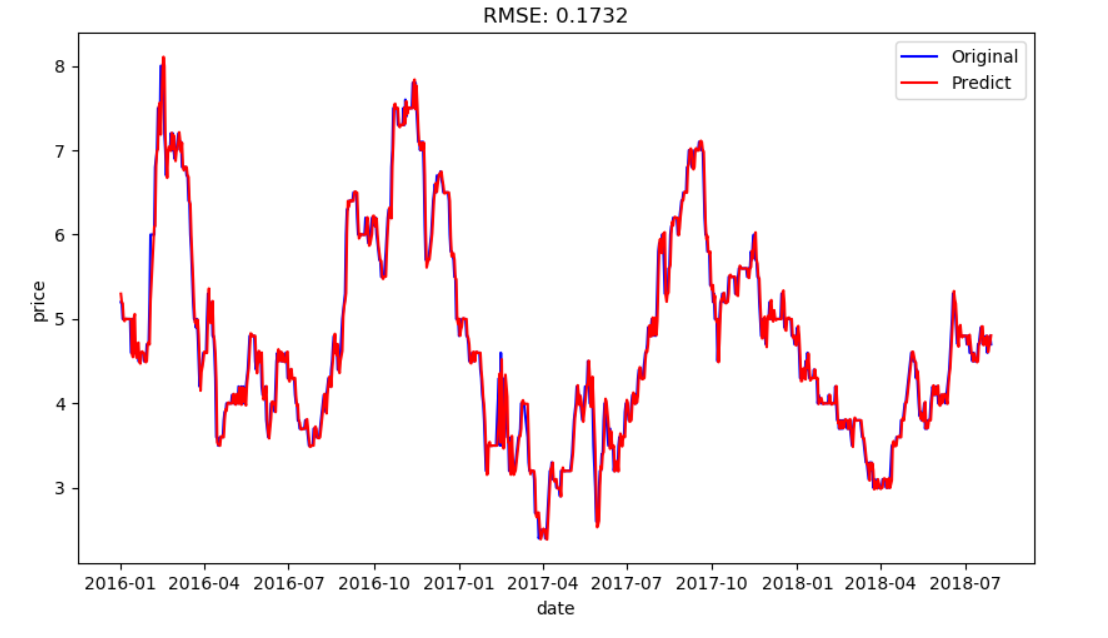


该图中RMSE极小，说明结果已经极为拟合，甚至有可能有过拟合的趋势，不过考虑到数据量的问题，还是可以理解。

p，q两个值的确定，也可以依据BIC准则识别模型的p, q值，通常认为BIC值越小的模型相对更优，综合考虑了残差大小和自变量的个数，残差越小BIC值越小，自变量个数越多BIC值越大，比如对一阶差分后的数据进行分析，最优解如下（依次为最优解的BIC值，p，q）：



即最优解的p，q为0，2，以这两个值重新拟合模型，得到如下：



RMSE的值相差不大。

之后若是进行预测的话，可以设置在添加新数据到达一定量，比如一周，一月，在这期间的预测值用同一个模型。之后再重新拟合一下模型，再不断重复以上操作。