

使用高斯回归过程进行时间序列分析

S2302W0405 杨晨伟

1、原理简介

高斯过程是一类灵活的非线性回归和分类模型。它们最初是在地质统计学中引入的，近年来在机器学习界引起了广泛关注。它们与神经网络的不同之处在于，它们采用完整的贝叶斯处理，提供完整的后验预测分布。对于回归，它们在计算上也相对简单。

假设 x 是某个过程的 $f(x)$ ，于是表示为：

$$f(x) \sim GP(m(\cdot), k(\cdot, \cdot))$$

其中 $m(\cdot)$ 和 $k(\cdot, \cdot)$ 分别是均值和协方差函数：

$$m(x) = E[f(x)]$$

$$k(x_1, x_2) = E[(f(x_1) - m(x_1))(f(x_2) - m(x_2))]$$

高斯过程回归是通过贝叶斯推理从而得到实现的，目的是在给定的先验和训练数据的情况下获得函数的后验分布。然后，给定新的测试输入，我们可以使用后验得出基于测试输入和训练数据的预测分布。为了方便，通常假设先验分布的均值为 0。

2、预测方法

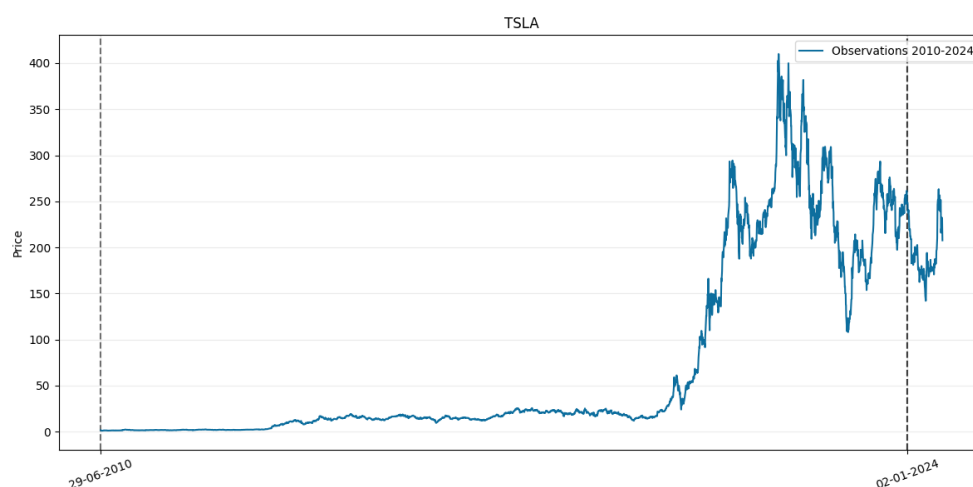
主要是避免将整个历史数据表示为一个时间序列。每个时间序列都被视为回归模型中的独立输入变量。在这个 case 中，每个序列代表了不同的年份，序列是交易期间特定价格的序列。基于股票市场的长度，一般就是 252 个结果的序列。而考虑到不完整的序列，比如今年，序列就会更少。因此，可以将过去序列的知识转移到要预测的新序列。每年的数据都被视为单独的时间序列，对应的年份就用作回归模型的独立变量。

3、数据与评估

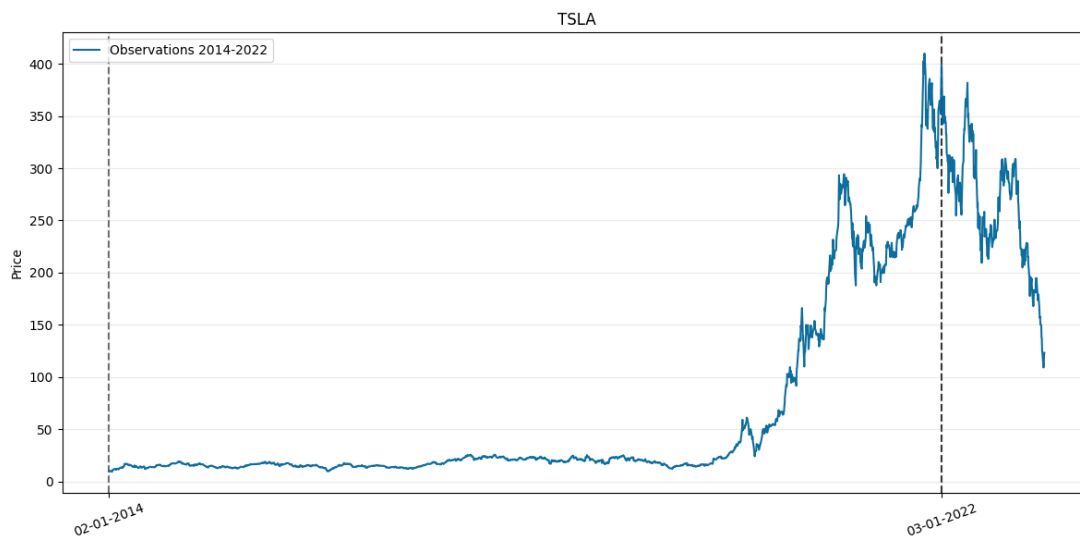
对于此项目选择了四个股票：特斯拉（TESLA），苹果（AAPL），微软（MSFT），英伟达（NVDA）。从财经网站上以 csv 格式下载了历史数据，通过 2014 到 2022 年的数据来预测 2023 年，通过 2014 到 2024 年的数据来预测 2024 剩下的走向。每年大约有 252 天交易时间，有些年份超过 252 天的则忽略 252 天以上的交易日，而交易日较少的年份则用年平均值来填充剩余的几天。使用了调整后的收盘价，旨在预测股票的趋势而不是价格。

1. 特斯拉公司

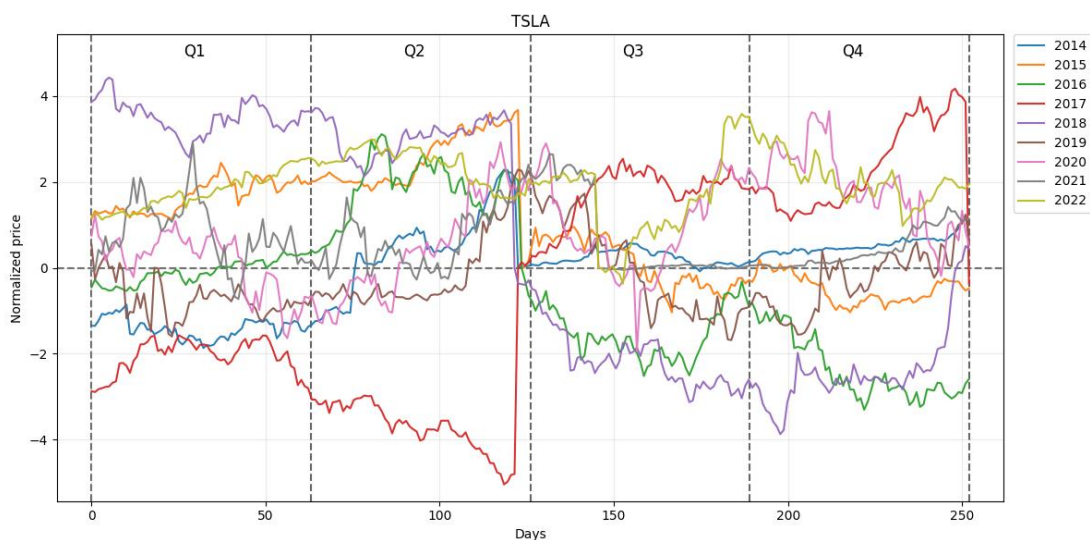
其历史的价格图表如下：



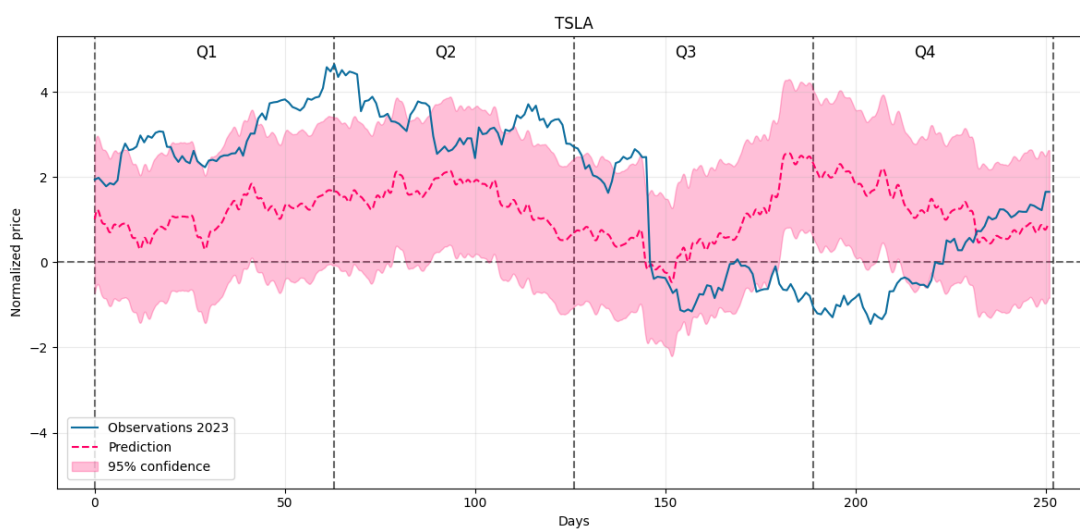
采用的样本区间为 2014 到 2022，图表如下：



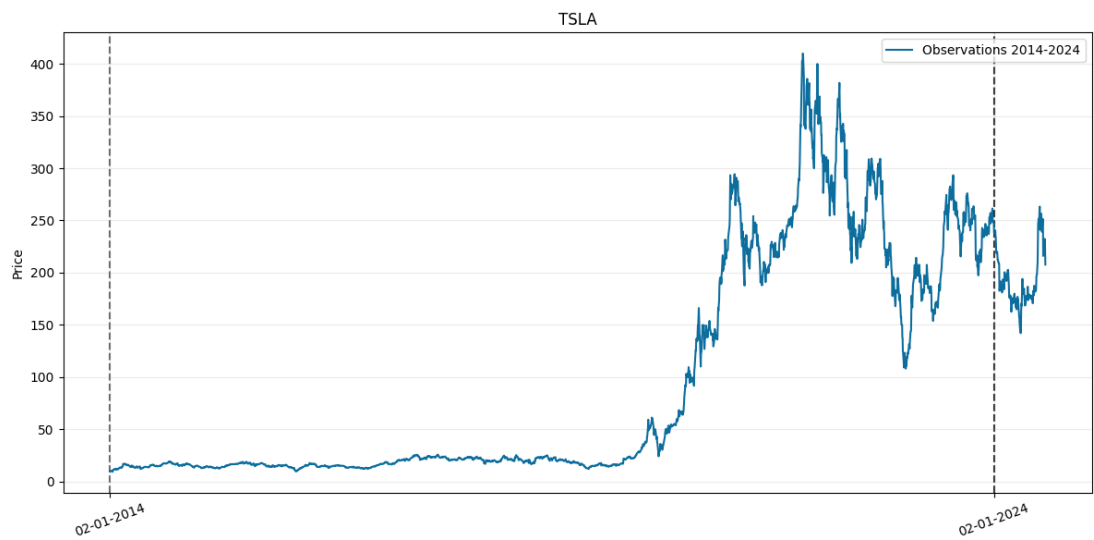
此期间的标准化价格分别是：



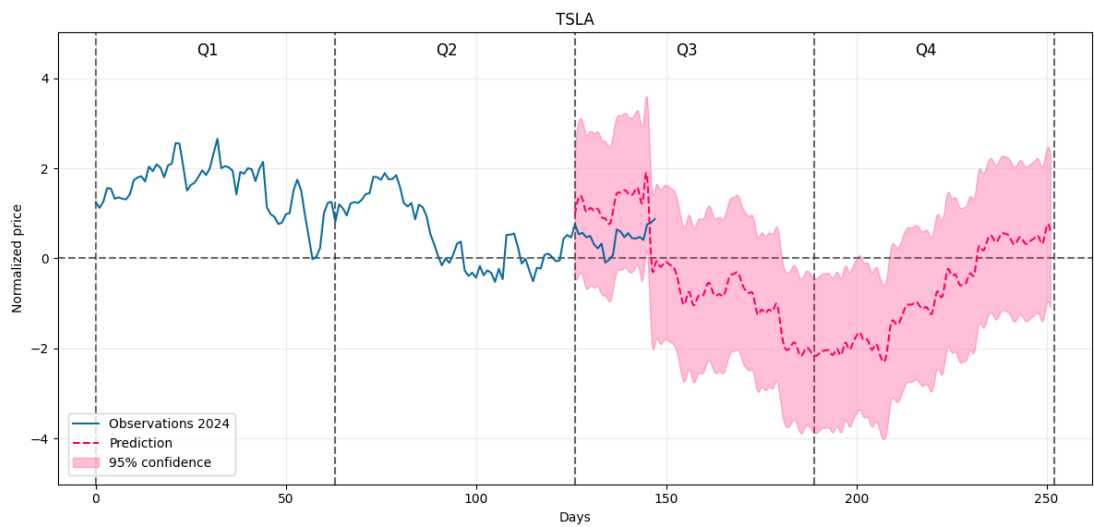
对于 2023 年全年的预测，置信区间为 95%，图表如下：



可以看出整体的趋势预测较好，Q4 的中期预测稍有失误。
接着看 2014 到 2024，图表如下：



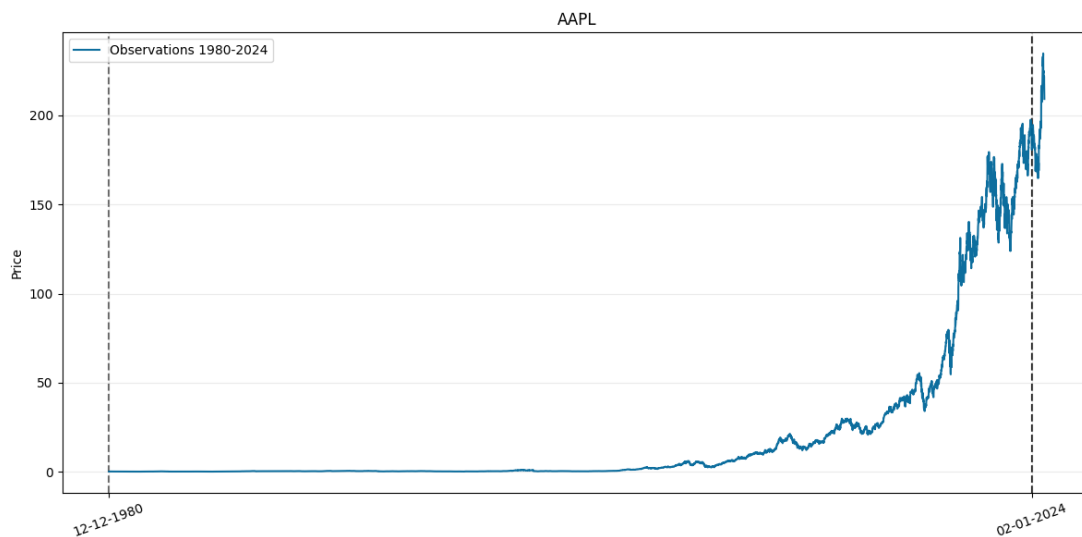
而对 2024 年剩下的预测如下：



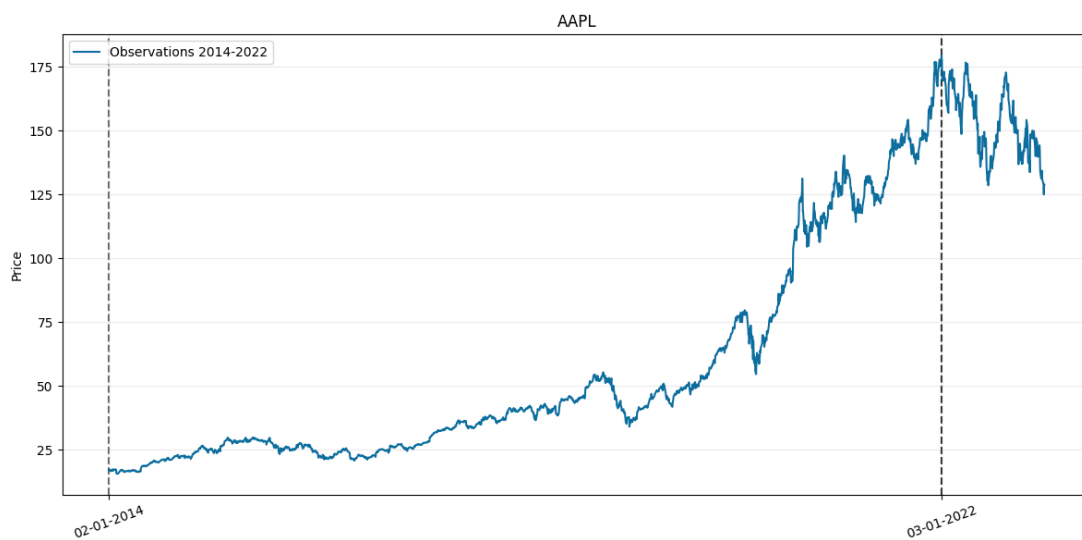
虽然偏离了价格，但是预测出来的初步趋势较好，今年剩下的时间可能会持续走低。

2. 苹果公司

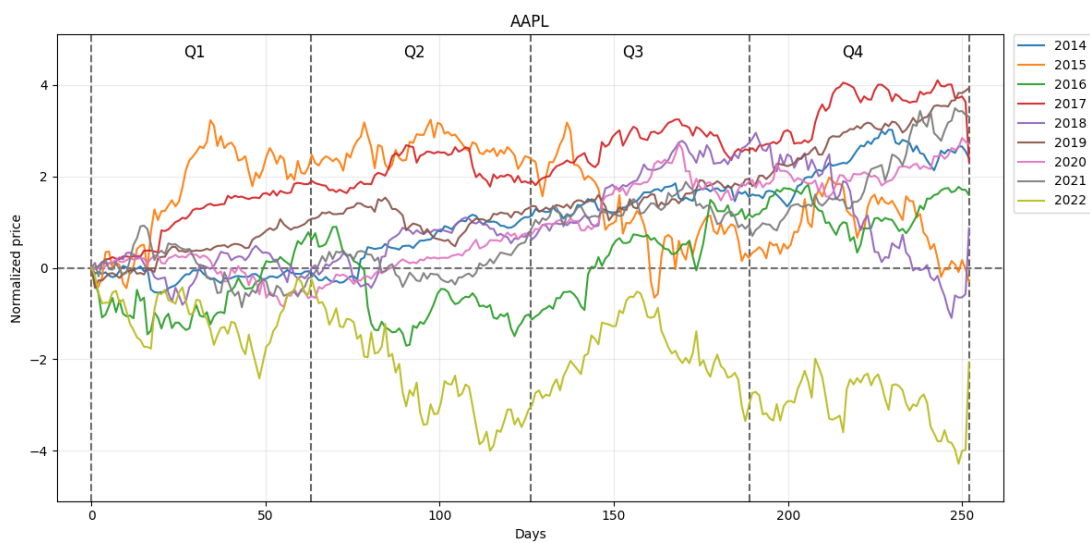
其历史价格图如下：



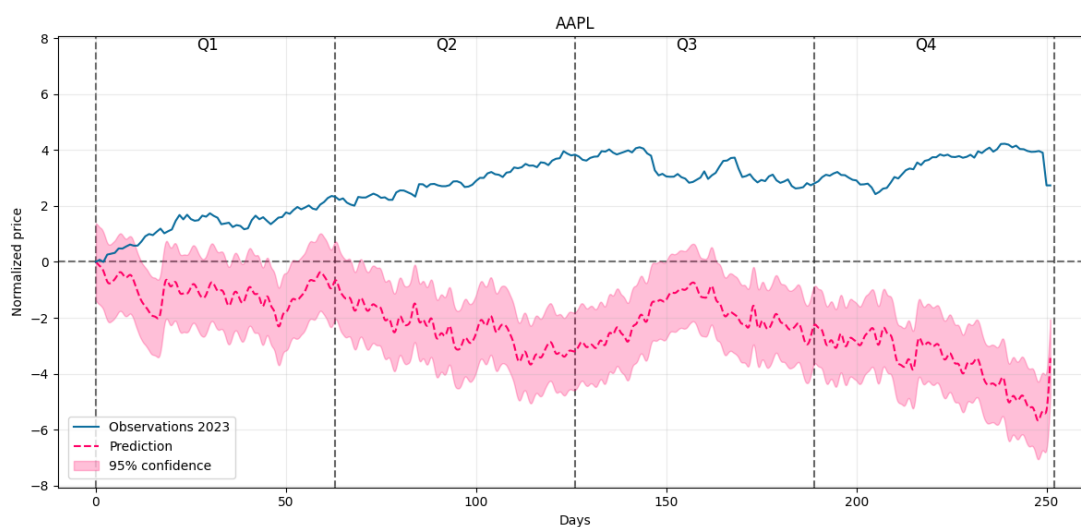
采用的样本区间为 2014 到 2022 年，图表如下：



期间标准化的价格为：

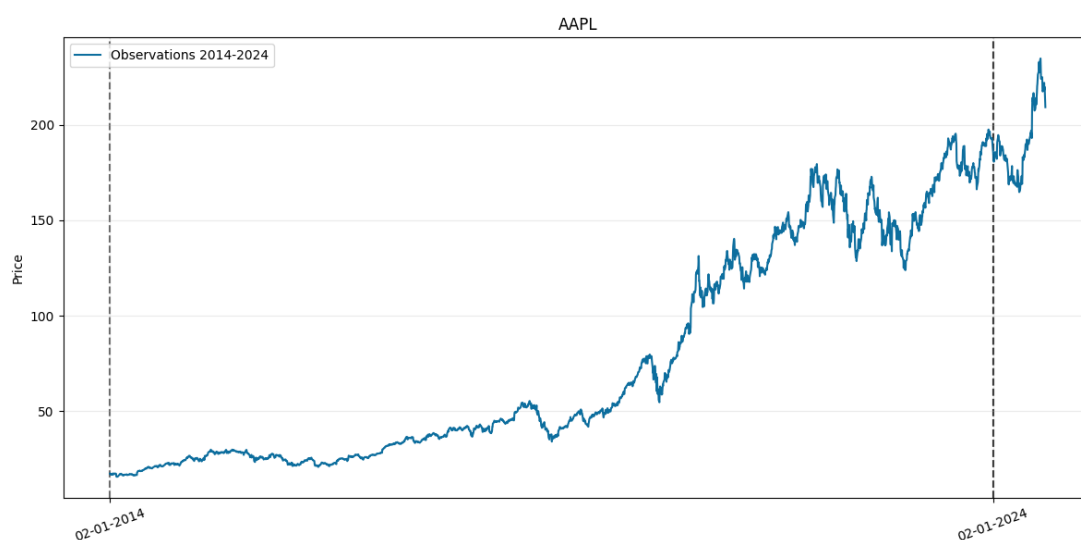


而我们对 2023 年全年的预测如下：



这次模型的预测结果不尽人意，Q1 中期和 Q3 以及 Q4 前期预测趋势较好，其他的时期大大低估苹果的走向，也许是 2022 年一年整体持续走低，而两个时间点很近，相关性较高，所以使得预测结果也认为苹果股价走向不乐观。

接着是 2014 到 2024 年图表如下：



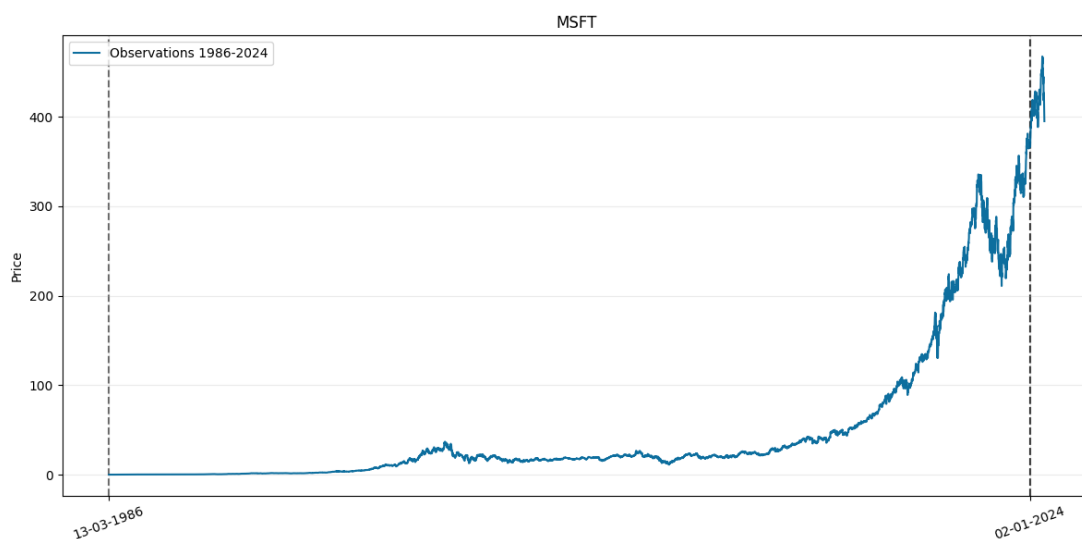
模型对于今年剩下的时间预测如下图所示：



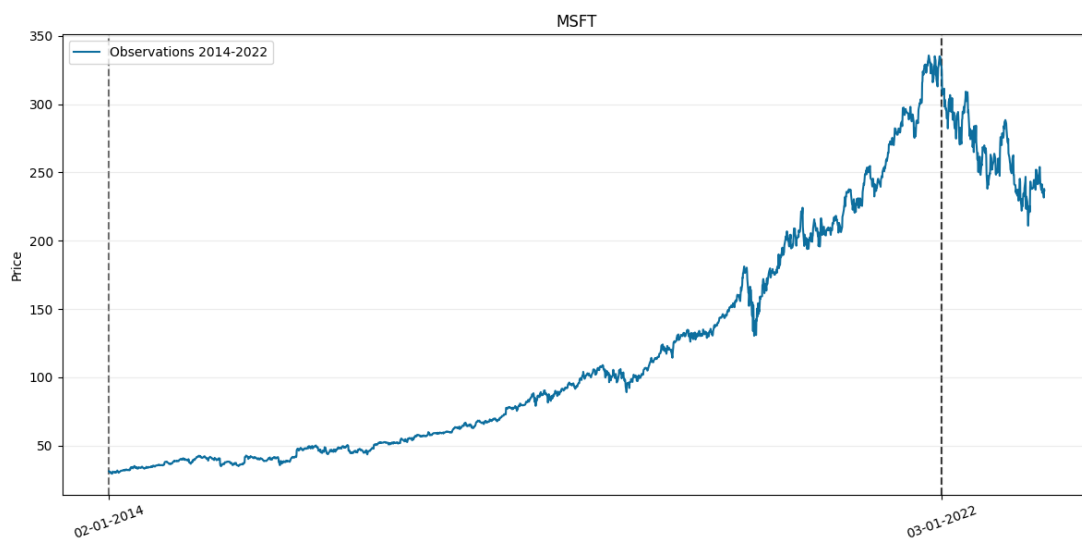
同样的，虽然价格上有出入，但是趋势相近，以及未来的走势是否像模型预测的那样涨幅还算平稳还需拭目以待。

3. 微软公司

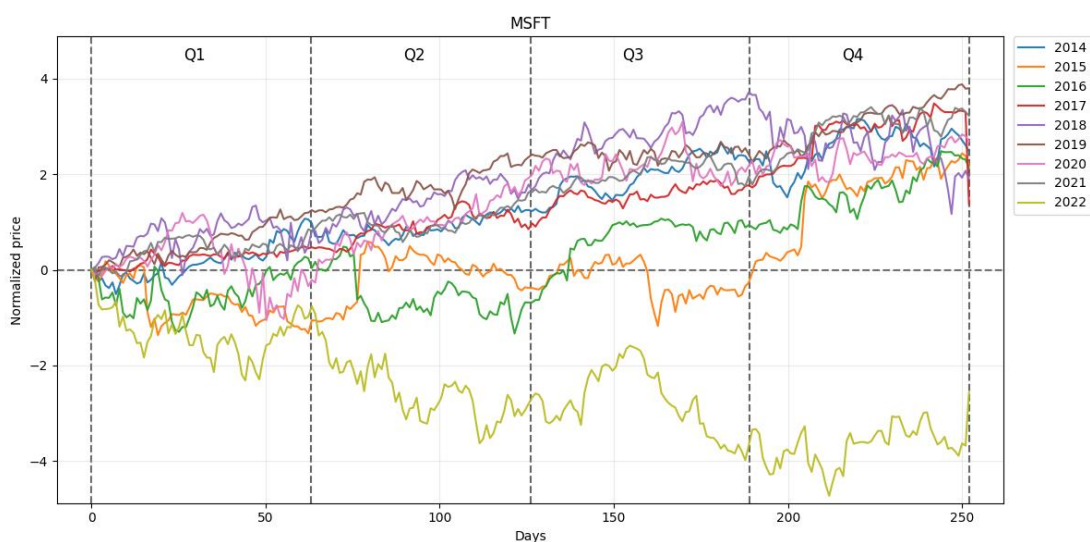
其历史价格如图所示：



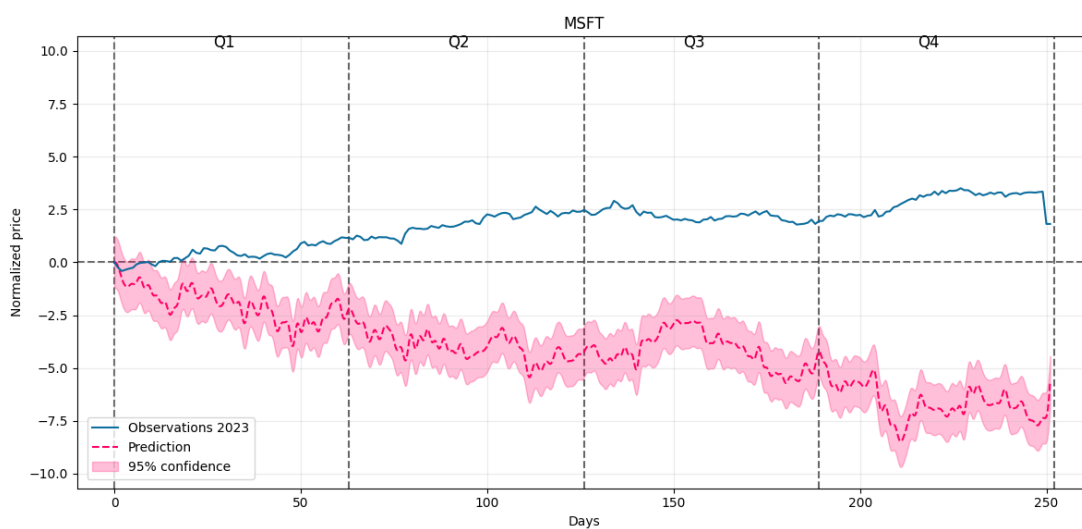
选取的样本区间为 2014 到 2022 年，图表如下：



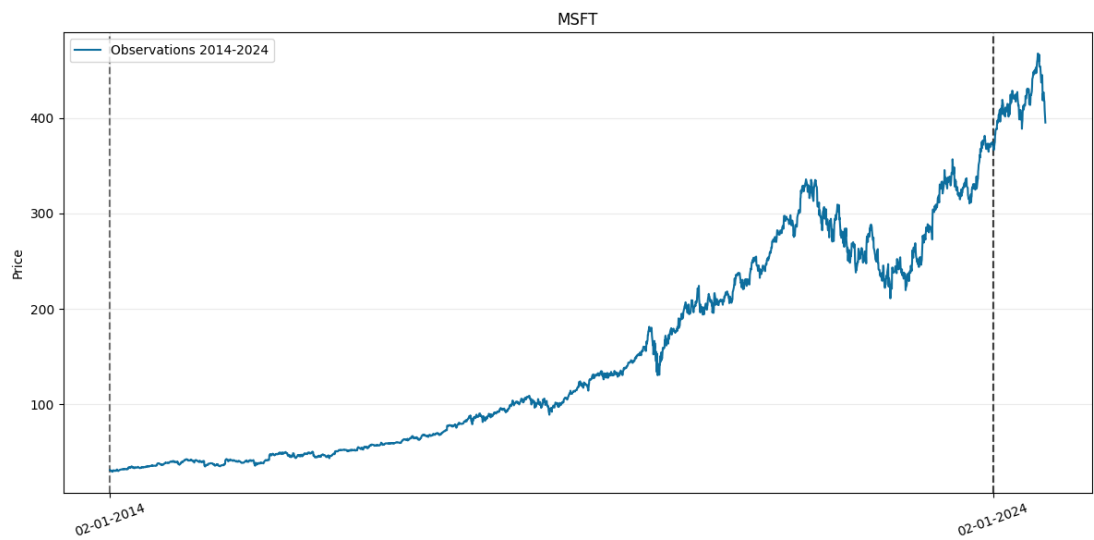
期间标准化价格如图：



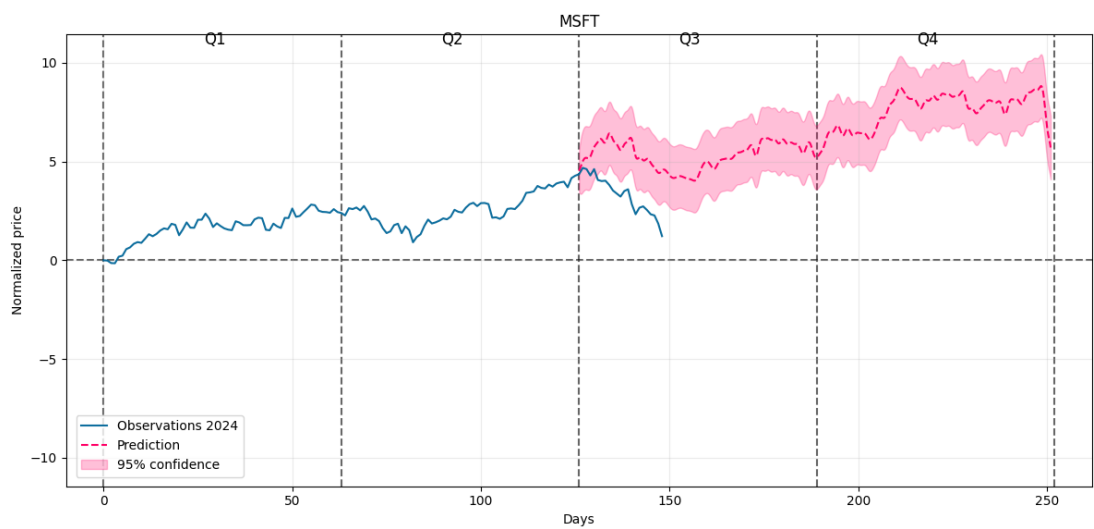
于是对于 2023 年全年的预测如下图所示：



也是相似的原因导致预测结果很不理想，预测结果偏差较大。
再基于以下区间范围：



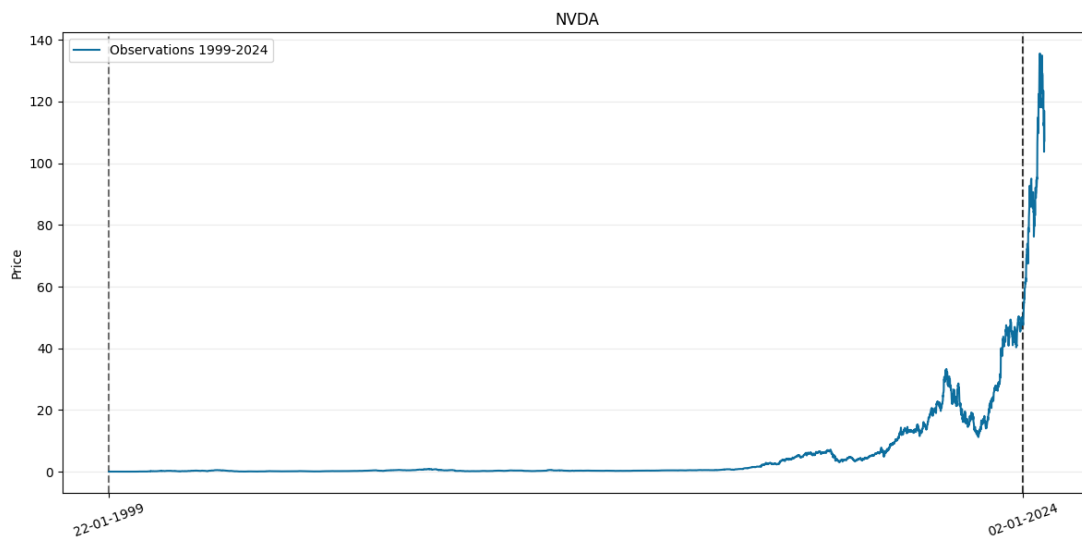
进行了相关预测，预测今年下半年的趋势，如图所示：



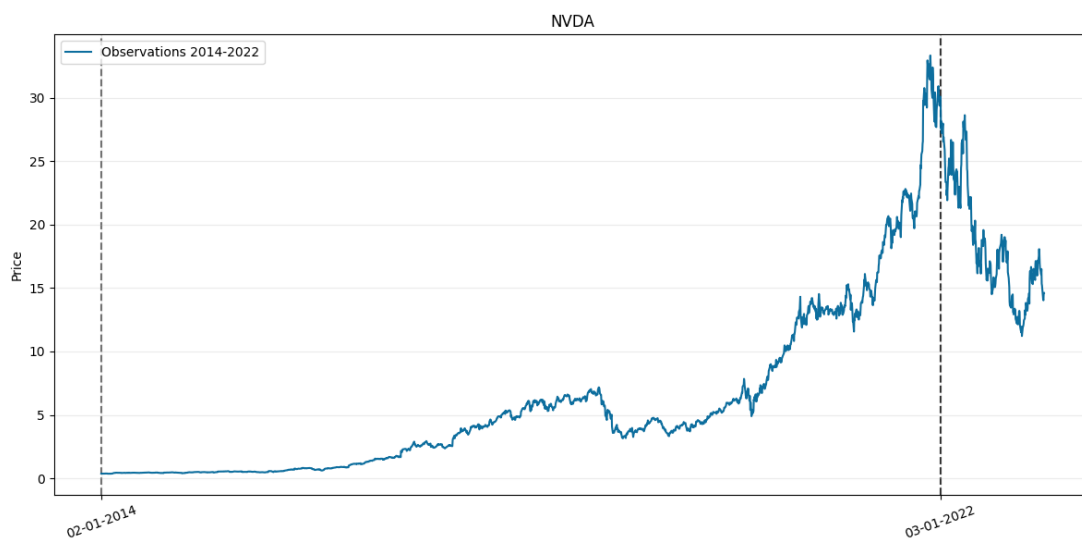
可以看出预测的价格和实际价格十分相近，并且过后一段时间的趋势也大致相似。

4. 英伟达

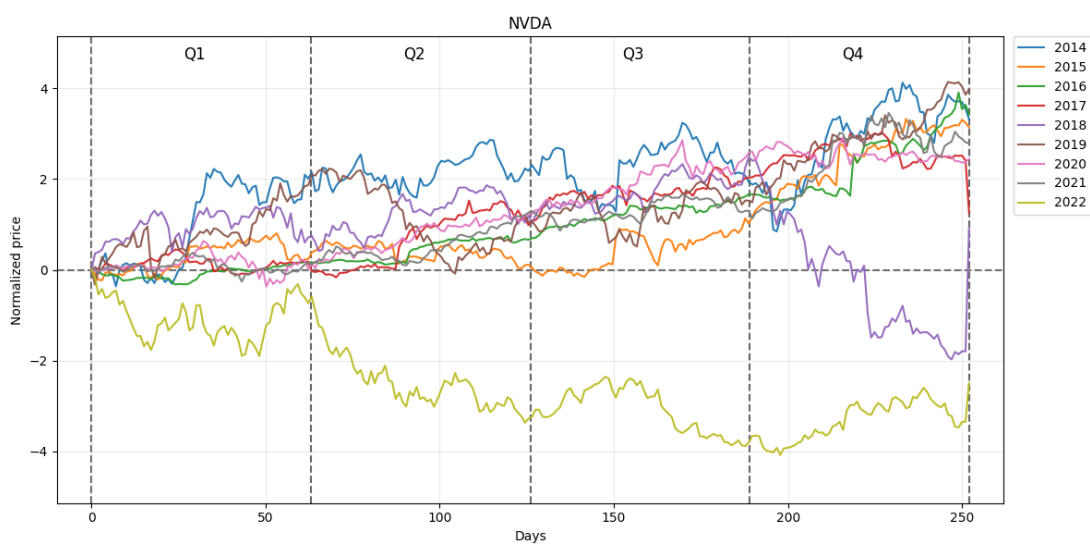
其历史价格如图所示：



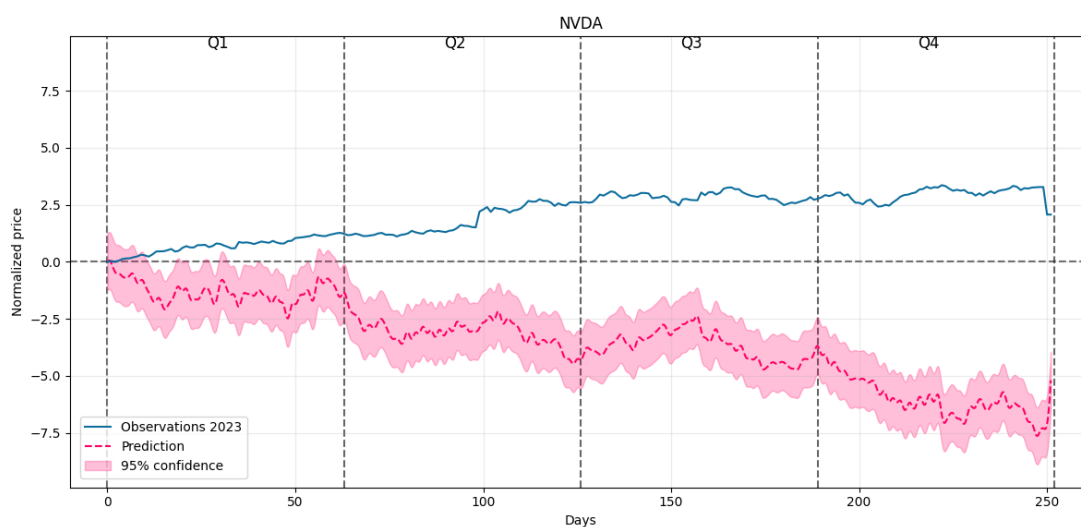
选取的样本区间为 2014 到 2022 年，图表如下：



其标准化价格如下图：

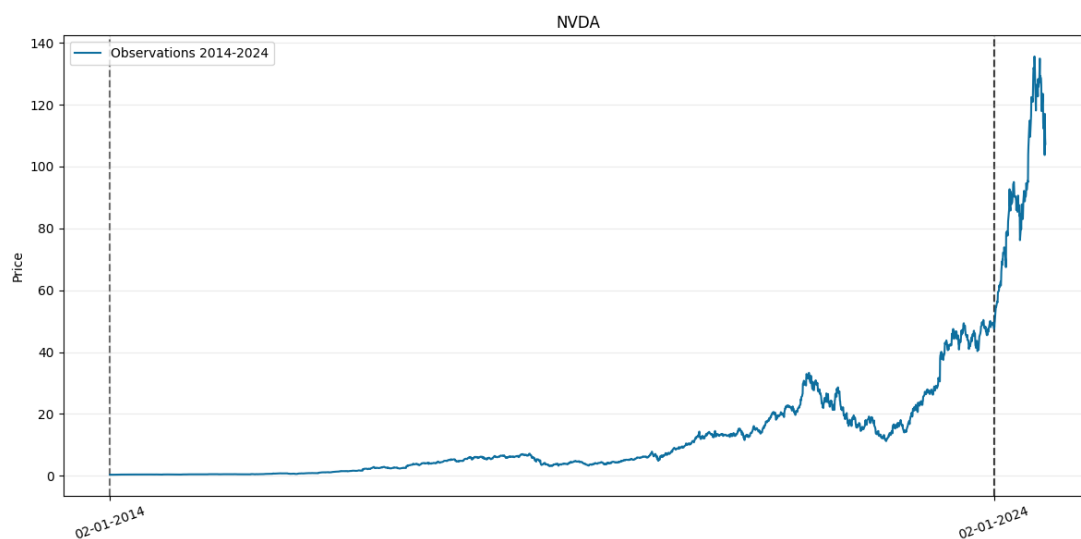


对于 2023 年的预测如下图所示：

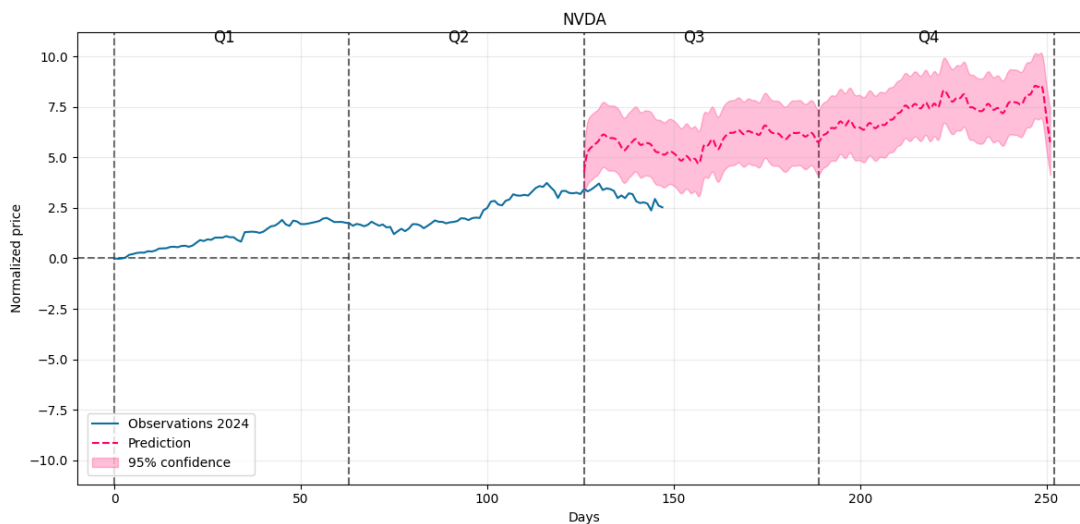


同样由于 2022 年的全年走低，预测结果也并不向好，但是 Q2 和 Q3 平稳的趋势相近，预测的水平尚可。

接下来看 2014 到 2024 年的数据：



对于 2024 年剩下时间的预测如下图：



同样是价格的预测稍高，但是开头的走势大致相同，后续的结果预测是否准确仍然需要持续关注。

4、概括

根据之前的预测结果可以看出，预测趋势结果尚可，如果前一年数据突然和之前数据相差较大，那么预测容易被时间序列相近的数据影响从而出现误差较大的情况。如果数据比较平均或者跟往年差别不大那么预测效果其实比较好。有时预测结果很好，也有预测情况很差，所以这个并不能成为投资时决定因素，不过可以当一些参考。也许和别的手段结合可以让它的结果更加可靠。

高斯回归过程在时间序列中的使用，可以用与金融数据分析，检测环境各项数据(例如气候，水质变化等)。