从回归角度看时间序列：

时间序列预测，前提假设是过去的表现可以预测未来。因此，本质上就是回归问题，只不过自变量是过去的表现，问题是选取过去的哪几期作为自变量，也就是**如何选取具有预测能力的特征**。

AR模型：将周期性特征差分了，再用过去的p期进行预测。此时，预测包含了两部分：

第一部分是过去最近几期的特征预测（预测差分后剩余的信息）：



第二部分是确定的周期性特征预测（根据差分计算原始的信息）：





回归模型：

方式一：

类似ARIMA模型，先去除确定信息（差分是简单的方法），再用模型提取剩余的信息，问题的关键在于**滞后p期的选择**，如何选择呢？

自相关和交叉项相关分析---构建线性模型，该模型能够拟合非线性关系

自相关和交叉项相关分析----构建非线性模型（如SVM模型，只需自相关分析）

方式二：

两者结合起来预测—问题的关键在于**滞后p期的选择**，如何选择呢？

#### Summary

1. 当预测值相比实际值滞后一期，说明选择的滞后变量（时间序列特征）过少或不具有预测能力，导致预测下一期时只用到了当期的信息。因此，需要尝试加入具有预测能力的特征。
2. 加入外生变量，可以减少实际值与预测值之间的差距。

注：对于未来确定的外生变量可以使用当期值（如节假日、温度、湿度等），未来不确定的外生变量则使用滞后一期或滞后多期值。如果外生变量与目标变量相关性较高，可以直接用外生变量进行预测，效果可能很好。

温度、湿度，通常公布未来七天数据

1. 一般步骤：

 the chi-square periodogram test确定周期；

STL分解；

季节性用STLM拟合，趋势性用多项式/分段样条/SVM拟合，残差用SVM/???拟合。直至模型拟合后的残差为白噪声序列。

注：多个周期时，可以先大周期差分，其余步骤同上。

重点在于，如何解释残差？是否含有相关信息？

Reference:

1Xu Cui » SVM regression on time series, is there a lag