BetaPlus 小组

专注于多因子模型和异象研究

时间序列分析小册子

- 作者：石川
- 版本：2021/01/03

未经授权，严禁转载。请务必阅读文末免责声明。

目录

1 引言

时间序列分析（time series analysis）是金融数学中的一门基本技术。**时间序列是指在一定时间内按时间顺序测量的某个变量的取值序列。**比如变量是股票价格，那么它随时间的变化就是一个时间序列；如果变量是股票的收益率，则它随时间的变化也是一个时间序列。**时间序列分析就是使用统计的手段对这个序列的过去进行分析，以此对该变量的变化特性建模，预测未来。**

时间序列分析要求使用者具备一定的高等数学知识。特别是其中一些高级的模型，如分析波动率的 ARCH/GARCH 模型、极值理论、连续随机过程、状态空间模型等都对使用者的数学水平有着极高的要求。因此，在很多人眼中，金融时间序列分析无疑带着厚厚的面纱，令人望而却步。

然而，如果学习的目的是为了解金融时间序列的特点、熟悉金融时间序列分析的目的、并使用线性但非常实用的模型（比如 ARMA 模型）对金融时间序列进行预测并以此制定量化策略，那么只要具备简单的统计学基础，就完全能够实现这些目标。本小册子将深入浅出的介绍金融时间序列分析的相关知识，设计的内容包括自相关性、白噪声和随机游走、AR、MA、ARMA、GARCH 等时间序列分析中常见的概念。

以下行文会避免过多罗列晦涩难懂的大数学（但会涉及必要的数学知识），希望带你走入金融时间序列分析的大门，为你今后学习更高级的模型奠定一些基础。在介绍具体内容之前，首先简要介绍下金融时间序列分析。

金融时间序列分析考虑的是金融变量（比如投资品收益率）随时间演变的理论和实践。任何金融时间序列都包含不确定因素，因此统计学的理论和方法在金融时间序列分析中至关重要。金融资产的时间序列常被看作是未知随机变量序列随时间变化的一个实现。通常假设该随机变量序列仅在时间轴上的离散点有定义，则该随机变量序列就是一个离散随机过程。比如股票的日收益率就是离散的时间序列。

在量化投资领域，研究的目标是通过统计手段对投资品的收益率这个时间序列建模，以此推断序列中不同交易日的收益率之间有无任何特征，以此来预测未来的收益率并产生交易信号。

一个时间序列可能存在的特征包括以下几种：

- **趋势：**趋势是时间序列在某一方向上持续运动（比如牛市时股市每天都在上涨，股票收益率持续为正；熊市时股市每天都在下跌，股票收益率持续为负）。趋势经常出现在金融时间序列中，特别是大宗商品价格；许多商品交易顾问（CTA）基金在他们的交易算法中都使用了复杂的趋势识别模型。

- **季节变化：**许多时间序列中包含季节变化。在金融领域，我们经常看到商品价格的季节性变化，特别是那些与生长季节或温度变化有关的商品，比如天然气。
- **序列相关性：金融时间序列的一个最重要特征是序列相关性（serial correlation），又称为自相关性（autocorrelation）。**以投资品的收益率序列为例，人们会经常观察到一段时间内的收益率之间存在正相关或者负相关。此外，**波动聚类（volatility clustering）**也是一种序列相关性，它意味着高波动的阶段往往伴随着高波动的阶段出现、低波动的阶段往往伴随着低波动的阶段出现。
- **随机噪声：**它是时间序列中除去趋势、季节变化和自相关性之后的剩余随机扰动。由于时间序列存在不确定性，随机噪声总是夹杂在时间序列中，致使时间序列表现出某种震荡式的无规律运动。

量化投资的交易者的目标是利用统计建模来识别金融时间序列中潜在的趋势、季节变化和序列相关性。利用一个好的模型，金融时间序列分析的主要应用包括：

- **预测未来：**为了成功交易，需要在**统计上**准确预测未来的投资品价格或者收益率。
- **序列模拟：**一旦发现了金融时间序列的统计特征，便可通过它们来模拟时间序列并进行场景分析。这对于估计交易次数、期望交易成本、期望收益率至关重要，从而最终定量的计算一个策略或者投资组合的风险分布和盈利水平。

金融时间序列的关系中，最重要的当属自相关性。这是因为通常很容易从一个时间序列中识别出趋势以及季节变换。当除去这些关系后，剩下的时间序列往往看来十分随机。然而对于金融时间序列，比如投资品的收益率，看似随机的时间序列中往往存在着自相关。**对自相关建模并加以利用能够大幅提高交易信号的准确性。**配对交易的均值回复策略就是这么一个例子。均值回复策略利用一对投资品价差序列的**负相关性**进行投资，产生做多或者做空的交易信号，实现盈利。

金融时间序列分析的核心就是挖掘该时间序列中的自相关性。

下一节就来介绍如何计算时间序列的自相关性。为此，先来看两个基础概念：协方差和相关系数。之后会谈及时间序列的平稳性，它是时间序列分析的一个必要前提。最后介绍时间序列的自相关性。

2 自相关性

2.1 协方差和相关系数

本节介绍概率论中的基础概念：协方差和相关系数。熟悉它们的读者可跳过。

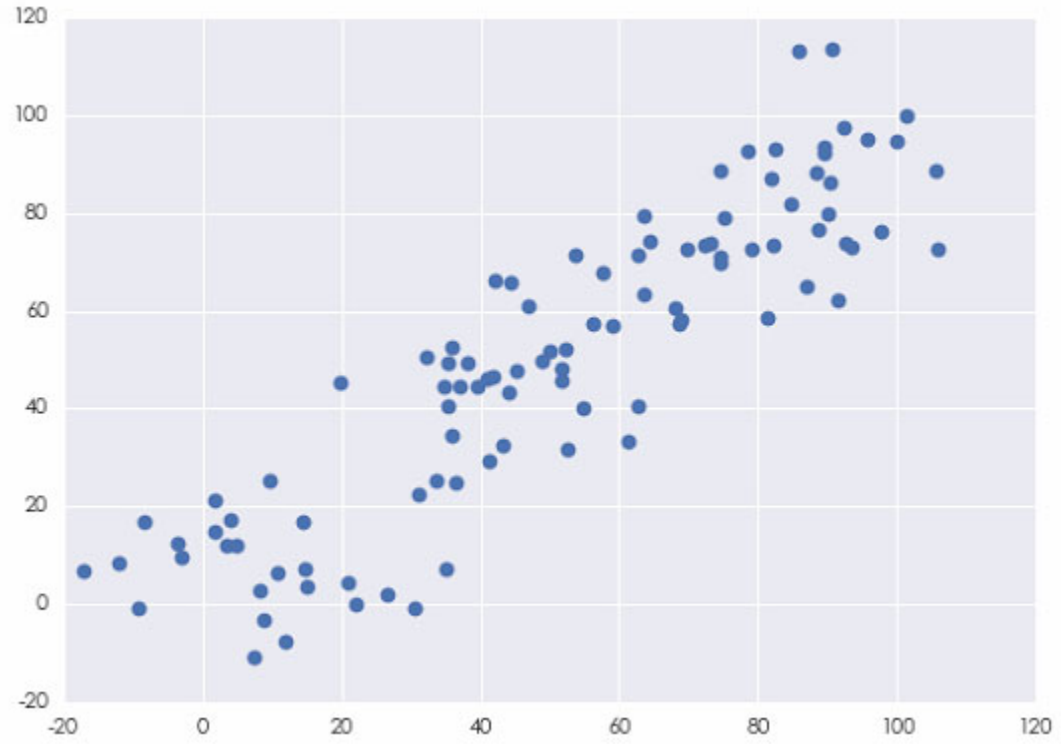
假设两个随机变量 X 和 Y 满足未知的概率分布（可以是同分布也可以是不同的分布）。$E[]$ 为求解数学期望的运算符。X 和 Y 的**总体协方差（population covariance）**为：

$$\sigma(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

其中，μ_X 和 μ_Y 分别为 X 和 Y 的**总体均值（population mean）**。**协方差告诉人们两个随机变量是如何一起移动的。**在实际中，由于总体的概率分布未知，只能通过 X 和 Y 的观测值来计算**样本均值（sample mean）**。假设各有 X 和 Y 的观测值 n 个，则它们的**样本协方差（sample covariance）**为：

$$\text{cov}(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

其中，$\bar{X}$ 和 $\bar{Y}$ 为 X 和 Y 的样本均值。上面公式中右侧之所以除以 $n-1$ 而非 n 的原因是，这么做可以保证样本协方差是（未知）总体协方差的一个**无偏估计量（unbiased estimator）**。假设随机生成两个随机变量 X 和 Y 的序列，它们的散点图如下。



按照上面的公式，X 和 Y 的样本协方差为 893.215203。它有什么意义呢？在回答这个问题之前，先来看另外两个变量，称之为 $X100$ 和 $Y100$。它们分别定义为 $X100 = 100 \times X$ 和 $Y100 = 100 \times Y$。可见，它们仅仅是 X 和 Y 各乘以 100 得到的。$X100$ 和 $Y100$ 的样本协方差为 8932152.03，这是 X 和 Y 的协方差的 10000 倍。然而，如果仅仅因此就得出 $X100$ 和 $Y100$ 的相关性高于 X 和 Y 的相关性就大错特错了。事实上，由于 $X100$ 和 $Y100$ 是由 X 和 Y 分别乘以 100 得到的，因此它们之间的相关性显然和 X 与 Y 的相关性相同。

上面这个例子说明使用协方差衡量变量相关性的缺点：**协方差是有量纲的，因此它的大小受随机变量本身波动范围的影响。**在上个例子中，当两个随机变量的波动范围扩大 100 倍后，它们的协方差扩大了 10000 倍。因此，人们希望使用某个和协方差有关，但是又是**无量纲**的测量来描述两个随机变量的相关性。最简单的做法就是用变量自身的波动对协方差进行标准化。**相关系数（correlation 或者 correlation coefficient）**便由此得来。

令 ρ 表示 X 和 Y 的**总体相关系数（population correlation）**，它的定义为：

$$\rho(X,Y)=\frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}=\frac{\sigma(X,Y)}{\sigma_X\sigma_Y}$$

其中 σ_X 和 σ_Y 分别为 X 和 Y 的**总体标准差（population standard deviation）**。通过使用 X 和 Y 的标准差对它们的协方差归一化，ρ 的取值范围在 -1 到 $+1$ 之间：

- $\rho(X,Y)=1$ 表示 X 和 Y 之间存在确切的**线性正相关**；
- $\rho(X,Y)=0$ 表示 X 和 Y 之间不存在任何**线性相关性**；
- $\rho(X,Y)=-1$ 表示 X 和 Y 之间存在确切的**线性负相关**。

值得一提的是，**相关系数仅仅刻画 X 和 Y 之间的线性相关性；它不描述它们之间的（任何）非线性关系。**在实际中，由于总体的概率分布未知，只能通过 X 和 Y 的观测值来计算 X 和 Y 的**样本相关系数（sample correlation）**：

$$\mathrm{corr}(X,Y)=\frac{\mathrm{cov}(X,Y)}{\mathrm{sd}(X)\mathrm{sd}(Y)}$$

其中，$\mathrm{sd}(X)$ 和 $\mathrm{sd}(Y)$ 分别为 X 和 Y 的**样本标准差（sample standard deviation）**。在上面的例子中，无论考虑 X 和 Y 还是 $X100$ 和 $Y100$（即无论如何缩放 X 和 Y），它们的相关系数都是 0.894655，这和预期相符。由于这个数值非常接近 1，它意味着二者之间存在很强的线性正相关。

2.2 时间序列的平稳性

平稳性（stationarity）是时间序列分析的基础。

为了通俗的理解平稳性，来看下面这个类比（这是我能想到的最好的例子）。假如某股票的日收益率由转轮盘赌决定：转到不同数字就对应不同的收益率。在每个时刻 t 都转同一个轮盘赌并确定收益率 r_t。只要这个轮盘不变，那么对于所有的 t，r_t 的概率分布都是一样的、不随时间变化。这样的时间序列 $\{r_t\}$ 就是（严格）平稳的。如果从某个时刻 t' 开始，轮盘发生了变化（比如轮盘上面的数字变多了），那么从 $t\geq t'$ 开始，r_t 的分布就便随之发生变化，因此时间序列 $\{r_t\}$ 就不是平稳的。

在数学上，时间序列的**严平稳（strictly stationary）**有着更精确的定义：它要求时间序列中任意给定长度的两段子序列都满足相同的联合分布。**这是一个很强的条件，在实际中几乎不可能被满足。**因此还有**弱平稳（weakly stationary）**的定义，**它要求时间序列满足均值平稳性（stationary in mean）和二阶平稳性（secondary order stationary）**。

如果一个时间序列 $\{r_t\}$ 满足以下两个条件，则它是弱平稳的：

1. 对于所有的时刻 t，有 $E[r_t] = \mu$，其中 μ 是一个常数。
2. 对于所有的时刻 t 和任意的间隔 k，r_t 和 r_{t-k} 的协方差 $\sigma(r_t, r_{t-k}) = \gamma_k$，其中 γ_k 与 t 无关，它仅仅依赖于间隔 k。特别的，当 $k = 0$ 时，这个特性意味着 $\sigma(r_t, r_t)$ —— r_t 的方差 —— 不随时间变化，等于一个与时间 t 无关的常数 γ_0，这称为**方差平稳性（stationary in variance）**。

弱平稳假设对于分析投资品收益率至关重要。

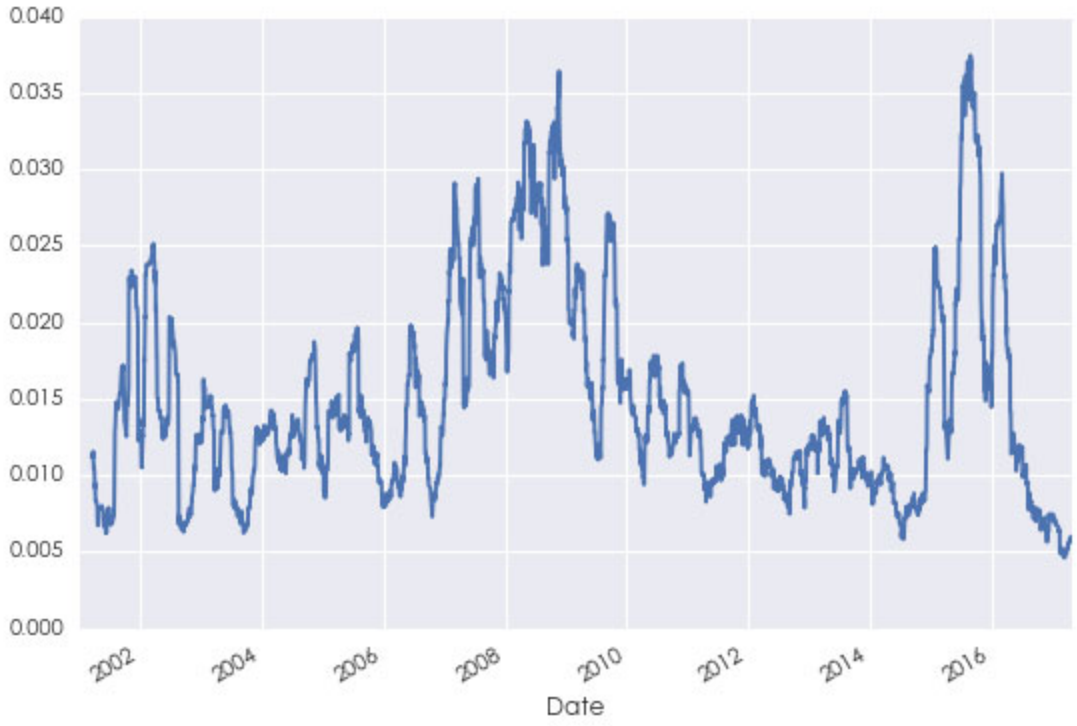
为了解释这一点，来看一个例子。假设我们想知道 2017 年 5 月 16 日这天上证指数收益率的均值是多少，而它是来自某个未知的分布。也许你会马上说“查一下 Wind 不就知道了？上证指数那天的收益率是 0.74%”。注意，0.74% 这个数值仅仅是那天上证指数未知收益率分布的一个实现（realization），它不是均值，因此从时间序列分析的角度来说仅仅知道 0.74% 远远不够。

对于一般的未知概率分布，只要通过进行大量重复性实验，就可以有足够多的独立观测点来进行统计推断（计算均值和方差这些统计量）。按照这个思路，必须把 2017 年 5 月 16 日这一天经历许多遍，得到许多个那天的收益率观测值，然后用这些观测值计算出收益率的均值。不幸的是，历史只发生一次，时间也一去不复返，人们只能实实在在的经历一次 2017 年 5 月 16 日，只能得到一个收益率的观测点，即 0.74%。因此这个方法对于金融数据是行不通的。

然而，如果假设上证指数的收益率序列满足弱平稳，就柳暗花明了。根据弱平稳假设，上证指数的日收益率序列 $\{r_t\}$ 的均值是一个与时间无关的常数，即 $E[r_t] = \mu$。这样便可以利用一段时间的历史数据来计算出日收益率的均值。比如我们可以对上证指数在 2017 年交易日的日收益率序列取平均，把它作为对总体均值 μ 的一个估计。根据弱平稳性，该平均值也正是 2017 年 5 月 16 日的收益率均值。

同样的道理，在弱平稳的假设下，可以根据历史数据方便的对时间序列的诸多统计量进行推断。在金融文献中，也通常假定投资品收益率序列是弱平稳的。只要有足够多的历史数据，这个假定可以用实证方法验证。比如，可以把数据分成若干个子集，并分别计算每个子集的统计量，然后通过统计的手段检验这些来自不同子集的统计量的一致性。

需要说明的是，即便是弱平稳性，有时金融数据也无法满足。下图给出了 2001 倒 2017 年之间上证指数日收益率**标准差**的时间序列。它清晰的说明标准差是随时间变化的，因而收益率序列不满足二阶平稳性。对于此，可以通过更复杂的非线性模型对波动率建模（比如 GARCH），又或者可以把时间段细分为更短的区间，使得在每个小区间内的收益率序列尽量满足弱平稳性。



有了上述铺垫，下面就来解释时间序列的自相关性。

2.3 自相关性和自相关函数

假设我们有弱平稳的投资品收益率序列 $\{r_t\}$。自相关性考察的是 t 时刻的收益率 r_t 和距当前任意间隔 k 时刻的收益率 r_{t-k} 之间的线性相依关系（k 的取值是所有 ≥ 0 的整数）。由于 r_t 和 r_{t-k} 来自同一个时间序列，因此将 2.1 节中的相关系数的概念应用到 r_t 和 r_{t-k} 上，便推广出**自相关系数（autocorrelation）**。

定义：r_t 和 r_{t-k} 的相关系数称为 r_t 的间隔为 k 的自相关系数。

在弱平稳假设下，这个间隔为 k 的自相关系数与时间 t 无关，而仅仅与间隔 k 有关，由 ρ_k 表示。由 2.1 节介绍的相关系数的定义可知：

$$\rho_k = \frac{\sigma(r_t, r_{t-k})}{\sigma_{r_t}\sigma_{r_{t-k}}} = \frac{\sigma(r_t, r_{t-k})}{\sigma_{r_t}\sigma_{r_{t-k}}} = \frac{\gamma_k}{\gamma_0}$$

上面的推导中用到了弱平稳的性质，即协方差和方差平稳性（换句话说，二阶平稳性）。从这个定义不难看出，当 $k = 0$ 时有：

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$

这表示 r_t 的间隔为 0 的自相关系数恒定为 1。此外，ρ_k 还有如下的性质：

$$\rho_k = \rho_{-k}; \quad -1 \leq \rho_k \leq 1$$

和 2.1 节一样，上面定义的 ρ_k 是总体的统计特性。实际中，仍然只能通过有限的样本数据来计算样本的统计特性。令 ζ_k 为与 ρ_k 对应的样本统计量，则有：

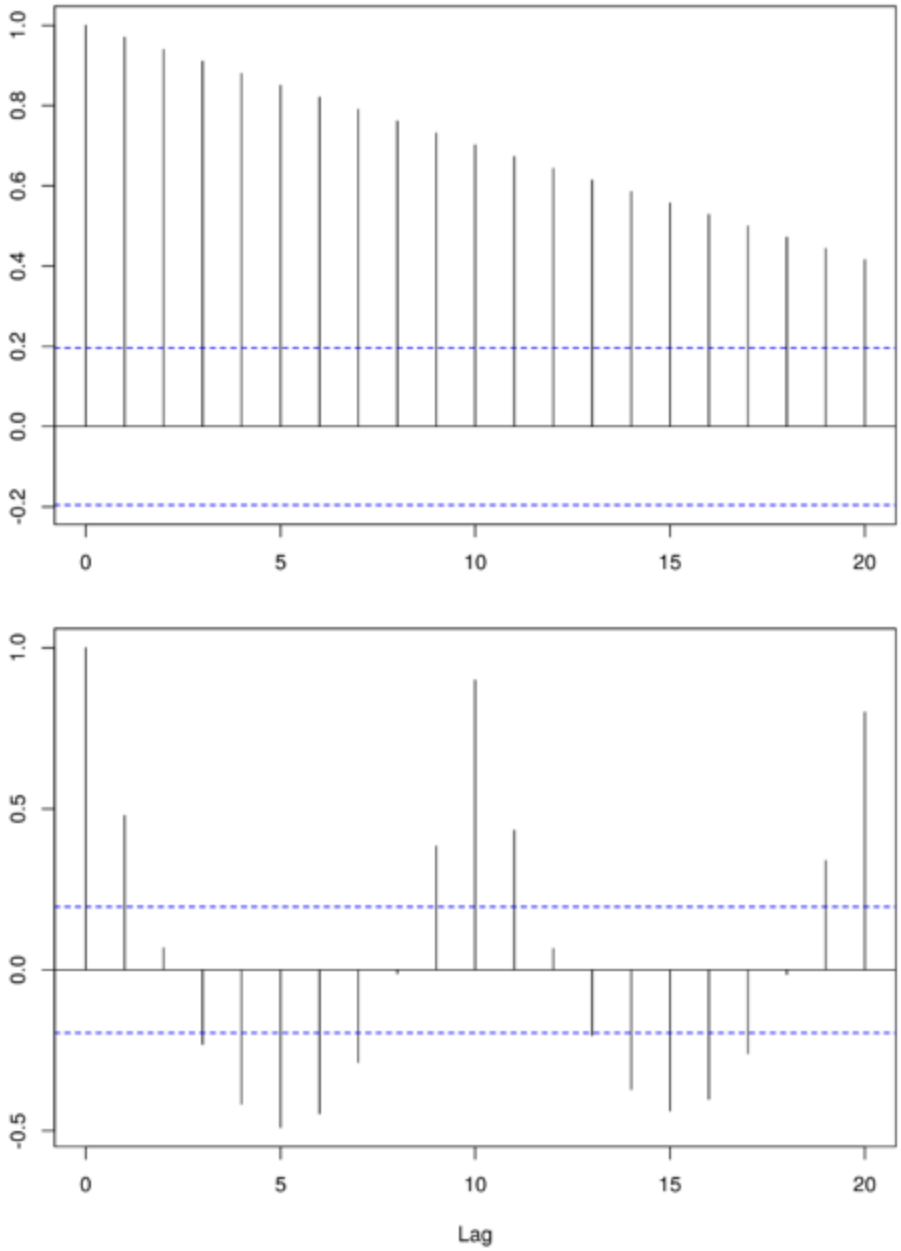
$$\zeta_k = \frac{c_k}{c_0}$$

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(r_t - \bar{r})(r_{t+k} - \bar{r})$$

上式中，c_k 是 r_t 的**间隔为 k 的样本自协方差（sample autocovariance of lag k）**；ζ_k 为 r_t 的**间隔为 k 的样本自相关系数（sample autocorrelation of lag k）**。

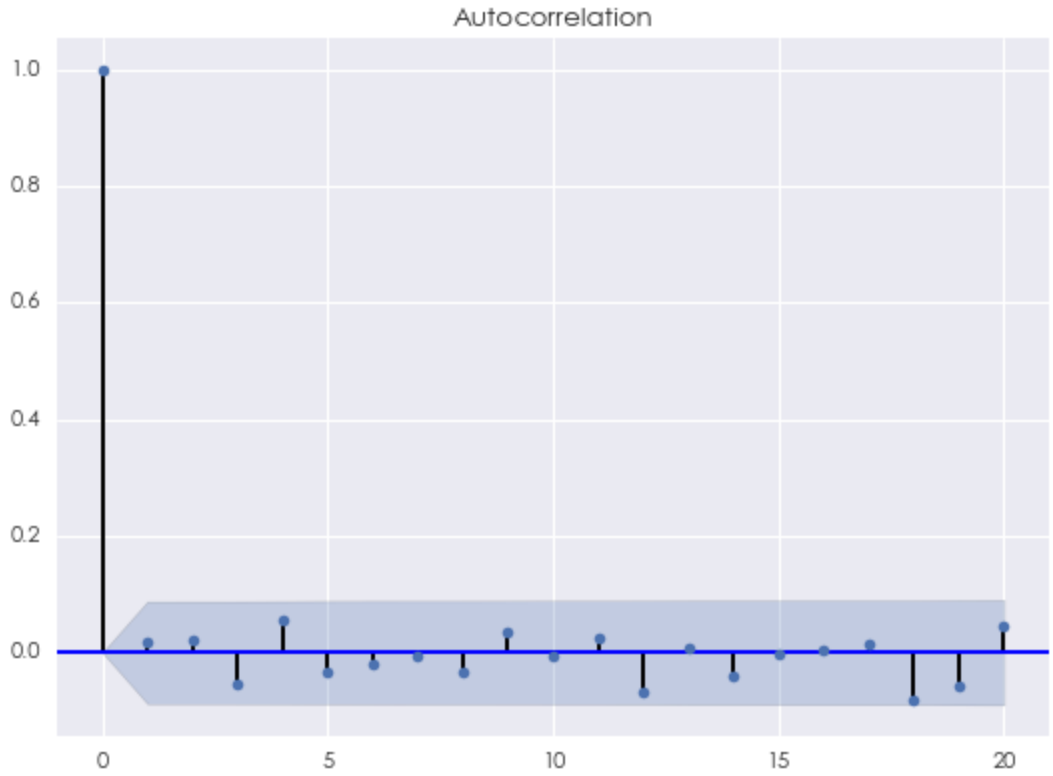
如果把 ζ_k 看作是 k 的函数，则它通常被称为**样本自相关函数**（sample autocorrelation function；同样的，ρ_k 为总体自相关函数），它刻画了时间序列的重要特性。利用相关图（correlogram）可以清晰地看到 ζ_k 是如何随间隔 k 变化的。

下图为两个假想时间序列的相关图。它们呈现出完全不同结构的自相关性。事实上，第一个相关图的时间序列存在明显的趋势；而第二个相关图的时间序列存在明显的周期性。这两个例子说明相关图可以传递出时间序列的很多内在特性。



金融时间序列的相关图虽然远没有这两个假象序列的相关图这么有结构，但相关图在对时间序列建模时至关重要。之前已经说过，金融时间序列，特别是收益率序列，最重要的特性是一些不容易被发现的自相关性。（通常股票的收益率序列没有季节性或者明显的趋势性；即便是弱趋势也可以由自相关性反应。）因此，拿来一个收益率序列，只要画出相关图，就可以检测该序列在任何间隔 k 有无统计上显著的自相关性。

对金融时间序列建模，最重要的就是挖掘出该序列中的不同间隔 k 的自相关性。相关图可以帮助人们判断模型是否合适。这是因为金融时间序列的特征中往往包括相关性和随机噪声。如果模型很好的捕捉了自相关性，那么原始时间序列与模型拟合的时间序列之间的残差应该近似的等于随机噪声。残差序列自然也是一个时间序列，因此可以对它画出相关图。一个标准随机噪声的自相关满足 $\rho_0 = 1$ 以及 $\rho_k = 0, k = 1, 2, 3, \cdots$ 即对于任意不为 0 的间隔，随机噪声的自相关均为 0。下图为一个随机噪声的相关图（用标准正态分布构造了有 500 个点的随机噪声序列）：



关于这个图：

1. 间隔为 0 的自相关系数为 1；
2. 对于任意的 $k \geq 1$，蓝色的阴影区域为 95% 的置信区间。因此，自相关系数只要没有超过蓝色阴影区域，我们就无法在 5% 的显著性水平下拒绝原假设（原假设为间隔为 k 的自相关系数为 0）。上图的结果说明当 k 不为 0 时，随机噪声的自相关系数为 0。

因此，在评价对金融时间序列的建模是否合适时，首先找到原始时间序列和它的拟合序列之间的残差序列；然后只要画出残差序列的相关图就可以看到它是否含有任何模型未考虑的额外自相关性：

- 如果残差的相关图和上面这个图相似，则可以认为残差是一个随机噪声，而模型已经很好的捕捉了原始时间序列中的自相关性；
- 如果残差的相关图体现了额外的自相关性，它们将为我们改进已有的模型提供依据，因为这些额外的自相关说明已有模型没有考虑原始时间序列在某些特定间隔上的自相关。

有了上述基础知识，下节将解读时间序列建模。首先会给出建模框架，紧接着从最简单的白噪声和随机游走出发，说明它们无法有效刻画投资品收益率序列中体现出来的自相关性。这会促使人们提出更高级的模型，包括 AR，MA，以及 ARMA 等。

3 时间序列建模

时间序列模型是一个可以解释时间序列中的自相关性的数学模型。

能够解释时间序列的自相关性在量化投资领域意义重大。人们假设金融时间序列（比如投资品的收益率）存在未知的自相关性（当然也伴随着噪声），而这种自相关性体现了该时间序列某种内在的特性（比如趋势、或者均值回复），而这种内在特性是可以延续的（至少在未来短时间内）。因此，我们希望通过对历史数据的拟合找到一个合适的模型，使得它能最大程度的解释该时间序列表现出来的自相关性。基于未来会重复历史的假设，在统计上预期这种自相关性存在于未来的序列中。由于这个模型考虑了这种自相关性，因此它将会帮助人们预测未来。时间序列分析为人们研究投资品收益率的行为提供了有力的统计学框架。

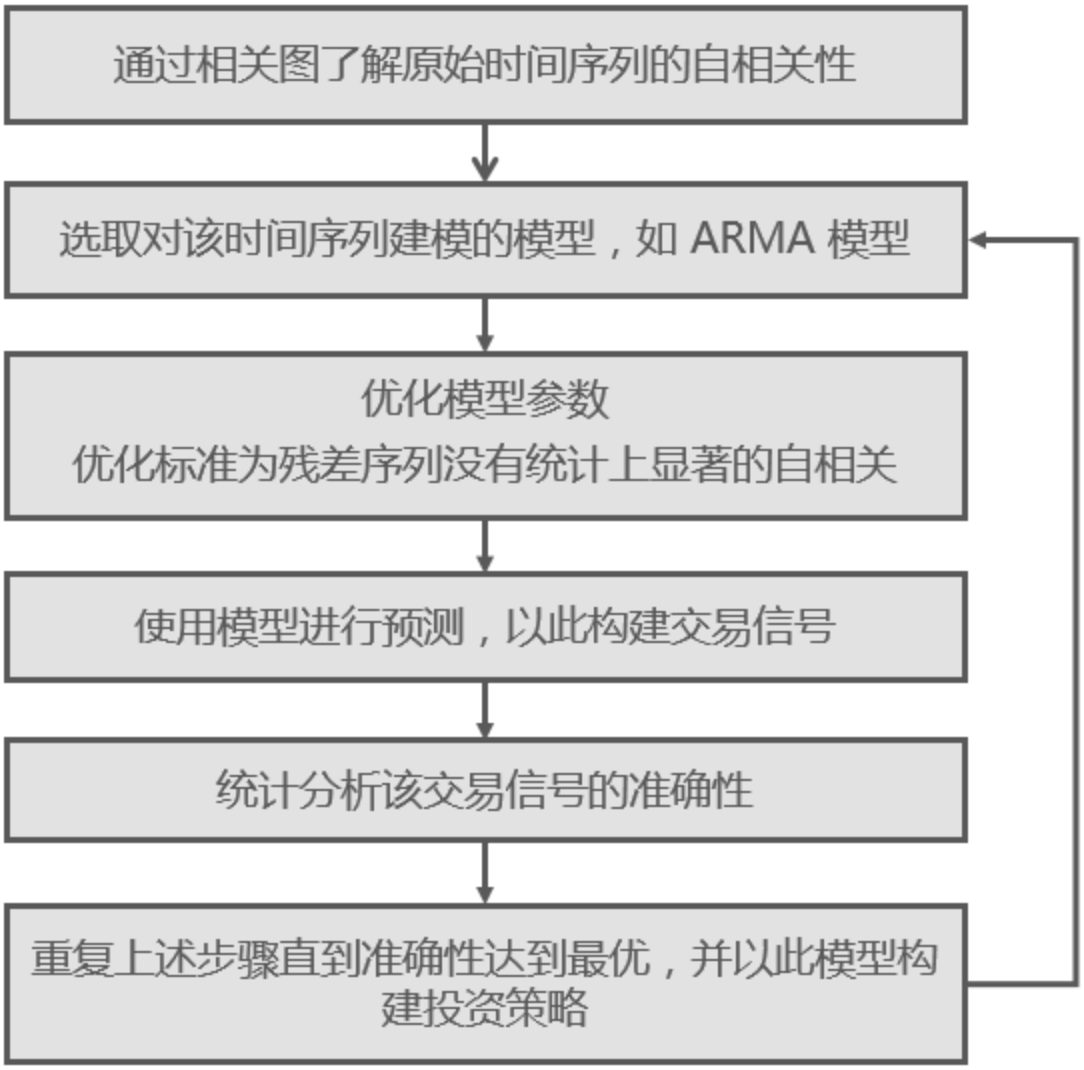
在投资中，如果能够正确预测投资品的涨跌，那么就能基于此构建一个交易策略；如果能够正确预测收益率的波动率，那么就可以进行风险管理（因此人们对时间序列的二阶统计量，如方差，同样感兴趣）。

假设原始时间序列为 $\{y_t\}$，模型拟合出来的序列为 $\{p_t\}$，则残差序列 $\{e_t\}$ 定义为原始序列和拟合序列的差值：

$$e_t = y_t - p_t$$

如果模型很好的捕捉了原始时间序列的自相关性，那么残差序列 $\{e_t\}$ 应该近似的为白噪声，对任何非零间隔 k，该残差序列的自相关系数 ρ_k 都应该在统计意义上不显著的偏离 0。当然，这仅仅是该模型被视为优秀模型的充分条件，因为一个好模型最关键的还是能产生赚钱的交易信号。因此，模型的检验最终还要看它在样本外预测的准确性。

时间序列建模的过程如下图所示。



对于一个时间序列，人们总是希望首先画出它的相关图来看看它存在什么样的自相关性。基于对其自相关性的认知，第二步则是选择合适的模型，比如 AR、MA 或者 ARMA 模型，甚至于更高级对波动率建模的 GARCH 模型等。

选定模型后，接下来便需要优化模型的参数，以使其尽可能解释时间序列的自相关性。在这一步，通过对残差进行自相关性分析来判断模型是否合适。在这方面，Ljung–Box 检验是一个很好的方法，它同时检验给残差序列各间隔的自相关系数是否显著的不为 0。在选定模型参数之后，仍需定量评价该模型在样本外预测的准确性。毕竟，对于样本内的数据，错误的**过拟合**总会得到“优秀”的模型，但它们往往对样本外数据的预测效果很差。因此，只有样本外预测的准确性才能客观的评价模型的好坏。如果模型的准确性较差，这说明该模型存在缺陷，无法充分捕捉原序列的自相关性。这时必须考虑更换模型。这就构成了上述步骤的反馈回路，直到最终找到一个既能解释原时间序列自相关性，又能在样本外有不错的准确性的模型。之后，该模型将被用来产生交易型号并构建量化投资策略。

4 白噪声模型

4.1 白噪声

对于资产收益率来说，**白噪声（white noise）**通常不是一个有效的模型。那么为什么还要研究它呢？这是因为它有一个重要的特性，即**序列不相关：一个白噪声序列中的每一个点都独立的来自某个未知的分布，它们满足独立同分布（independent and identically distributed）**。

一个（离散）白噪声的定义如下：**考虑时间序列 $\{w_t：\ t=1,\cdots,n\}$。如果该序列的成分 w_t 满足均值为 0，方差 σ^2，且对于任意的 $k\geq 1$，自相关系数 ρ_k 均为 0，则称该时间序列为一个离散的白噪声。**

上面的定义并没有假设 w_t 来自正态分布。事实上，白噪声对分布没有要求。**当 w_t 来自正态分布时，该序列又称为高斯白噪声（Gaussian white noise）。**

根据白噪声的定义，一个白噪声序列满足平稳性要求。它的均值和二阶统计量为：

$$\begin{aligned}\mu_w &= 0\\ \rho_k &= \begin{cases}1 & \text{if} \quad k=0\\ 0 & \text{if} \quad k\neq 0\end{cases}\end{aligned}$$

前文已经多次强调，当一个模型很好的捕捉了原始时间序列的自相关性，它的残差序列就应该没有任何（统计意义上显著的）自相关性了。换句话说，一个优秀模型的残差序列应该（近似）为一个白噪声。因此，使用白噪声序列的性质可以帮助人们确认残差序列中没有任何相关性了，一旦残差序列没有相关性便意味着模型是原始时间序列的一个良好的拟合。在白噪声模型中，唯一的参数就是方差 σ^2。

4.2 随机游走

将白噪声模型进行一步延伸，便得到**随机游走（random walk）**模型，它的定义如下：**对于时间序列 $\{x_t\}$，如果它满足 $x_t=x_{t-1}+w_t$，其中 w_t 是一个均值为 0、方差为 σ^2 的白噪声，则序列 $\{x_t\}$ 为一个随机游走。**

由定义可知，在任意 t 时刻的 x_t 都是不超过 t 时刻的所有历史白噪声序列的总和，即：

$$x_t=w_t+w_{t-1}+w_{t-2}+\cdots+w_0$$

随机游走的序列均值和方差为：

$$
\begin{aligned}
\mu_{x_t} &= 0 \\
\\
\mathrm{var}(x_t) &= \mathrm{var}(w_t) + \mathrm{var}(w_{t-1}) + \cdots + \mathrm{var}(w_0) \\
&= t \times \mathrm{var}(w_t) \\
&= t\sigma^2
\end{aligned}
$$

虽然均值不随时间 t 改变，但是由于方差是 t 的函数，因此随机游走不满足平稳性。随着 t 的增加，x_t 的方差增大，说明其波动性不断增加。对于任意给定的 k，通过以下推导给得出随机游走的自协方差：

$$
\begin{aligned}
\mathrm{cov}(x_t, x_{t+k}) &= \mathrm{cov}(x_t, x_t + w_{t+1} + \cdots + w_k) \\
&= \mathrm{cov}(x_t, x_t) + \sum_{i=t+1}^{k} \mathrm{cov}(x_t, w_i) \\
&= \mathrm{cov}(x_t, x_t) + 0 \\
&= t\sigma^2
\end{aligned}
$$

上述推导中使用了独立随机变量的方差可加性。有了自协方差和方差，便可以方便的求出随机游走的自相关函数：

$$
\rho_k(t) = \frac{1}{\sqrt{1 + k/t}}
$$

由上式可知，自相关系数既是时间 t 又是间隔 k 的函数。**ρ 的表达式说明，对于一个足够长的随机游走时间序列（t 很大），当考察的自相关间隔 k 很小时，自相关系数近似为 1。**这是随机游走的一个非常重要的特性，不熟悉它往往容易造成不必要的错误。

举个例子。我们通常假设股价的对数收益率符合正态分布，因此股价对数是一个布朗运动（随机游走的一种特殊形式）。如果当前的（对数）股价是 x_t，由随机游走的特性可知，$t+1$ 时刻的股价的条件期望为 $E[x_{t+1}|x_t] = x_t$，即对下一时点的股价的最好的猜测就是当前的价格。随机游走是一个鞅（martingale）。

假如某个股价预测模型就是用 t 时刻的股价作为对 $t+1$ 时刻的股价的预测，则该模型的预测值和实际值之间的相关系数就等于股价序列的间隔为 1 的自相关系数。如果股价近似的为随机游走，那么由它的性质可知，间隔为 1 的自相关系数非常接近 1。因此上述股价预测模型 —— 用今天的价格作为明天的价格的预测 —— 的预测值和实际值之间的相关系数也非常接近 1。这会给我们造成错觉：这个模型相当准确。不幸的是，这个模型猜测的收益率在任何时刻都为 0，因此它对于我们构建交易信号毫无作用。

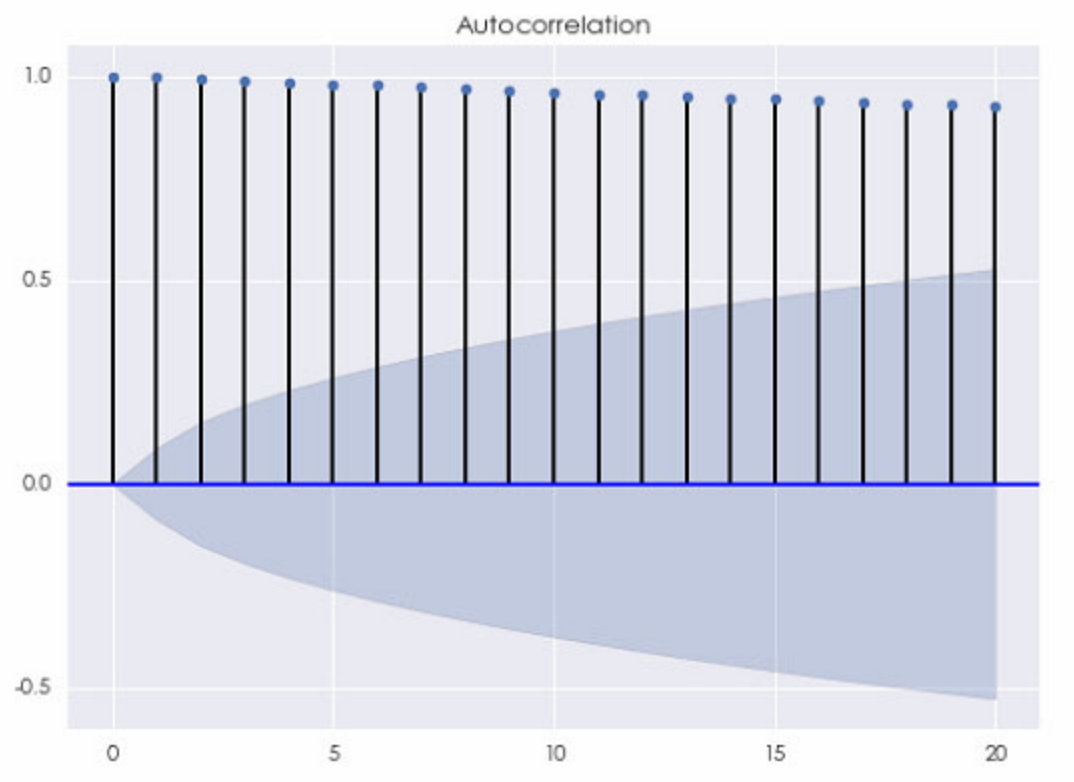
我看到过无数的学术论文（大多是硕士论文）中，针对投资品价格本身构建**自回归**模型。**独立变量就包括历史价格**，用它们和其他一些基本面或宏观经济数据来预测下一个交易日的股价。

从上面的分析可知，这样的模型将会“精准的毫无用处”，因为回归模型中历史价格的系数之和将会非常接近 1。

任何价格序列的自回归模型都是耍流氓。

利用上述白噪声序列，便可以构建一个人工随机游走序列的例子。它的轨迹如下图所示。

不出意外，当间隔 k 相对于时间序列的长度很小时，它的自相关系数（下图）非常接近 1，这源自随机游走的性质。不要忘了，随机游走是对股价的对数建模。因此，这种自相关性对于基于收益率预测的投资策略并没有帮助。



事实上，如果（对数）股价严格的符合随机游走，那么该时间序列的方差将会随时间线性增长。这说明，长期来看它将呈现出巨大的波动。下图为来自同一个分布的 15 条随机游走的轨迹。随着时间的推进，这些轨迹上对应观测值的波动越来越大，充分的展现出随机性。

4.3 对收益率建模

如果股票的对数收益率为白噪声，那么它的自相关系数应该在任何非零的间隔上都在统计意义上等于零。下面就来看看真实的股票收益率是否满足这一点。为此，考虑一支个股（万科）和一个股指（上证指数）。

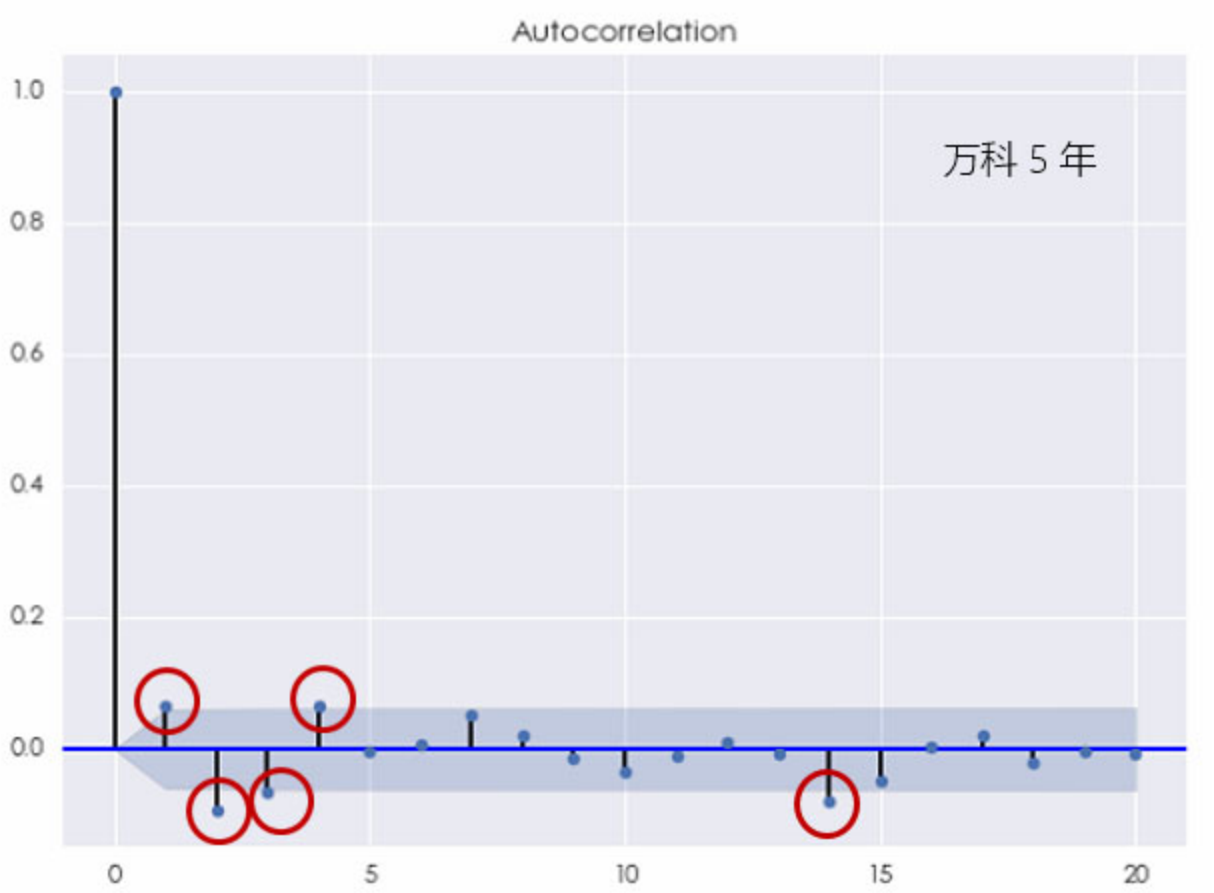
以日频为例，通过交易日的复盘后收盘价可以算出对数收益率：

$$r_t = \ln x_t - \ln x_{t-1}$$

首先来看看万科，当考察期为过去 10 年时，万科的对数收益率的相关图为：

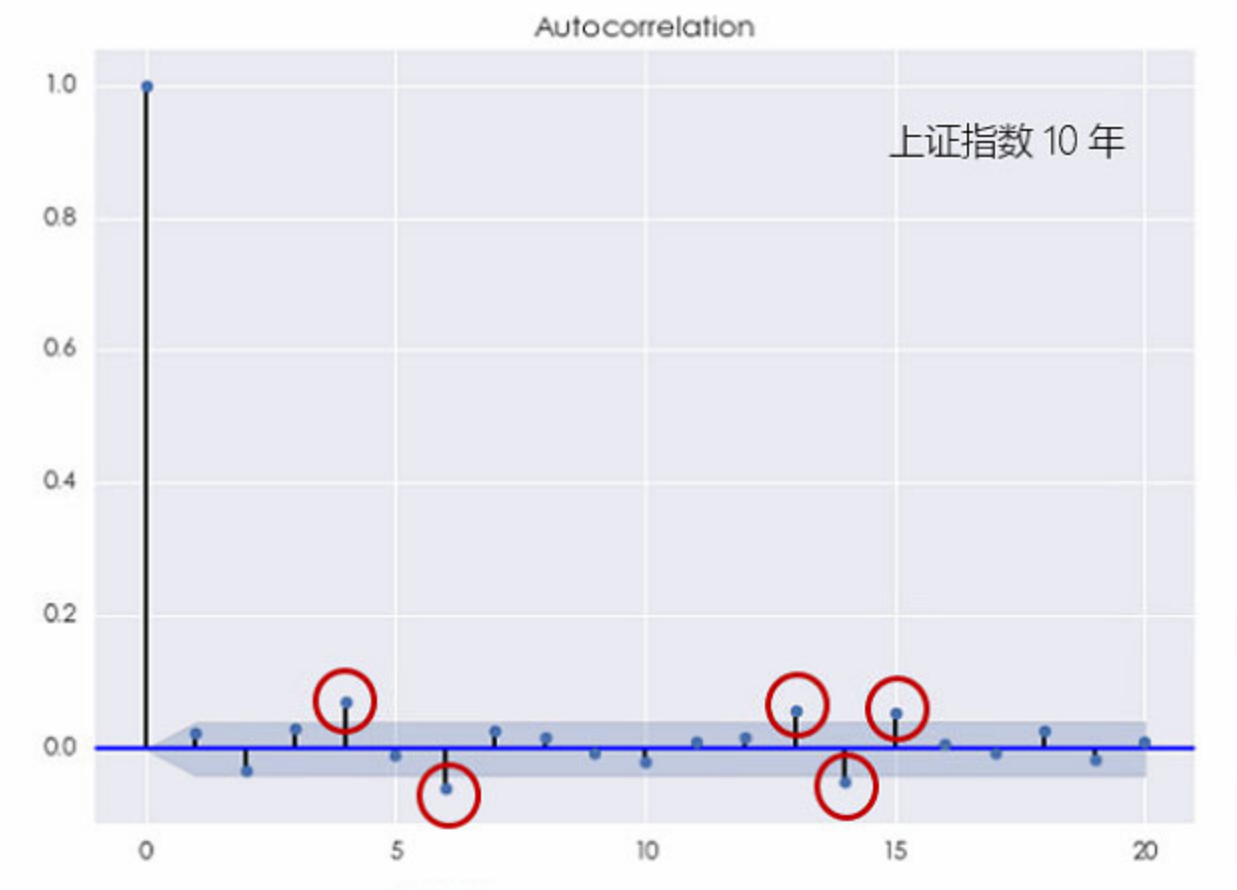
上图指出，在间隔为 2 和 4 时，该收益率序列表现出了统计意义上显著的相关性。当然，由于图中的蓝色区域仅仅是 95% 的置信区间，因此仅仅根据随机性也很可能出现在一个或者两个间隔上的自相关系数处于置信区间之外的情况。因此，根据上面的结果，我们并不能一定就说白噪声不是万科收益率的一个适当的模型。

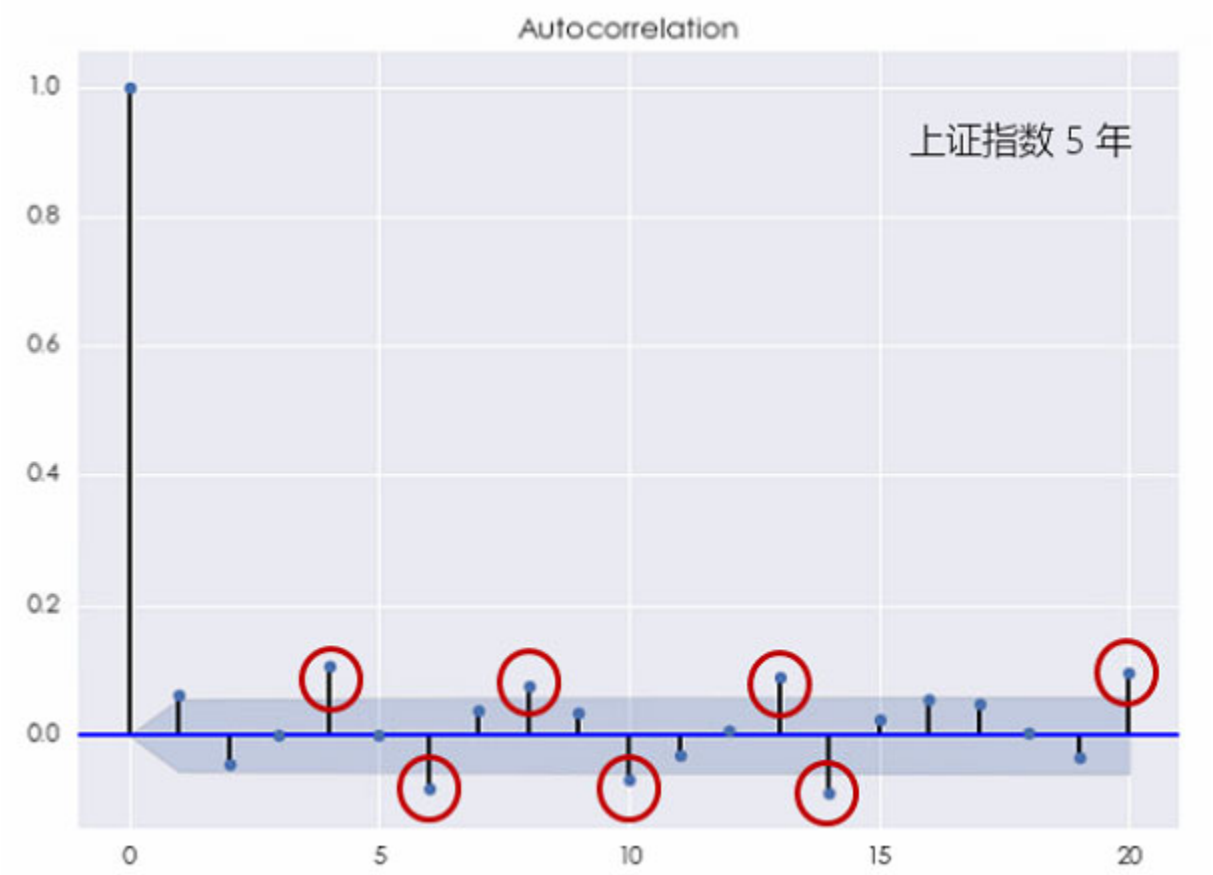
如果把考察的窗口缩短到过去 5 年，则万科的对数日收益率序列的相关图变为：



当 $k = 1, 2, 3, 4$ 以及 14 的时候，自相关系数都超过了置信区间，即在 5% 的显著性水平下不为零。我们无法再无视这样的结果而把它们都归结于随机性。该相关图清晰地说明白噪声不能有效的解释收益率序列中的自相关性。

对于上证指数，这种结论则更加明显。无论是考察 10 年还是 5 年的窗口，上证指数的对数收益率均在不同的间隔上表现出了显著的自相关（下图），且它比个股的自相关性更加显著。





这个结果说明上证指数的对数收益率序列无法用白噪声来建模。更有意思的是，当 k 较小或者较大时，上证指数的收益率均表现出了自相关性，这说明它既有短记忆又有长记忆。

上述分析引出如下的结论，**无论对于个股或是指数，它们的收益率序列中都存在某种自相关性，不满足白噪声模型。**因此，必须考虑更加高级的时间序列模型来对自相关性建模。在这方面，自回归模型（AR）和滑动平均模型（MA），以及它们二者的组合 —— 自回归滑动平均模型（ARMA）—— 都是非常有力的工具。它们将是下节的内容。

5 AR、MA 以及 ARMA 模型

5.1 自回归模型

对于 A 股的收益率，人们往往有这样的感受：

- 在大牛市的时候，股票天天涨（每个交易日的收益率都是正的、鲜有回调），万民欢腾；
- 在大熊市的时候，股票日日跌（每个交易日的收益率都是负的、拒绝反弹），戾气冲天；
- 在震荡市的时候，股票时涨时跌，一买就跌，一卖就涨，颇有价格在某个区间内震荡、收益率呈现均值回复之意。

这些感受给人们的启发是，收益率序列的前后观测点之间往往不是独立的，而是以某种自相关性联系在一起。因此，一个很自然的问题就是：能不能用过去的收益率序列对未来的收益率建模？答案是肯定的。这便引出了**自回归模型（autoregressive model，AR 模型）。**数学上，满足如下关系的时间序列 $\{r_t\}$ 被称为一个 p 阶的自回归模型，记为 AR(p) 模型：

$$r_t = \alpha_1 r_{t-1} + \alpha_2 r_{t-2} + \cdots + \alpha_p r_{t-p} + w_t = \sum_{i=1}^{p} \alpha_i r_{t-i} + w_t$$

这是一个典型的线性回归模型。它和传统线性回归的不同之处在于自变量是序列自身（历史观测值），而非其他变量，这就是自回归中“自”的由来。另外，p 阶的意思是模型使用当前时刻 t 之前的 p 个观测值作为自变量对 r_t 建模。这个模型的含义是，r_t 可以表达为 t 时刻之前的 p 个收益率观测值的线性组合以及一个 t 时刻的随机误差 w_t。p 的取值可以是任何一个正整数，因此最简单的自回归模型就是 AR(1) 模型（$p = 1$）。

在上面这个定义中，并没有考虑截距项。如果截距项对于待研究的时间序列是必要的，则可以在上面的公式右侧加入一个常数项 c。

另外需要特别说明的是，自回归模型**不一定**都满足平稳性。举一个最简单的例子，前述随机游走模型其实就是一个 1 阶自回归模型，满足：$x_t = x_{t-1} + w_t$。由于 x_t 的方差是时间 t 的函数，因此该序列不满足平稳性。

对于一个 p 阶自回归模型，由它的回归系数 α_i 可以写出它的**特征方程（characteristic equation）**：

$$1-\alpha_1 x-\alpha_2 x^2-\cdots-\alpha_p x^p=0$$

它是一个 p 次多项式，有 p 个解，其中可能既包括实数解又包括复数解；这 p 个解的倒数称为该方程的**特征根（characteristic roots）。自回归模型平稳性要求模型特征方程的所有特征根的模都小于 1**（可以通过 augmented Dickey-Fuller test 检验）。在上面的随机游走例子中，该模型的特征方程为 $1-x=0$，它的特征根为 1。由于它不满足模小于 1 这个条件，因此该模型不满足平稳性。

对于一个满足平稳性、且假设没有截距项的 p 阶自回归模型，它的均值为 0（如果有截距项的话，该时间序列的均值就是 c）；它的不同间隔 k 的自协方差 γ_k 和自相关系数 ρ_k 可以表达为如下的递归方程，又称为 Yule-Walker equations：

$$\begin{cases} \gamma_k &= \sum_{i=1}^{p} \alpha_i \gamma_{k-i}, \quad k>0 \\ \rho_k &= \sum_{i=1}^{p} \alpha_i \rho_{k-i}, \quad k>0 \end{cases}$$

在实际中，想要使用自回归模型对收益率建模，必须确定模型的阶数 p。

5.2 滑动平均模型

滑动平均（moving average，即 MA）模型是另一个常见的线性时间序列模型。在自回归模型中，我们将收益率 r_t 看作是给定阶数 p 下历史收益率序列的线性组合。与自回归模型不同，滑动平均模型将收益率 r_t 看作是历史白噪声的线性组合。

这听起来也许有些费解。但它背后的逻辑也符合人们的认知。以美股指数（比如 S&P500 指数）为例，它给我们的印象是它的收益率有一个微弱的但是大于零的漂移率（drift），形成一个常年慢牛的走势。除了这个 drift 项之外，它的收益率呈不规则的波动。在这种背景下，自回归模型仿佛不是那么好用。而滑动平均模型则是对漂移率之外“随机噪声”建模，它把这些噪声理解为不同时刻出现的影响收益率的新息或者冲击（shocks）。通过对“噪声”建模来预测当前时刻 t 的“噪声”，再和漂移率结合，作为 t 时刻的收益率预测。

数学上，满足如下关系的时间序列 $\{r_t\}$ 被称为一个 q 阶的滑动平均模型（为了简化表达式，我们假设漂移率项为 0，即该模型不考虑截距项），记为 MA(q) 模型：

$$r_t = w_t + \beta_1 w_{t-1} + \beta_2 w_{t-2} + \cdots + \beta_q w_{t-q}$$

与自回归模型不同，滑动平均模型一定满足平稳性。它的序列均值为 0（如果考虑截距项，则可以在上式右侧加入一个常数 c 代表漂移率，这时序列均值变为 c）。它的各间隔 k 的自相关系数满足：

$$\rho_k = \begin{cases} 1 & \text{if} \quad k = 0 \\ \sum_{i=0}^{q-k} \beta_i \beta_{i+k} / \sum_{i=0} \beta_i^2 & \text{if} \quad k = 1, \cdots, q \\ 0 & \text{if} \quad k > q \end{cases}$$

其中 $\beta_0 = 1$。

5.3 自回归滑动平均模型

5.1 和 5.2 两节分别讨论了自回归和滑动平均模型。前者用收益率的历史对未来收益率做预测，它背后的逻辑是捕捉市场参与者的有效性（或者非有效性）造成的市场的动量或者反转效应；而后者对噪声建模，其逻辑为突发信息对收益率将会造成冲击（比如上市公司超出预期的财报或者内部交易丑闻等）。

将一个 p 阶的自回归模型和一个 q 阶的滑动平均模型组合在一起，便得到了一个阶数为 (p, q) 的**自回归滑动平均模型（autoregressive moving average model，即 ARMA）**，它将 AR 和 MA 模型的优势互补起来。由于 AR 和 MA 模型都是线性模型，因此它俩的线性组合，即 ARMA 模型，也是线性模型。

数学上，满足如下关系的时间序列 $\{r_t\}$ 被称为一个阶数为 (p, q) 的自回归滑动平均模型（为了简化表达式，假设模型中的不含常数项），记为 ARMA(p,q) 模型：

$$r_t = \alpha_1 r_{t-1} + \cdots + \alpha_p r_{t-p} + \beta_1 w_{t-1} + \cdots + \beta_q w_{t-q} + w_t$$

相比较单一的 AR 或者 MA 模型，ARMA 模型拥有更多的参数。因此它出现过拟合的危险就更高。虽然它能够捕捉到两个单一模型各自所代表的时间序列自回归性，但是在确定阶数 p 和 q 的时候，应时刻谨记防止过拟合。

下面就来看看如何利用**信息量准则（information criterion）**和**残差自相关检验**可以被用来确定 AR、MA 以及 ARMA 模型的阶数。

5.4 确定模型阶数

在实际中使用 AR、MA 或 ARMA 模型对收益率建模，必须确定模型的阶数 p 以及 q。p 或者 q 越大，则模型的参数越多，越有可能捕捉到时间序列中不同间隔 k 的自相关性。但是，参数太

多的话容易造成过拟合。因此在选择阶数时，必须同时考虑拟合的准确性和防止过拟合。

在确定模型阶数时，常用的工具是使用信息量准则，包括**赤池信息量准则**（Akaike information criterion，简称 AIC，由日本统计学家赤池弘次创立）以及**贝叶斯信息量准则**（Bayesian information criterion，简称 BIC）。

这两个信息量准则的目的都是寻找可以最好地解释数据但包含最少自由参数的模型。它们均使用模型的似然函数、参数个数以及观测点个数来构建一个标量函数，以此作为评价模型好坏的标准。它们的区别是标量函数的表达式有所不同。

令 L, k, n 表示模型的似然函数、滞后阶数以及样本个数，则 AIC 和 BIC 的定义分别为：

$$\begin{aligned} \mathrm{AIC} &= -2\ln L + 2k \\ \mathrm{BIC} &= -2\ln L + k\ln n \end{aligned}$$

从定义可知，AIC 和 BIC 都由两部分组成：第一部分衡量模型的拟合度，第二部分是对参数个数的惩罚（防止过拟合）。当一个模型能够很好的解释（样本内）数据时，它的似然函数很大，因此第一项 $-2\ln L$ 就会越小；如果模型的参数越少，则第二项也越少。**所以 AIC 和 BIC 总是越小越好。**

随着模型阶数 p 和 q 的增多，模型对样本内的数据的解释程度越来越高，即 $-2\ln L$ 变小。但是解释度的提高是以参数增多（过拟合风险增大）为代价，因此 $2k$ 或者 $k\ln n$ 增大。所以 AIC 和 BIC 是在这两者之间做权衡。最终选出的最佳参数 $p^\star$ 和 $q^\star$ 可以使它们对应的 AIC 或者 BIC 比其他任何参数 p 和 q 对应的 AIC 或者 BIC 更小。

值得说明的是，AIC 和 BIC 的表达式虽然长得差不多，但是还是有细微的差别。因此在实际中，有可能 AIC 对应的最优阶数（即使得 AIC 最小）和 BIC 对应的最优阶数（即使得 BIC 最小）略有差别。具体选择哪个信息量准则则取决于使用者自身。

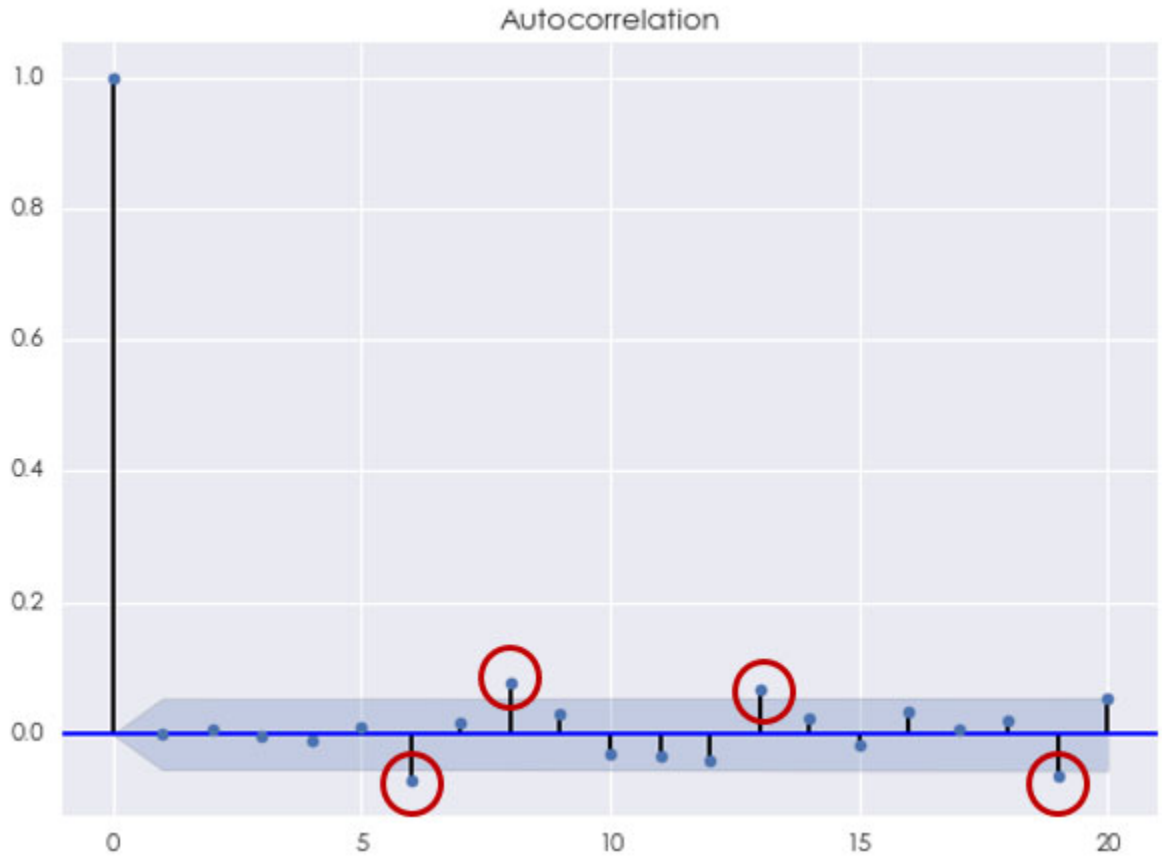
一旦通过 AIC 或 BIC 确定模型的最优阶数，便可以对时间序列建模。但是，我们仍然需要检验该模型是否很好的捕捉了时间序列的自相关性。前文反复强调过，如果一个模型和原时间序列的残差满足白噪声，那么该模型就是合适的。因此，只需要检验残差序列是否在任何间隔 k 上呈现出统计意义上显著的自相关性。在这方面，Ljung–Box 检验是一个很好的方法，**它同时检验残差序列各间隔的自相关系数是否显著的不为 0。**Ljung–Box 检验构建了一个满足卡方分布（chi-squared distribution）的统计量，然后计算它出现的概率，以此来判断是否可以在给定的显著性水平下拒绝原假设。

5.5 对收益率建模

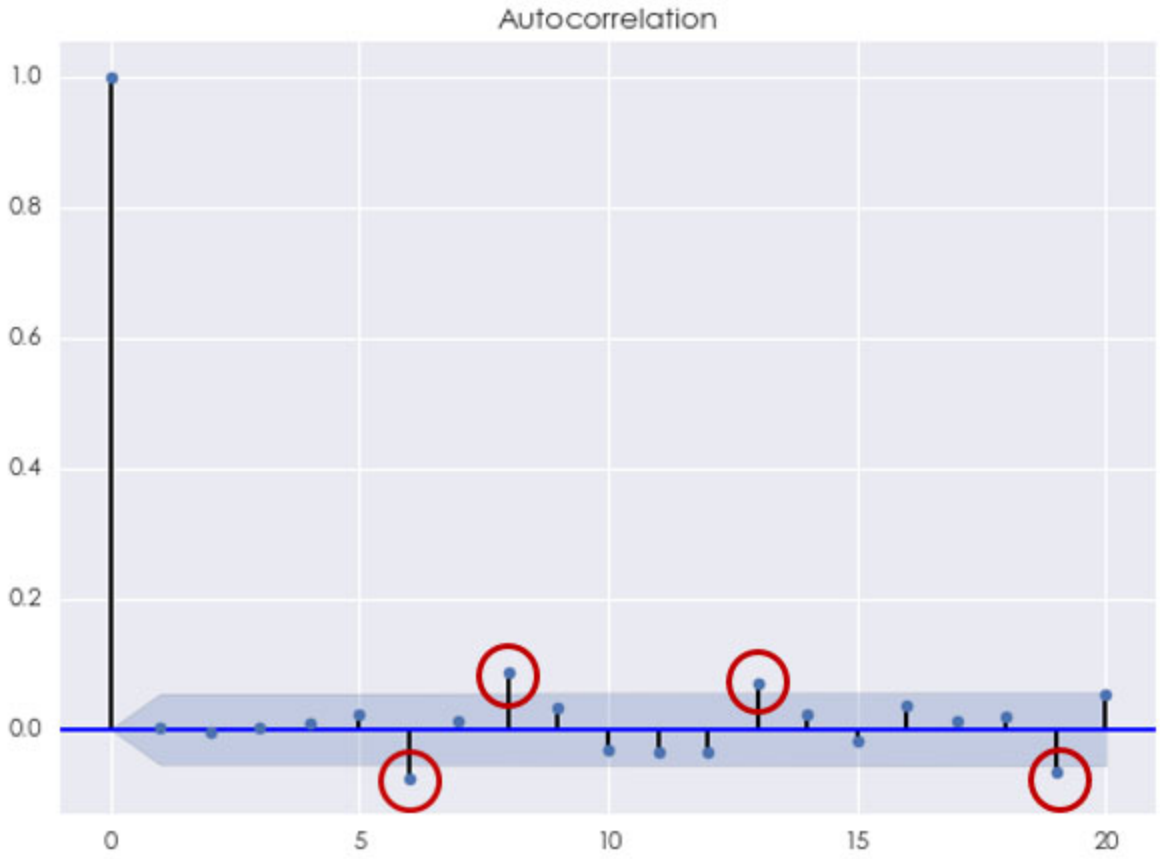
以下利用 AR、MA 以及 ARMA 对上证指数的对数收益率建模。实验考虑 2012 年 4 月 24 日到 2017 年 4 月 24 日这五年之中上证指数的日收益率。在确定模型阶数时，在给定的 p, q 参数区间内使用不同的参数取值建模，并采用 AIC 准则进行参数选择，在建模时让保留常数项。p 和 q 的区间分别为：

- AR 模型：p 的取值范围为 1 到 5；
- MA 模型：q 的取值范围为 1 到 5；
- ARMA 模型：p 和 q 的取值范围为 1 到 5。

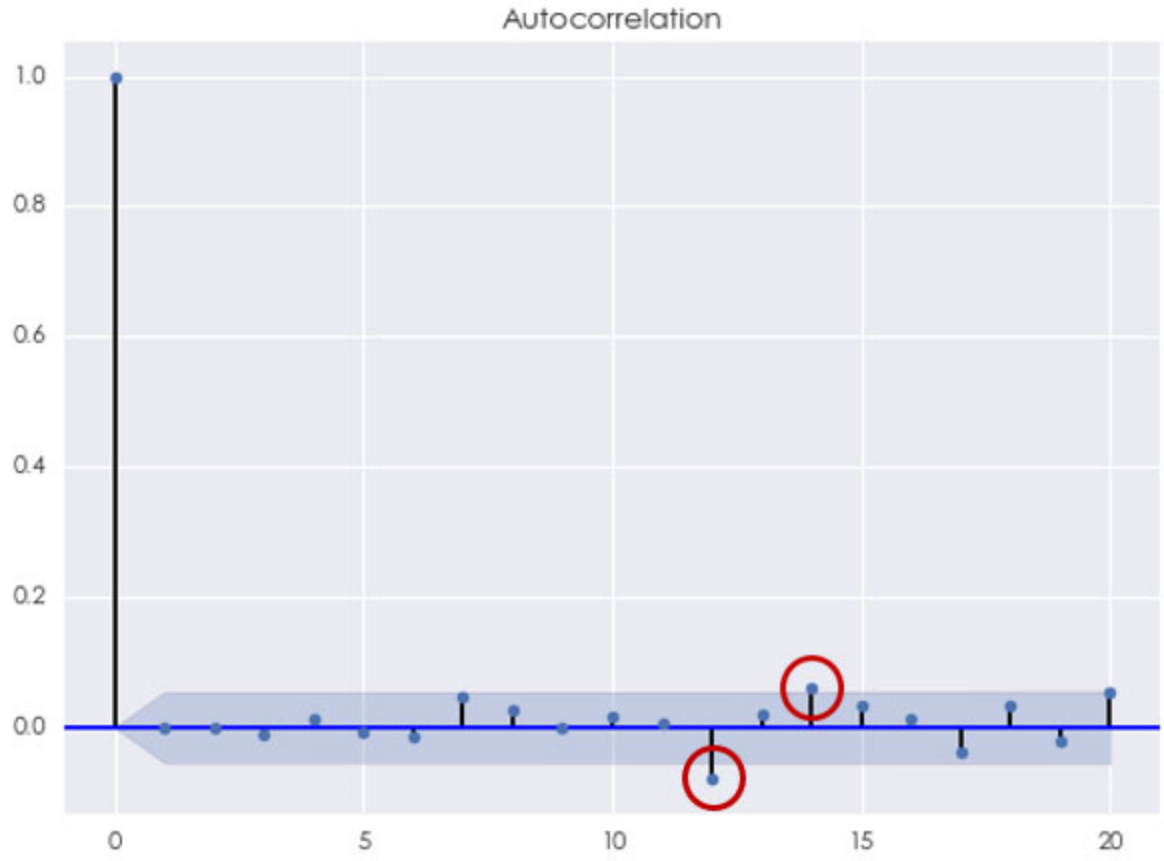
首先来看 AR 模型。根据 AIC 准则，最优的阶数 $p^\star = 4$，此时 AIC = -7305.31。使用 Ljung-Box 检验原始对数收益率序列和 AR(4) 模型的残差是否在 20 以内的间隔上有任何自相关性，统计量的 p-value 为 0.005132，说明可以在 1% 的显著性水平下拒绝原假设。这意味着残差中存在相关性。事实上，这可以从残差序列的相关图中看到，它说明残差序列在间隔 k 等于 6、8、13 和 19 时仍然有 AR(4) 模型未捕捉到的自相关性。



再来看看 MA 模型。根据 AIC 准则，最优的阶数同样为 $q^\star = 4$，此时 AIC = -7302.70。使用 Ljung-Box 检验原始对数收益率序列和 MA(4) 模型的残差是否在 20 以内的间隔上有任何自相关性，统计量的 p-value 为 0.001371。同样在 1% 的显著性水平下拒绝原假设。从下面的残差相关图不难发现，与 AR(4) 模型类似，MA(4) 模型的残差序列在间隔 k 等于 6、8、13 和 19 时仍然有模型未捕捉到的自相关性。



最后来看看 ARMA 模型。根据 AIC 准则，最优的阶数为 $p^\star = 5, q^\star = 4$，此时 AIC = -7330.43。使用 Ljung-Box 检验原始对数收益率序列和 ARMA(5,4) 模型的残差是否在 20 以内的间隔上有任何自相关性，统计量的 p-value 为 0.103462。这说明不能在 10% 的显著性水平下拒绝原假设。它意味着间隔 20 以内，该模型的残差序列没有统计上显著的自相关。从残差序列的相关图中看到，虽然当 k 等于 12 和 14 时自相关系数超过了 95% 置信区间，但我们无法从统计上否定它们可能是来自随机误差。

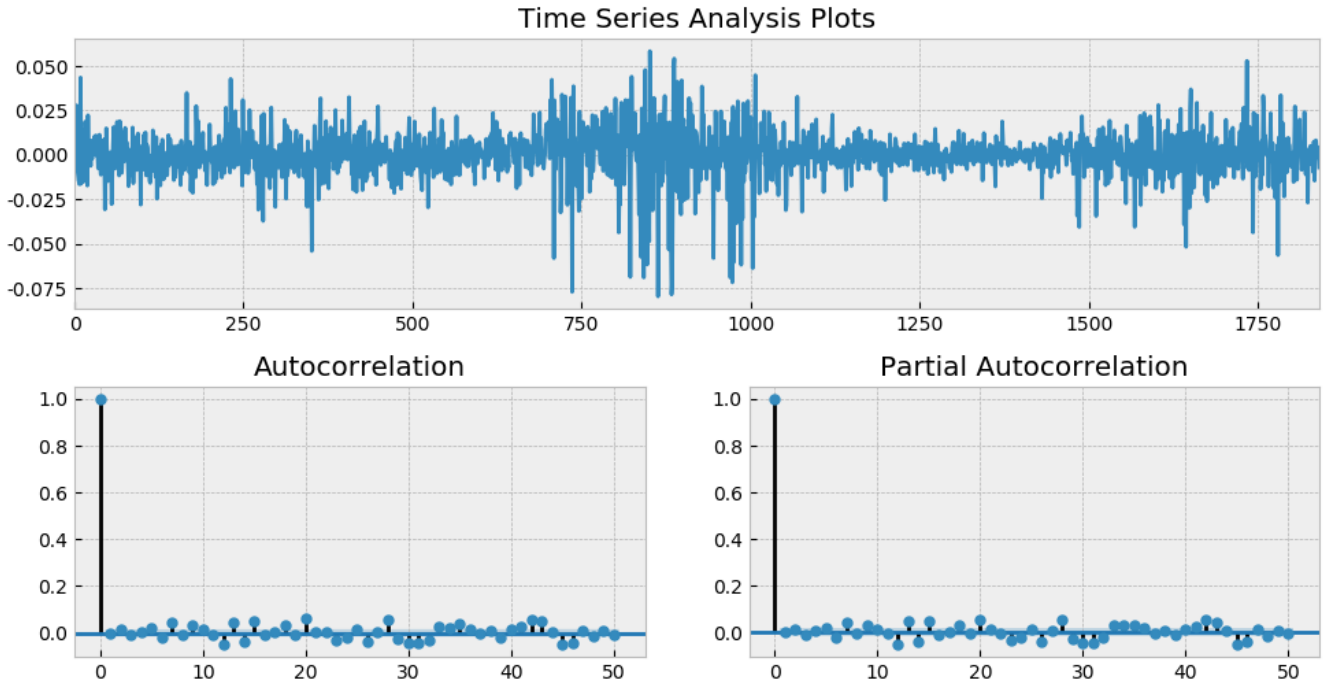


从残差的自相关性分析来看，ARMA 模型比 AR 和 MA 模型单独使用更有效的捕捉了收益率序列中的自相关性。

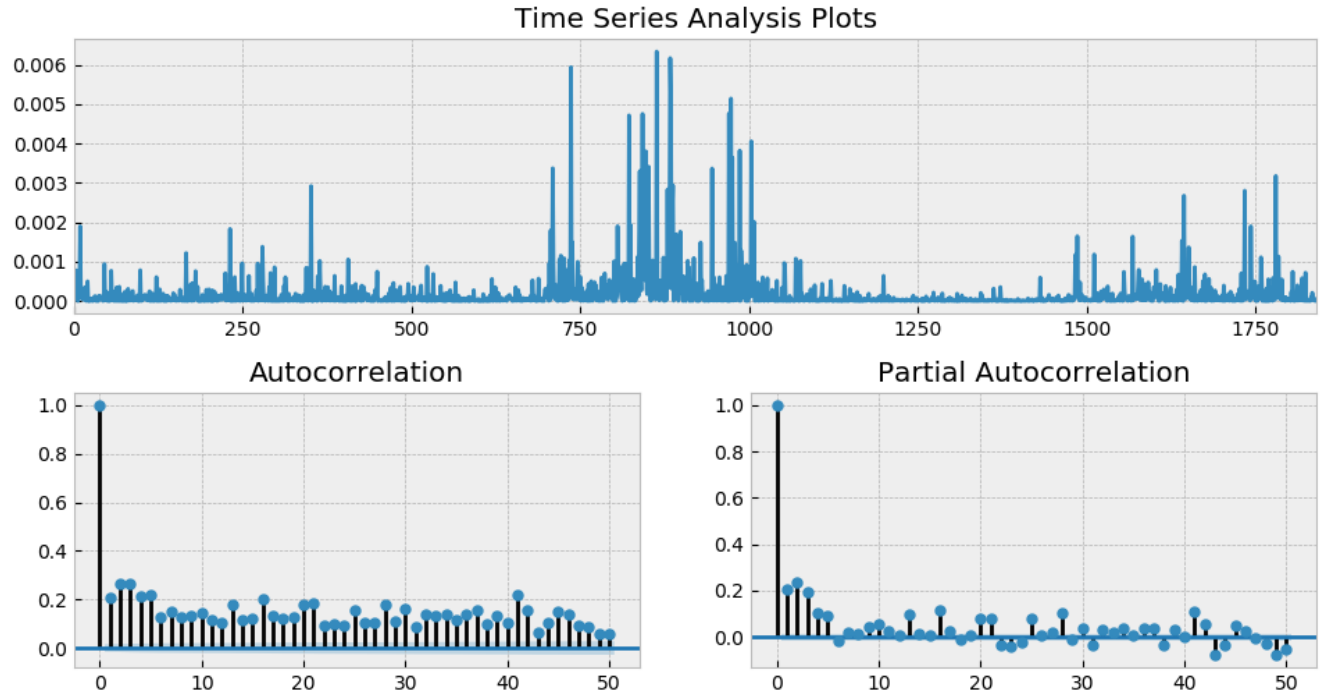
6 GARCH 模型

投资品的收益率序列具有一个重要属性，即**波动聚类（volatility clustering）**。这意味着收益率的波动率是随时间变化的（它是对收益率序列的二阶平稳性假设的直接挑战），这种波动率行为的术语称为**条件异方差（conditional heteroskedasticity）**。

以 2012 年 1 月 1 日到 2019 年 7 月 31 日上证指数日频对数收益率为例，假设使用 ARMA(3, 2) 对其建模，并考察其残差。下图展示了残差时序以及它的 ACF 和 Partial ACF（PACF）。



从 ACF 和 PACF 上不难看出，在很多 lags 上，自相关系数是超过 95% 的置信区间的；而从最上面一副图中也能明显看出收益率序列的一大特征 —— 波动率聚类。如果把残差取平方，并再次作图，上述波动率聚类则会变得更加直观。它在数学上被称为条件异方差。



上述结果意味着，仅使用 ARMA 对收益率序列建模是不够的，它对条件异方差无能为力。针对波动率的特性，人们需要对收益率的平方直接建模。这时，可以使用**自回归条件异方差**（Autoregressive Conditional Heteroskedastic，又称 ARCH）模型和**广义自回归条件异方差**（Generalized Autoregressive Conditional Heteroskedastic，又称 GARCH）模型。GARCH 模型主要用于预测风险，在量化投资中应用广泛。

6.1 GARCH 模型的结构

首先来看看“条件异方差”一词。波动率聚类说明不同阶段收益率的方差是不同的，这就是异方差性（heteroskedastic）。而很多时候，资产收益率表现出高波动伴随着高波动时期（大牛市或者股灾的时候），而低波动又往往伴随着低波动，因此波动率之间是存在序列相关性的，这就是“条件”一词的来源。将二者结合就有了条件异方差。

使用 GARCH 建模，是为了在 r_t 的线性自相关性之上考虑其方差之间的相关性，即把均值模型和波动率模型放在一个整体框架中考虑（Tsay 2010）。假设 $t-1$ 时刻所有已知的信息为 F_{t-1}，则当给定 F_{t-1} 时，t 时刻收益率的条件均值和条件方差可写为：

$$\begin{aligned} \mu_t &= E[r_t|F_{t-1}] \\ \sigma_t^2 &= \mathrm{var}(r_t|F_{t-1}) = E\left[(r_t-\mu_t)^2|F_{t-1}\right] \end{aligned}$$

对于条件均值 μ_t，它可以是一个常数，也可以使用我们已经掌握的 ARMA 模型对其建模。一旦有了 μ_t 的模型，r_t 可以写作：

$$r_t = \mu_t + \varepsilon_t$$

上式中 ε_t 是 t 时刻的扰动或者新息。结合上式和条件方差的定义可知，t 时刻收益率 r_t 的条件方差由 ε_t 的方差决定：

$$\sigma_t^2 = \mathrm{var}(r_t|F_{t-1}) = \mathrm{var}(\varepsilon_t|F_{t-1})$$

从模型结构不难看出，为了考虑条件异方差则需要对 ε_t 建模，而这正是 GARCH 的目标。

6.2 ARCH 和 GARCH

在介绍 GARCH 之前不妨先来看看 ARCH，毕竟 GARCH 只是在它前面加了一个 G（generalized）从而将其推广了。ARCH 由 Engle (1982) 提出，它是第一个对波动率建模的系统性框架。

对于 ε_t，考虑如下模型（其中 w_t 表示均值为 0、方差为 1 的白噪声）：

$$\begin{aligned} \varepsilon_t &= \sigma_t w_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \end{aligned}$$

把 σ_t 的表达式带回到 ε_t 中可得：

$$\varepsilon_t = w_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}$$

这个关于序列 $\{\varepsilon_t\}$ 的模型称作一阶自回归条件异方差模型，也就是最简单的 ARCH(1) 过程 —— 括号里的系数 1 表明自回归模型中只考虑了 lag = 1 阶。为了直观看出方差序列之间的关系，将上式两边平方：

$$\begin{aligned}\mathrm{var}(\varepsilon_t) &= E[\varepsilon_t^2] - (E[\varepsilon_t])^2 \\ &= E[\varepsilon_t^2] \\ &= E[w_t^2]E[\alpha_0 + \alpha_1\varepsilon_{t-1}^2] \\ &= E[\alpha_0 + \alpha_1\varepsilon_{t-1}^2] \\ &= \alpha_0 + \alpha_1\mathrm{var}(\epsilon_{t-1})\end{aligned}$$

上式清晰的显示了 $\mathrm{var}(\varepsilon_t)$ 和 $\mathrm{var}(\varepsilon_{t-1})$ 之间的关系。前面提到，$\{\varepsilon_t\}$ 的模型是一个 ARCH(1) 过程。从 $\mathrm{var}(\varepsilon_t)$ 和 $\mathrm{var}(\varepsilon_{t-1})$ 的关系可知，**一个 ARCH(1) 过程的方差，即 $\mathrm{var}(\varepsilon_t)$，正是一个 AR(1)，即一阶自回归过程。**接下来照猫画虎，将 ARCH(1) 简单推广到多阶 lags，就得到 ARCH(p) 过程：

$$\varepsilon_t = w_t\sqrt{\alpha_0 + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2}$$

类似的，可以说一个 ARCH(p) 过程的方差是一个 AR(p)，即 p 阶自回归过程；这相当于对方差使用 AR(p) 来建模。既然能对方差用 AR(p) 来建模，那么很自然的一个问题就是，为什么不把 MA(q) 也加上得到方差的 ARMA(p, q) 模型呢？如此便引出了 GARCH(p, q)。对于 ε_t，考虑如下模型：

$$\begin{aligned}\varepsilon_t &= \sigma_t w_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2\end{aligned}$$

这就是大名鼎鼎的 **GARCH(p, q) 模型** —— (p, q) 阶的广义自回归条件异方差模型。有了 GARCH 就可以用它对收益率建模了。

6.3 使用 ARMA 和 GARCH 对收益率建模

本小节来看看如何在 6.1 节介绍的体系下使用 ARMA(p, q) 和 GARCH(p', q') 对 r_t 联合建模。为了区分条件均值模型和条件方差模型中的自回归阶数，此处特意用了 (p, q) 和 (p', q') 表示。

将前面的内容整合到一起得到关于 r_t 的模型如下：

$$
\begin{aligned}
r_t &= \mu_t + \varepsilon_t \\
\mu_t &= \theta_0 + \sum_{i=1}^{p} \theta_i r_{t-i} + \sum_{j=1}^{q} \eta_j \varepsilon_{t-j} \\
\varepsilon_t &= \sigma_t w_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p'} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q'} \beta_j \sigma_{t-j}^2
\end{aligned}
$$

当对 r_t 建模时，需要同时指定 mean model（对 μ_t 建模）以及 volatility model（对 ε_t 建模）。上式使用了 ARMA(p, q) 作为 mean model，但根据实际问题也可以使用更简单的模型，比如 μ_t 为常数；使用了 GARCH(p', q') 作为 volatility model。最后使用已有的数据对这两个模型的参数进行联合估计。

在实际应用中，无论使用 python 还是 R 的相关 package，在调用时都要指定 mean model 和 volatility model。举个例子，在 quantstart.com 上一篇使用 ARMA 对 mean 建模、用 GARCH 对 volatility 建模来交易 S&P500 指数的例子中，作者对两个模型同时进行了设定。

In the next code block we are going to use the rugarch library, with the GARCH(1,1) model. The syntax for this requires us to set up a ugarchspec specification object that takes a model for the variance and the mean. The variance receives the GARCH(1,1) model while the mean takes an ARMA(p,q) model, where p and q are chosen above. We also choose the sged distribution for the errors.

Once we have chosen the specification we carry out the actual fitting of ARMA+GARCH using the ugarchfit command, which takes the specification object, the k returns of the S&P500 and a numerical optimisation solver. We have chosen to use hybrid, which tries different solvers in order to increase the likelihood of convergence:

```
>     spec = ugarchspec(
>               variance.model=list(garchOrder=c(1,1)),
>               mean.model=list(armaOrder=c(final.order[1], final.order[3]), include.mean=T),
>               distribution.model="sged")
>
>     fit = tryCatch(
>       ugarchfit(
>         spec, spReturnsOffset, solver = 'hybrid'
>       ), error=function(e) e, warning=function(w) w
>     )
```

出处：https://www.quantstart.com/articles/ARIMA-GARCH-Trading-Strategy-on-the-SP500-Stock-Market-Index-Using-R

在具体 GARCH 建模时可以遵如下步骤（Tsay 2010）：

1. 使用 ARMA 对 r_t 建模以消除任何线性依赖，确定最优参数 $p^\star$ 和 $q^\star$（可以利用 AIC/BIC 来确定）；
2. 对上述模型的残差进行 GARCH 分析；
3. 如果残差中表现出显著的条件异方差，则给定一个波动模型 GARCH(p', q')；
4. 使用历史数据对第一步中的 ARMA(p, q) 和第三步中的 GARCH(p', q') 进行联合参数估计；
5. 仔细检验第四步中拟合出的模型，如有必要则对其进行修改。

以上五步构成了对条件均值和条件方差的联合建模，使用得到的模型就可以对未来的 r_t 以及 $\mathrm{var}(r_t)$ 进行预测。在离开本节之前，我们再来介绍两个使用 GARCH 建模时**不十分正确**的做法（希望能帮你排雷）。

错误做法一：用 ARMA(p, q) 的阶数作为 GARCH(p', q') 的阶数

网上一些资料中提过这样的做法：首先是用 ARMA 对 r_t 建模、确定参数 p 和 q；然后将它们作为波动率模型的阶数，即 GARCH(p, q)，同时在联合建模时仅假设 mean model = constant。这种做法使用从 r_t 线性关系找到的 p 和 q 去对 r_t 的波动率的关系建模，然后又假设 mean model 是常数，着实令人费解。

错误做法二：将 mean model 和 volatility model 拆开估计

这种做法听上去更"靠谱"一些。首先是用 ARMA 对 r_t 建模，确定参数 p 和 q；然后使用 ARMA 模型的残差为被解释变量，对其进行 GARCH(p', q') 建模；第二步中因为被解释变量是残差，因此 GARCH 模型的 mean model = 0，即假设残差均值为零。

这种做法看似合理，但是从条件均值角度来说，它也仅仅是利用了 ARMA 这一步（第二步的 GARCH 建模由于假设 mean model = 0 因此对条件均值不再有影响），而没有利用 ARMA + GARCH 的联合估计考察异方差对收益率序列的影响。通常来说，**就 ARMA 的参数而言，仅使用 ARMA 和联合使用 ARMA + GARCH 的结果是有差异的。**举个例子：使用 AR(2) 和 AR(2) + GARCH(1, 1) 两种方法对收益率建模。下表展示了两种方法建模时，AR(2) 的参数，可以看出它们之间的差异。

	AR(2)	AR(2) + GARCH(1, 1)
θ_0	0.0003 (0.562)	0.0003 (0.932)
θ_1	0.0634 (2.117)	0.0319 (1.051)
θ_2	-0.0513 (-1.714)	-0.0199 (-0.601)

所以，GARCH 模型虽好，但是 use with care。我们应时刻搞清楚是在对什么建模、怎么建模，mean model 是什么、volatility model 又是什么。

7 应用举例

最后通过一个 toy example 总结以下本小册子介绍的内容，介绍 ARMA + GARCH 的应用。

以下对上证指数自 2012 年 1 月到 2019 年 7 月的日频对数收益率进行时间序列建模，并使用该模型预测下一个交易的收益率。如果预测为正则选择持有上证指数，反之则空仓；假设以收盘价成交且不考虑任何交易成本。

在构建策略时，采用长度为 T 的滚动窗口历史数据。首先是用 AR 对收益率建模（因为 python arch package 不支持 ARMA 作为 mean model，所以仅使用 AR(p) 模型），并根据 AIC 选择最优 p 值（p 取值范围为 0 到 5）；然后以该 AR(p) 作为 mean model，并使用 GARCH(1, 1) 模型为 volatility model，进行联合参数估计。使用最终的模型预测下一个交易日收益率。此外作为比较，我们也考虑仅采用 AR(p) 来对收益率建模，而不考虑条件异方差的影响。

首先来看 $T = 60$ 个交易日的情况。下图展示了 AR 和 AR + GARCH 两种策略的净值和回撤曲线。就表现而言，它们均战胜了上证指数本身（benchmark）。但是在股灾之后（波动率变大了），这两种模型的表现发生了分化，就这个简单实证而言，AR 的效果比 AR + GARCH 更好。

	上证指数	AR	AR + GARCH
年化收益率（%）	1.71	9.42	6.69
夏普率	0.15	0.69	0.51
最大回撤	-57.01	-20.12	-30.87

再来看看把滚动窗口长度换到 $T = 252$ 的情况。结果和上面接近，依然是 AR 战胜了 AR + GARCH 的组合。

	上证指数	AR	AR + GARCH
年化收益率（%）	1.71	9.82	6.94
夏普率	0.15	0.69	0.51
最大回撤	-57.01	-26.94	-36.96

从本小节的例子来看，加入了 GARCH 的策略似乎并没有仅使用 AR 的策略优异。从实证结果来看，加入 GARCH 似乎没什么效果。但不要忘了，我们并没有对 GARCH 的参数进行任何优化，也没有额外利用其对波动率的建模来添加更加复杂的规则 —— 比如 volatility scaling。因此，仅仅基于这个简单的例子难以对 GARCH 的贡献做出任何正确的评判。

最后，在上述举例中，因 python arch package 的功能所限，实证中的 mean model 仅采用了 AR 模型。感兴趣的小伙伴不妨尝试 R 的相关 packages，对收益率采用 ARMA 模型。

8 结语

对于量化投资的研究来说，构建出策略并看到回测出来的净值曲线无疑是最令人激动的。然而，真正研究工作的核心却在于搞懂每个模型的原理以及它的作用，而这个过程注定是枯燥的。希望本小册子带你走进时间序列分析，并通过它更好的研究资产的收益率。

本小册子的写作也重点参考了 quantstart.com 上时间序列的一系列文章（链接见参考文献），特此说明。Quantstart Team 在其时间序列分析系列文章（以及其他系列）中不厌其烦的介绍每个基础模型，从简单到复杂，像搭积木一样为读者构建知识体系，令人敬佩。建议感兴趣的读者去读一读，一定会有启发。

现在，我们有了时间序列分析的各个 building blocks。但是，能够用它们做什么、如何去更科学的对收益率分析、预测，还需有经验的积累。最后，我想以下面这段出自 Quantstart 的话作为本小册子的结束，也希望与各位共勉。

> *True quantitative trading research is careful, measured and takes significant time to get right. There is no quick fix or "get rich scheme" in quant trading.*

参考文献

- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. ***Econometrica*** *50*(4), 987 – 1008.
- Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed). Wiley.
- https://www.quantstart.com/articles/topic/time-series-analysis/

免责声明

本网站所含内容由 BetaPlus 小组基于公开信息而提供。入市有风险，投资需谨慎。在任何情况下，本网站中的内容、信息及数据或所表述的意见并不构成对任何人的投资建议。在任何情况下，BetaPlus 小组不对任何人因使用本网站的任何内容所引致的任何损失负任何责任。本网站中内容所表述的意见不代表 BetaPlus 小组成员所属机构观点。