提出原假设：

1. 存在某种周历、月历效应，属于以随机试探的形式寻找具体可能存在的该类日历效应，对该现象出现的原因在未做进一步数据分析时不作任何主观猜测；
2. 存在假日效应，休市的时间内股票暂停交易，但与股市行情高度相关的信息仍然在被随机地释放，且无法动用的购置股票所用资金的机会成本随休市时间同步增长，休市后的首个交易日股市有较高概率出现收益率的异常波动。

所用数据与处理方法：

本文所用的基础数据包括上证指数、深证成指与创业板指2010年XXXX至XXXX的日收盘价数据与日内五分钟高频交易数据以及其他辅助数据。本文分别对三个股指对数收益率及其已实现波动率进行R-GARCH建模分析相应的日历效应并解析其背后的影响因素。鉴于所选时间内股市行情差别较大，对收益率及已实现波动建立隐马尔科夫模型，判断其背后的股市风格（隐态），并将股市风格发生转变节点作为数据分段点，以分别研究相应的日历效应。

模型选择：

本文将基于GARCH模型对中国股市的日历效应进行检验以及后续的影响因素分析，GARCH模型对收益率与其波动率联合建模的方法更加符合资产收益率的真实情况，采用该类方法检验日历效应更为准确。

Engle（1982）首次提出了自回归条件异方差模型（Autoregressive Conditional Heteroscedastic，ARCH），为波动率建模提供了系统框架，其认为资产收益率的扰动是序列不相关的，但不是独立的，且这类不独立性可以用扰动滞后项的简单二次函数来描述。

Bollerslev（1986）为了使ARCH模型拥有更灵活的滞后结构，类似AR过程扩展为ARMA过程将ARCH过程扩展为广义ARCH（Generalized-ARCH，GARCH）过程。

为了在模型中体现正负收益率的非对称性，Nelson（1991）与Glosten、Jagannathan和Runkle（1993）分别提出了指数GARCH模型（Exponential-ARCH，EGARCH）与门限GARCH模型（Threshold-ARCH，TGARCH）。

Hansen、Huang和Shek（2011）提出，在高频金融数据已广泛使用的背景下，现有文献介绍了许多已实现波动率的测量，例如由Andersen和Bollerslev提出的已实现波动率（1998），这类已实现波动率相比于以往收益率（扰动）平方而言，在对波动率建模以及预测上都更加实用、更具有信息性，并以此建立了已实现GARCH模型（Realized-GARCH，R-GARCH）。

结合以上各类GARCH模型的特点，本文将使用R-GARCH（Realized-GARCH）模型来检验中国股市的日历效应：首先，已实现GARCH模型由于添加了日内已实现波动率对股指收益率有更好的拟合能力（Christoffersen et al.，2010），相比于普通GARCH模型建模更加充分；其次，R-GARCH模型中的测度方程能识别波动的非对称性，与股市正、负向波动带来的收益率变动非对称性相符；再者，使用GARCH族模型进行建模在理论上使得模型贴合股票收益率的条件异方差性，且可识别波动率的聚集特点；最后，在对股指收益率拟合能力更好的模型中加入日历效应后，其检验到的日历效应更具有准确性。

模型描述与解释：

第一个方程为均值方程，该方程左边是股指对数收益率，方程右边是ARMA(p,q)部分加代表日历效应的其他外生变量，是独立同分布且均值为0方差为的收益率扰动项（error项），的具体分布可根据建模需要而变化；

第二个方程为波动率方程，即对收益率扰动的对数方差进行建模，方程右边为之后q阶的对数已实现波动率以及滞后p阶的对数方差，类似GARCH模型的波动率方程，只是将收益率扰动替换为已实现波动率，该方程表明过去更大的已实现波动率与收益率的波动会导致未来更大的收益率波动，依然符合资产收益率波动率聚集现象；

第三个方程为测度方程，是对数已实现波动率表示为收益率扰动的方差与标准化收益率扰动函数的具体形式，是的杠杆函数，具体形式取，预期为负且为正，来自的负向变动导致的对数已实现波动率变动并传导至方差（将第三个方程代入第二个方程右边）的变动大于来自的正向变动；

方程之所以采用对数线性形式，是由于天然地,其中是收益率扰动项，实际也是,其期望值即日收益率的方差，自然十分接近日内已实现波动（对日收益率方差的近似），故而第三个方程扩展地写成是与的函数形式，能够更准确地测量估计。

额外加入测度方程的R-GARCH模型相比于普通GARCH模型多了已实现波动率这一桥梁，若将第三个方程代入第二个方程右边的，得到的新方程与能够反应收益率扰动正负变化带来波动非对称变化的指数GARCH模型十分接近，只是杠杆函数略有不同。

确定R-GARCH模型的波动方程阶数：

相较于ARCH模型可使用均值方程残差项平方序列的偏自相关函数来确定阶数，GARCH族模型的阶数不易确定，且高阶GARCH模型会出现不稳定的情况，对较为严格的拟合或预测任务通常通过排列组合的方式来结合贝叶斯信息准则（Bayesian Information Criterion, BIC）或者赤池信息准则（Akaike Information Criterion, AIC）确定GARCH项与ARCH项的滞后阶数。本文的重点不在于精确拟合或者预测，无需建立过于复杂的R-GARCH模型，因此所有模型的波动方程初始均设定为GARCH（1,1）的形式，仅当回归结果的均值方程残差项未通过充分建模检验时才重新设定更为复杂的模型。

模型中均值方程收益率扰动服从分布的选择：

标准化的收益率扰动通常被认为服从标准正态分布，由于本文建模的对象是具有尖峰厚尾特点的金融时间序列，故本文首先检验何种分布更适合所选用的数据，将从标准正态分布、具有厚尾特征的学生t-分布、GARCH模型常用的广义误差分布三种分布中选择。

分布的选择遵循以下步骤：（1）检验所选股指的对数收益率序列是否具有GARCH效应；（2）取用某一种分布，在均值方程中不加入任何代表日历效应的外生变量情况下得到回归估计结果；（3）检验R-GARCH模型是否建模充分；（4）比较三种分布下标准化残差项与所用分布的QQ图以及KS检验结果并选择最优分布。