以建模模板template.r为例说明建模流程和思路，并提醒注意事项。

建模流程分为以下五步：载入报表数据（Y）、读取行业数据（X）、合并数据、建模和输出。

下面分别予以说明。

# 载入报表数据（Y）

针对某一指标（如净利润、营收、现金流等）做清洗，整理为每个季度各公司的平均值

作为建模的Y

报表数据来自于report.rda，根据申万行业分划提取需要的子行业。报表中的公司财务指标均为累计值，且某些公司未能每季度均给出报表。

利用report\_clean函数做清洗，整理为每个季度各企业的指标均值。

对于季报齐全的公司，将每季报告的累计值做差后得到各季净值

对于季报不齐全的公司，用平均法补齐

而后对所有公司各季度的报告值做平均，得出该季度的行业均值。

# 读取行业数据（X）

读入行业数据（命名为chain），将第一列处理为时间格式。

利用Ave函数将行业数据整理为季度数据，此处允许不同更新频率的同时出现，但除时间外必须都为数值型

函数输入：

dat为待整理的数据框，允许不同更新频率的同时出现，也允许空值和0

t为时间标签的列号，默认在第一列，格式为POSIXlt，包含年月信息。此列不可有缺。

ac为FALSE时计算流量数据，ac为TRUE时计算累计值数据。注意数据框既有流量数据又有累计值数据时需要将两部分分开分别调用。

当行业数据来自多个表时，将它们分别读入处理后合并即可

# 合并数据

将行业数据中有缺失的变量整列删除。

考虑到外界影响的传递时间，对行业数据做提前期处理，作为待选变量。

考虑指标的自回归性，对y做提前期处理，作为待选变量。

添加季节因子，在建立模型后可以根据实际需要合并季节，减少模型变量数。

# 建立线性回归模型

通过Model函数建立模型。

由于变量过多所以根据marginal-rsquare和aic准则进行筛选。将所有变量按marginal-rsquare排序，逐个尝试加入模型。模型是否接受该变量的标准是AIC准则。

为限制自相关性，利用vif进行限制。若加入新变量后模型内任意变量vif值大于设定的阈值，则拒绝该新变量。

函数参数：

Y 回归因变量

X 回归自变量

must 手动设定加入模型的变量列号。注意该列号是X的列号，而非全数据集的列号。

vif 根据vif限制自相关的阈值，默认为15

method aic或adj.r2，模型评价的指标（建议使用aic）

模型输出：

model 函数筛选得到的模型

index 入选模型的变量列号，同样是X的列号而非全数据集列号

模型检查和调整

根据变量p值、诊断图等手段调整，删除不显著的变量或调整合并季节因子。

# 模型输出与预测

利用修改后的模型和已搜集到的行业数据，对2017二季度的指标进行预测。

csv文件为真值与预测值，txt文件为模型变量和参数，以及对下一期的预测

注意对季节因子合并后，预测可以正常进行，但输出的模型不会显示出因子合并的相关规则，目前也没有输出说明，需手动在代码中查看。

此模板适用于各个行业的各个指标，只需修改工作路径和变量名，并根据需要做细微调整即可。