1. 用以训练基础模型的因子：

开盘价，最高价，最低价，收盘价，成交量，成交额，涨跌幅，振幅，均价，总市值，市盈率，市净率，市销率，市现率，股息率。

1. 训练参数：
   1. 模型：多项式贝叶斯
   2. 交叉验证：80%作为训练集，20%作为测试集，交叉验证100次
   3. 数据处理：向前查看N（0~15之间）天+下列任意一种处理方式
      1. 标准化：令数据的均值为0，方差为1
      2. 中心化：令数据的均值为0，协方差阵不变
      3. 极差标准化：均值为0，极差为1
      4. 离散化：默认离散化为10类
   4. 选择条件：交叉验证100次后，获得原模型准确度的均值xm\_before与标准差xstd\_before，加入因子后，如果其均值超过了xm\_before + threshold\*xstd\_before，则认为该因子有效。其中threshold为人工控制的有效性阈值，当有效性条件较高时，可以增大阈值；较低时减小阈值。
2. 因子测试：
   1. 测试次数：5次，投票+人工筛选方式寻找有效因子。当5次测试中，有3次或以上满足条件，即认为该因子有效。
   2. 因子组：一致预期EPS、2年复合增长率、一致预期归属母公司净利润、一致预期PE、一致预期PE/G、一致预期净利同比、一致预期ROE、一致预期净资产、一致预期PB、干扰数据
      1. 均值筛选结果：一致预期EPS、2年复合增长率、一致预期归属母公司净利润、一致预期PE、一致预期净利同比、一致预期ROE、一致预期净资产、干扰数据
      2. 均值/方差筛选结果：一致预期EPS、2年复合增长率、一致预期归属母公司净利润、一致预期净资产
   3. 因子组：滚动净利润、滚动PE、干扰数据
      1. 均值筛选结果：滚动净利润、干扰数据
      2. 均值/方差筛选结果：干扰数据
   4. 因子组：滚动净资产、滚动市净率、滚动净利润、滚动PEG、干扰数据
      1. 均值筛选结果：全部有效
      2. 均值/方差筛选结果：滚动净资产、滚动市净率、干扰数据
3. 测试报告：

在只考虑均值的情况下，基本上所有的因子都是有效的，所以引入了0.1倍标准差的阈值作为有效性判断条件，在给定阈值下，最终筛选结果有效的因子为：

一致预期EPS、2年复合增长率、一致预期归属母公司净利润、滚动净利润、滚动净资产、滚动市净率。

在具体测试过程中，引入了random的干扰数据，在测试过程中基本都是有效的（而且在一定情况下，其效果比一般的因子更好）；说明模型无法排除随机数据的干扰，在下一步模型改进过程中应当着重对这方面进行改进。