

HMM 模型择时及配置策略

机器学习资产配置系列之一

投资要点:

► 资产配置的本质是资产选择以及市场择时

资产配置就是在给定的预期收益条件下，期望组合风险最小；在给定期望组合风险条件下，期望投资收益最大。实现其目标主要包括资产选择和权重配置两部分。

► 隐马尔可夫模型的本质是对复杂数据进行降维，学习低维隐藏状态的转移进行预测

隐马尔可夫模型将每一个可观测值的产生对应着市场状态的变化，通过观测序列得知对于目前市场状态建模，进而帮助择时判断。市场状态是隐藏状态不可观测的，其通过可观测变量来进行推断。模型本质上是将市场观测状态用隐藏状态进行降维，之后再对低维的隐藏状态空间拟合状态转移，根据状态转移概率对未来市场状态形成预测。

► 基于隐马尔可夫模型的择时策略具有高收益，高胜率，高夏普比率，低回撤率等特点

由于隐马尔可夫模型可以对市场变化有着较高的拟合优度，所以可以根据模型对市场变化的判断进行交易，通过在沪深 300 上的回测表明模型具有高收益，高胜率，高夏普比率，低回撤率的优点。优秀的择时策略是资产配置策略的基础。

► 基于 HMM 模型择时在风险和收益两个维度有着很好的表现

由于 HMM 模型对市场变化趋势有着较高的拟合度，所以可以根据模型对市场变化的判断进行交易。对沪深 300 指数的回测结果表明模型具有高收益，高胜率，高夏普比率，低回撤率的优点。

风险提示

模型基于对历史数据统计，仅作为投资参考。

评级及分析师信息

证券分析师：王湘
SAC NO: S1120520080004

证券分析师：杨国平
SAC NO: S1120520070002

助理分析师：周游
邮箱: zhouyou2@hx168.com.cn

正文目录

1. 隐马尔可夫模型基本理论.....	4
1.1. 隐马尔可夫模型的定义.....	4
1.2. 隐马尔可夫模型的 3 个基本问题.....	4
1.3. 学习算法.....	5
1.3.1. 监督学习算法.....	5
1.3.2. 无监督学习算法.....	6
1.4. 预测算法.....	6
1.4.1. 近似算法.....	6
1.4.2. 维特比算法.....	7
1.5. 小结.....	7
2. 隐马尔可夫模型的择时建模.....	7
2.1. 模型合理性的讨论.....	8
2.2. 模型优缺点的讨论.....	8
2.3. 策略设计思路.....	9
2.3.1. 建模方法.....	9
2.3.2. 预测方法.....	9
2.3.3. 数据及参数选择.....	9
2.4. 策略具体过程及回测分析.....	10
2.4.1. 隐藏状态的马尔可夫性检验.....	11
2.4.2. 不定长滚动训练分析.....	12
2.4.3. 定长滚动训练分析.....	15
2.5. 择时策略讨论.....	17
3. 基于隐马尔可夫择时的资产配置策略.....	17
3.1. 资产配置.....	17
3.2. 基于隐马尔可夫择时的资产配置.....	18
3.3. 资产配置策略讨论.....	21
4. 总结.....	21
5. 风险提示.....	22

请仔细阅读在本报告尾部的重要法律声明

图目录

图 1 隐马尔可夫示意图.....	5
图 2 持仓时间平均收益率.....	10
图 3 5 日平均收益率.....	10
图 4 5 日和持仓期的平均成交量之比.....	10
图 5 持仓时间长度内夏普比率.....	10
图 6 隐藏状态数量为 3 时候通过维特比算法对观测变量进行解码效果.....	13
图 7 隐藏状态数量为 6 通过维特比算法对观测变量进行解码效果.....	14
图 8 隐藏状态为 3 时的隐状态分时买入策略净值图.....	14
图 9 隐藏状态为 6 时候的分时买入策略净值图.....	14
图 10 定长滚动训练回测.....	16
图 11 不定长滚动训练回测.....	16
图 12 定长滚动训练回测累积超额收益率.....	16
图 13 不定长滚动训练回测累积超额收益率.....	16
图 14 等权配置模型不同调仓周期累计收益走势图.....	18
图 15 均值方差配置模型不同调仓周期累计收益走势图.....	19
图 16 风险平价配置模型不同调仓周期累计收益走势图.....	20

表目录

表 1 隐藏状态数量变化进行马氏性检验.....	11
表 2 持仓周期变化进行马氏性检验.....	12
表 3 观测变量组合对回测结果的影响.....	12
表 4 隐藏状态数量对回测结果的影响.....	13
表 5 持仓周期对回测结果的影响.....	15
表 6 定长滚动训练和不定长滚动训练的比较.....	17
表 7 等权配置下的策略表现.....	19
表 8 均值方差模型配置下的策略表现.....	19
表 9 风险平价模型配置下的策略表现.....	20

请仔细阅读在本报告尾部的重要法律声明

1. 隐马尔可夫模型基本理论

隐马尔可夫模型 (Hidden Markov Model, HMM) 作为序列模型, 已经成为时间序列数据建模的一种主流模型。其已在多个领域得到广泛应用, 例如语音识别, 文本分类, 生物信息等。HMM 模型是一种可以用于标注问题的统计学习模型, 其描述了一种由隐藏的马尔可夫链来生成观测序列的过程, 是一种生成模型。这一节我们主要介绍 HMM 的基本理论, 包括模型定义和算法。

1.1. 隐马尔可夫模型的定义

隐马尔可夫模型主要包括两个序列, 一个是状态序列 (state sequence), 一个是观测序列 (observation sequence)。其描述由一个隐藏的马尔可夫链来随机生成不可观测的状态序列, 再由各个状态来生成观测进而得到观测序列的过程。序列每一个位置代表了一个时刻。

隐马尔可夫模型主要由三个参数来确定, 一个是初始概率分布, 一个是状态转移概率分布, 一个是观测概率分布。我们对整个模型其进行形式化如下:

我们设 Q 为所有隐藏状态的集合, V 为所有可能的观测的集合: $Q = \{q_1, q_2, \dots, q_N\}$, $V = \{v_1, v_2, \dots, v_M\}$ 。这里 N 为所有可能的隐藏状态的数量, M 是所有可能的观测数。我们设 I 为状态序列, 其长度为 T , O 为其观测序列: $I = (i_1, i_2, \dots, i_T)$, $O = (o_1, o_2, \dots, o_T)$ 。我们设 A 为隐藏状态的转移矩阵: $A = [a_{ij}]_{N \times N}$, 这里, $a_{ij} = P(i_{t+1} = q_j | i_t = q_i)$, $i = 1, 2, \dots, N; j = 1, 2, \dots, N$ 。其表示在时刻 t 时候处于隐藏状态 q_i 的情况下在时刻 $t+1$ 转移至隐藏状态 q_j 的概率。我们设 B 为观测矩阵, 也叫发射概率矩阵: $B = [b_j(k)]_{N \times M}$, 这里, $b_j(k) = P(o_t = v_k | i_t = q_j)$, $k = 1, 2, \dots, M; j = 1, 2, \dots, N$ 。其表示在时刻 t 的时候隐藏状态为 q_j 的时候生成观测值 v_k 的概率。我们设 π 为初始状态: $\pi = (\pi_1)$, 这里, $\pi_i = P(i_1 = q_i)$, $i = 1, 2, \dots, N$ 。其表示在时刻 t 为 1 的情况下隐藏状态为 q_i 的概率。

由以上我们可以得知隐马尔可夫模型主要由三个参数即初始状态 π , 状态转移矩阵 A 以及观测矩阵 B 来决定。我们设隐马尔可夫模型为 λ , 其可以用上述三个参数来表示: $\lambda = (A, B, \pi)$ 。这里隐藏状态由两个参数确定, A, π ; 观测矩阵 B 确定如何从隐藏状态来生成观测, 整个模型的示意图如图 1 所示。

我们可以发现隐马尔可夫模型其有三个重要假设: (1) 当前观测值只由当前隐藏状态确定, 而其他隐藏状态或观测值无关: $P(o_t | i_T, o_T, i_{T-1}, o_{T-1}, \dots, i_{t+1}, o_{t+1}, i_t, i_{t-1}, o_{t-1}, \dots, i_1, o_1) = P(o_t | i_t)$ 。(观测的独立性假设); (2) 当前隐藏状态由其前一个隐藏状态决定, 与其它时刻的状态和观测无关: $P(i_t | i_{t-1}, o_{t-1}, \dots, i_1, o_1) = P(i_t | i_{t-1})$, $t = 1, 2, \dots, T$ 。(一阶马尔可夫假设); (3) 隐藏状态之间的转换函数概率不随时间变化 (转换函数稳定性假设)。

隐马尔可夫模型可以用于标注问题, 标注问题就是在给定观测的序列的情况下来预测对应的标注序列, 若假设数据是由隐马尔可夫模型生成的, 则隐藏状态序列即标注序列。由于隐马尔可夫模型是一个生成模型, 也可来描述观测序列的生成过程。

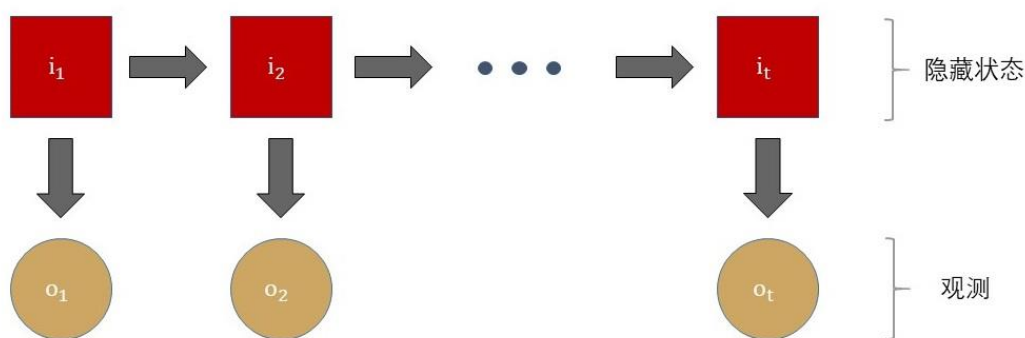
1.2. 隐马尔可夫模型的 3 个基本问题

请仔细阅读在本报告尾部的重要法律声明

对于隐马尔可夫模型来讲，其研究的有三个基本问题：

1. 概率计算问题：当给定模型 $\lambda = (A, B, \pi)$ 和观测序列 $O = (o_1, o_2, \dots, o_T)$ ，计算在模型 λ 已知的情况下观测序列 O 出现的概率 $P(O | \lambda)$ 。
2. 学习问题：在已知观测序列 $O = (o_1, o_2, \dots, o_T)$ 的情况下来估计模型 $\lambda = (A, B, \pi)$ 的参数，从而使得在该模型下观测序列概率 $P(O | \lambda)$ 最大，也就是使用极大似然估计的方法来进行参数估计。
3. 预测问题：即解码问题，是在已知模型 $\lambda = (A, B, \pi)$ 和观测序列 $O = (o_1, o_2, \dots, o_T)$ ，来求得对于给定的观测序列条件概率 $P(I | O)$ 最大的隐藏状态的序列 $I = (i_1, i_2, \dots, i_T)$ ，也就是给定观测序列，求得对应的隐藏状态序列。

图 1 隐马尔可夫示意图



资料来源：华西证券研究所

对于资产择时来讲，我们主要关注后两个问题，也就是学习问题和预测问题，即通过观测序列来估计模型参数并利用模型来进行预测。我们这里简述对于后两个问题的相应算法。

1.3. 学习算法

对于上一节我们所述的学习问题，其需要通过算法来求解。那么根据训练数据是包括观测序列和对应的隐藏状态序列还是只有观测序列，对应 HMM 模型训练算法可以分为两种，一种是监督学习算法，一种是无监督学习算法，也就是 Baum-Welch 算法 (EM 算法)。

1.3.1. 监督学习算法

如果我们假设训练数据中包括 S 个长度相同的观测序列和对应的隐藏状态序列，那么可以通过极大似然估计的方法来估计模型参数，其方法如下所述：

1. 对转移概率 a_{ij} 进行估计：设样本中时刻 t 处于隐藏状态 i 接着在时刻 $t+1$ 转移到状态 j 的频数为 A_{ij} ，则对于转移概率 a_{ij} 的估计为： $\hat{a}_{ij} = \frac{A_{ij}}{\sum_{j=1}^N A_{ij}}$ ， $i = 1, 2, \dots, N; j = 1, 2, \dots, N$

请仔细阅读在本报告尾部的重要法律声明

2. 对观测概率 $b_j(k)$ 进行估计：设样本中隐藏状态为 j 并且观测为 k 的频数为 B_{jk} ，则隐藏状态为 j 观测为 k 的概率 $b_j(k)$ 的估计为： $\hat{b}_j(k) = \frac{B_{jk}}{\sum_{k=1}^M B_{jk}}$ ， $j = 1, 2, \dots, N; k = 1, 2, \dots, M$
3. 对初始状态 π_i 进行估计：即计算 S 个样本中初始状态为 q_i 的频率。

以上为有监督情况下的学习算法，但由于隐藏状态需要人工来进行标注，所以这种算法的局限性比较大。更为实用的为无监督学习算法。

1.3.2. 无监督学习算法

无监督学习算法也就是 Baum-Welch 算法，其假设训练数据值包含 S 和长度为 T 的观测序列，而没有对应的隐藏状态的序列，那么我们的目标是来学习隐马尔可夫模型的参数 $\lambda = (A, B, \pi)$ 。我们设观测数据为 O ，隐藏状态序列为 I ，那么： $P(O | \lambda) = \sum_I P(O | I, \lambda)P(I | \lambda)$ ，其参数通过 EM 算法进行学习，其过程如下：

1. 首先要确认完全数据的对数似然函数：完全数据指的是观测数据和隐藏状态数据的联合： $(O, I) = (o_1, o_2, \dots, o_T, i_1, i_2, \dots, i_T)$ 。那么完全数据的对数似然函数为 $\log P(O, I | \lambda)$ 。
2. 接下来是 EM 算法的 E 步：求得 Q 函数 $Q(\lambda, \bar{\lambda})$ ： $Q(\lambda, \bar{\lambda}) = \sum_I \log P(O, I | \lambda)P(O, I | \bar{\lambda})$ 。这里 $\bar{\lambda}$ 表示隐马尔可夫模型的当前估计值， λ 为要极大化的隐马尔可夫模型参数， $P(O, I | \lambda) = \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) \dots a_{i_{T-1} i_T} b_{i_T}(o_T)$ 。则函数 $Q(\lambda, \bar{\lambda})$ 可以进行展开如下： $Q(\lambda, \bar{\lambda}) = \sum_I \log \pi_{i_1} P(O, I | \bar{\lambda}) + \sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}} \right) P(O, I | \bar{\lambda}) + \sum_I \left(\sum_{t=1}^T \log b_{i_t}(o_t) \right) P(O, I | \bar{\lambda})$ 。
3. 下面是 EM 算法的 M 步：对 Q 函数 $Q(\lambda, \bar{\lambda})$ 进行极大化来求得模型参数 $\lambda = (A, B, \pi)$ 。由于要极大化的参数是单独出现的，所以对各项分别进行最大化，可以得到： $\pi_i = \frac{P(O, i_1 = i | \bar{\lambda})}{P(O | \bar{\lambda})}$ ， $a_{ij} = \frac{\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j | \bar{\lambda})}{\sum_{t=1}^{T-1} P(O, i_t = i | \bar{\lambda})}$ 以及 $b_j(k) = \frac{\sum_{t=1}^T P(O, i_t = j | \bar{\lambda}) I(o_t = v_k)}{\sum_{t=1}^T P(O, i_t = j | \bar{\lambda})}$

以上是对于学习问题的无监督算法。通过上述算法可以对观测序列进行模型参数估计。

1.4. 预测算法

对于 1.3 节提到的预测问题，也就是解码问题，求得给定观测序列下最有可能的隐藏状态序列。对于这个问题，主要有两种算法，一种是近似算法，一种是维特比算法，下面我们简述这两种算法。

1.4.1. 近似算法

近似算法的本质思想是在每个时刻 t 来选择该时刻最有可能出现的状态 i_t^* ，进而得到一个状态序列 $I^* = (i_1^*, i_2^*, \dots, i_T^*)$ ，此为预测结果。

当给定隐马尔可夫模型 $\lambda = (A, B, \pi)$ 以及观测序列 O ，在时刻 t 处于状态 q_i 的概率 $\gamma_t(i)$ 为： $\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{P(O | \lambda)} = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)}$ 。那么在每一时刻 t 最有可能的状态 i_t^* 为： $i_t^* =$

请仔细阅读在本报告尾部的重要法律声明

$\arg \max_{1 \leq i \leq N} [\gamma_t(i)]$, $t = 1, 2, \dots, T$ 。进而据此得到状态序列 $I^* = (i_1^*, i_2^*, \dots, i_T^*)$ 。近似算法计算非常简单，但其并不能保证是全局最优的，即预测的状态序列很可能是实际上不发生的。

1.4.2. 维特比算法

维特比算法是一种通过动态规划的方法来解决隐马尔可夫模型的预测问题，此时一条路径对应一个最大概率的序列。

首先导入两个变量 δ 和 ψ ，我们定义在时刻 t 状态为 i 的所有单个路径 (i_1, i_2, \dots, i_t) 中概率最大值为： $\delta_t(i) = \max_{h_1, i_2, \dots, i_{t-1}} P(i_t = i, i_{t-1}, \dots, i_1, o_t, \dots, o_1 | \lambda)$, $i = 1, 2, \dots, N$ 。那么 δ 的递推公式为： $\delta_{t+1}(i) = \max_{1 \leq j \leq N} P(i_{t+1} = i, i_t, \dots, i_1, o_{t+1}, \dots, o_1 | \lambda) = \max_{1 \leq j \leq N} [\delta_t(j) a_{ji}] b_i(o_{t+1})$, $i = 1, 2, \dots, N; t = 1, 2, \dots, T-1$ 。那么定义在时刻 t 隐藏状态为 i 的所有单个路径 $(i_1, i_2, \dots, i_{t-1}, i)$ 中概率最大的路径的第 $t-1$ 个结点为： $\psi_t(i) = \arg \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}]$, $i = 1, 2, \dots, N$ 。

我们下面介绍维特比算法：

算法的输入：模型 $\lambda = (A, B, \pi)$ 以及观测序列 $O = (o_1, o_2, \dots, o_T)$ ；

算法的输出：最有的路径 $I^* = (i_1^*, i_2^*, \dots, i_T^*)$ 。

1. 进行初始化： $\delta_1(i) = \pi_i b_i(o_1)$, $i = 1, 2, \dots, N$; $\psi_1(i) = 0$, $i = 1, 2, \dots, N$
2. 进行递推计算，对于 $t=2, 3, \dots, T$: $\delta_{t+1}(i) = \max_{1 \leq j \leq N} [\delta_t(j) a_{ji}] b_i(o_{t+1})$, $i = 1, 2, \dots, N$; $\psi_t(i) = \arg \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}]$, $i = 1, 2, \dots, N$ 。
3. 终止计算： $P^* = \max_{1 \leq i \leq N} \delta_T(i)$; $i_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)]$ 。
4. 最后是进行路径回溯：对于 $t=T-1, T-2, \dots, 1$, $i_t^* = \psi_{t+1}(i_{t+1}^*)$ 。据此求得最优路径 $I^* = (i_1^*, i_2^*, \dots, i_T^*)$ 。

1.5. 小结

隐马尔可夫模型包含两个随机过程，一般随机过程以及马尔可夫过程。这其中马尔可夫过程对状态转移进行刻画，运用转移概率矩阵来进行描述。一般随机过程用概率分布函数表示。隐马尔可夫模型作为一个非线性时间序列模型在许多领域有着较为成熟的应用，其在模式识别以及状态预测等方面已经被证明了有着很大的优势。模型的本质就是通过训练数据来对未来进行预测。我们从证券市场获得的观测数据大多都具有时间趋势，借助隐马尔可夫模型来对进行资产价格走势预测是具有很强的现实意义的。

2. 隐马尔可夫模型的择时建模

我们知道，股价是可以被每个人观察到，但股价背后的意义却并不能直接被观察到，即使是相同的股价通常也会有着不同的意义。例如在 2015 年 6 月和 2018 年 1 月

请仔细阅读在本报告尾部的重要法律声明

上证指数都为 3000 点，但这两个 3000 点有着截然不同的含义，2015 年时候的 3000 点是在一段持续性下跌的过程中，而 2018 年的 3000 点确实指数触底后的一段上涨过程中，两者有着完全相反的内在含义。而股价背后的状态和含义是需要通过可观测的状态来进行估计的，隐马尔科夫模型对这样的包含因变量的时序过程有着很好的建模，我们这里来研究基于隐马尔科夫模型的择时建模。

2.1. 模型合理性的讨论

从理论层面上来讲，利用隐马尔科夫模型进行建模需要满足其三个核心假设：

(1) 当前观测值只由当前隐藏状态确定，而与其他隐藏状态或观测值无关(观测的独立性假设)；(2) 当前隐藏状态由其前一个隐藏状态决定，与其它时刻的状态和观测无关：(一阶马尔科夫假设)；(3) 隐藏状态之间的转换函数概率不随时间变化(转换函数稳定性假设)。

首先我们讨论观测的独立性假设，其认为每个观测变量只由其隐藏状态来决定，核心思想就是不同状态对应的观测数据来源于不同的概率分布，这个概率分布值依赖于其所属隐藏状态是什么而和其它状态无关。这个假设是一个较强的假设，但对于资产价格数据来讲是可以较为合理的，因为牛市和熊市下的各种指标有着显著不同，通过构建合适的输入观测数据使得牛熊等不同市场情况下有着比较明显的区分将会较好的满足这个假设。

接着我们讨论一阶马尔科夫假设，这个假设认为当前隐藏状态仅有前一个隐藏状态决定，而与其它隐藏状态无关。其仅考虑了一阶的关联性，而没有考虑高阶或者长期关联性，较为简单的模型会有着比较强的泛化能力，但也会带来模型的拟合能力不足，我们这里先以一阶马尔科夫假设进行研究，通过对输入数据的调整，例如选取合适的持仓周期来使得输入数据尽可能满足此假设等方法来提高模型的表现。

最后我们讨论转换函数稳定性假设，其认为隐藏状态之间的转换函数不随时间变化，这意味着对于训练数据的时间长度内隐藏状态的转移遵循相同的规律。那么这个假设将和训练样本的时间长度和调仓时间息息相关。

后两个假设是关于隐藏状态转移是否符合马尔科夫链。我们对于训练样本进行检验以判断隐马尔科夫模型是否适合择时建模。

2.2. 模型优缺点的讨论

隐马尔科夫模型有着较高的可解释性，其通过对隐变量进行建模，将复杂的观测用隐藏状态来进行解释，这有着一定的可解释性。其可以理解为将时间序列进行降维，低维度的隐藏状态则更具有可解释性，并且利用低维变量进行预测会有着更高的泛化能力和预测准确度。另外对于推断出的隐藏状态可以和经验进行比对，判断每个隐藏状态的含义。

隐马尔科夫模型核心调节参数比较方便，主要需要调整的超参数为隐藏状态的个数，相比于其它模型来讲其超参数数量较少，隐藏状态比较少时候会出现欠拟合的效应，隐藏状态比较多时候会出现过拟合，通过调节这一个超参数即可来调整模型，比较容易和直观。

隐马尔科夫模型有着比较高的模型表示能力，其作为一个非线性模型，相比于传统线性模型来讲会有着更高的模型表示能力，可以对更复杂的输入进行表示，并且作为一个概率模型可以刻画不确定性。

请仔细阅读在本报告尾部的重要法律声明

但如我们在 2.1 节中所述，如果期望模型有着比较好的表现，其需要满足模型的基本假设，如何对数据进行调整以满足模型假设是一个需要细致探索和研究的内容。另外模型的学习算法即 EM 算法并不能保证会收敛至全局最优，这也需要进行额外的考虑来制定合适的模型。

2.3. 策略设计思路

我们的策略的主要设计思路如下：

1. 设置主要的模型构建方法，即进行滚动训练以及设置执行方案；
2. 根据隐马尔可夫模型的原理来设置预测方案；
3. 对数据及参数进行选取；
4. 通过回测来确定最优模型。

2.3.1. 建模方法

从理论上讲，隐马尔可夫模型一般假设可观测序列服从泊松分布或者正态分布，其中泊松分布用来描述离散型随机变量，正态分布用来描述连续型随机变量，股票的时间序列为连续的，这里我们选取的模型假设可观测序列为正态分布，即正态隐马尔可夫模型。

由于资产价格数据随着时间的推移在不断增长，所以这里采用滚动训练的方法，即当获取到新的数据时候就把其加入训练数据中。由于隐马尔可夫模型的一阶马尔可夫假设，我们这里提出两种滚动训练的方法：一种是从起始数据开始，新的数据不断加入训练数据中，即训练样本长度不断增加，即不定长滚动训练；另一种是设定固定的训练样本长度，新的样本的加入则同时将样本周期最前的数据剔除。我们这里将会对两种方案分别进行测试。

模型的执行步骤为：

对模型进行训练，做参数估计。选取模型的训练集，如上所述，一种是不定长滚动训练集，一种是定长滚动训练集，将其特征指标作为输入变量，使用 Baum-Welch 算法来对模型进行训练，得到估计的参数。

将训练集观测向量数据输入训练好的模型使用 Viterbi 算法进行样本内解码，获得最大概率的隐藏状态序列，根据隐藏状态来对所处市场状态进行标注，再根据解码的状态来对未来状态进行预测，根据预测结果进行交易操作。

2.3.2. 预测方法

做预测是基于未来数据和历史数据存在类似的规律假设。基于隐马尔可夫模型的基本原理，可以构造预测的基本方法，即首先根据样本来估计模型参数，之后计算观测样本的在该参数模型下隐藏序列，假定当前为第 i 个调仓期间，欲估计第 $i+1$ 个持仓期间的市场状态，也就是统计第 i 个持仓期间所属隐藏状态在训练样本中后一个状态的期望收益，其作为 $i+1$ 持仓期间的预测，根据此预测收益率来判断是否进行交易。

2.3.3. 数据及参数选择

关于预测标的的选取我们使用沪深 300 指数（000300）的日数据来进行测试。

请仔细阅读在本报告尾部的重要法律声明

隐马尔科夫模型的参数主要包括两部分，一个是隐藏状态数量的选择，即模型的超参数；一个是观测变量的选择。

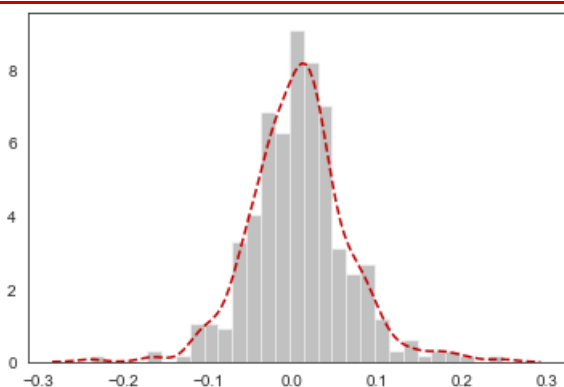
隐藏状态数量参数选择：从证券市场的周期来看可以粗略分为牛市，熊市和震荡市。而从价格来看又大致可以分为上涨启动阶段，上涨加速阶段，上涨末端，下跌初始阶段，下跌中继阶段，下跌末端。由于隐藏状态数量的设置过少会使得模型拟合能力下降，而过多又会欠拟合，我们将以隐藏状态为 6 为中心，在其附近进行参数寻优。

观测变量选择：一般而言隐马尔科夫模型的观测变量数量通常为 4，变量数量过少会使得市场信息不能被有效包含，变量数量过多会使得模型复杂，噪音变大，观测变量应该可以反应价格因素的变化，成交量变化以及收益率的波动情况。由于我们主要以收益率来作为交易的依据，我们这里以收益率作为基础观测变量，关于最佳观测变量的选择可以通过不断增加观测变量的数量来判断最佳观测变量的选择。对于更严谨观测变量选择我们这里提出枚举的方法，即首先构造观测变量因子池，构造所有的观测变量因子组合，对这些不同的组合在初始训练样本进行模型参数学习，选取一定长度的样本作为验证样本，在其上进行验证来判断最优因子组合。这里我们限于篇幅仅对 4 个观测变量因子进行研究。

2.4. 策略具体过程及回测分析

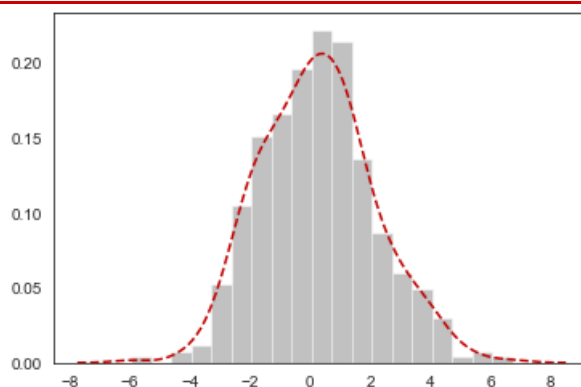
我们这里以沪深 300 作为交易标的，首先获取观测值的时间序列，这是后续进行模型计算的基础。我们首先将观测变量分为价格类，成交量类和综合类。对于收益率类观测变量，我们选取持仓时间长度收益率，交易日前二分之一持仓时间长度收益率作为价格类观测变量；持仓期间夏普比率作为综合类观测变量；5 日和持仓期平均成交量之比作为成交量类观测变量。如下图所示，我们以持仓时间长度为 10 日为例，画出数据的频率分布直方图：

图 2 持仓时间平均收益率



资料来源：华西证券研究所

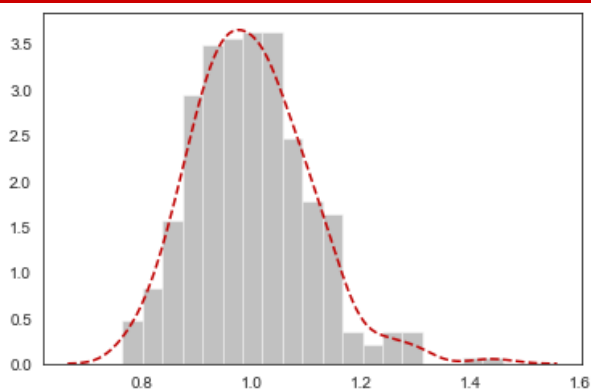
图 3 5 日平均收益率



资料来源：华西证券研究所

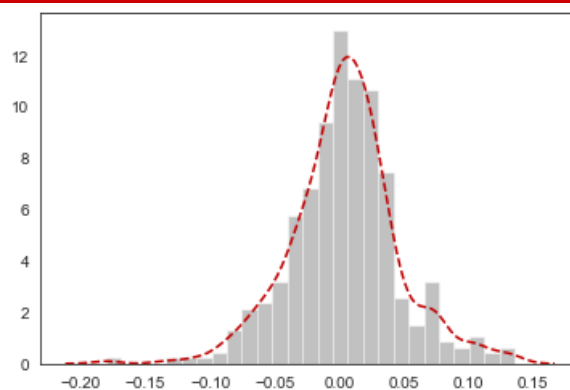
请仔细阅读在本报告尾部的重要法律声明

图 4 5 日和持仓期的平均成交量之比



资料来源：华西证券研究所

图 5 持仓时间长度内夏普比率



资料来源：华西证券研究所

由于我们这里使用正态隐马尔可夫模型，即假设观测变量服从正态分布，观察频率直方图可判断数据基本符合我们的假设，如果输入不符合关于正态分布的输入假设则还需进行输入变换，我们以此作为模型的输入。

这里数据的维度为 4 维，将 4 维的历史时间序列数据作为观测变量输入到模型中，通过 EM 算法得出模型的参数 $\lambda = (A, B, \pi)$ 。由于 EM 算法并不能保证找到全局最优解，所以我们进行对模型进行多次迭代，选取对数据拟合效果最好的模型参数。之后固定模型参数和观测变量，利用维特比算法进行更新，以求得最大概率的隐藏状态序列。隐藏状态的最后一项即为对于交易日前持仓期的状态推断。由于隐马尔可夫模型的状态转换函数不变性假设，我们对推断出的隐藏状态和转移函数进行相乘，求得下一个持仓阶段收益率的期望，此为未来收益率的预测，根据未来收益率预测来判断在下个持仓周期内交易操作。当新的观测数据到来时候，重新进行以上步骤，进而实现滚动训练预测。

2.4.1. 隐藏状态的马尔可夫性检验

检验随机变量序列是否具有“马氏性”，是应用马尔科夫链模型分析和解决实际问题的必要前提。隐马尔科夫模型假设隐藏状态的转移符合“马氏性”，这里我们首先对这个假设进行检验。目前绝大多数方法应用隐马尔科夫模型方法去解决实际问题时，忽视了检验“马氏性”这一步骤，这是不科学的，也是不严谨的。我们这里给出马氏性检验定理。

定理 1：设所讨论的指标值序列包含 m 个可能的状态，用 f_{ij} 表示指标值序列中从状态 i 经过一步转移到达状态 j 的频率。将转移频数矩阵的第 j 列之和除以隔行割裂的总和所得值称之为边缘概率，记为 $p_{\cdot j}$ ，即 $p_{\cdot j} = \sum_{i=1}^m f_{ij} / \sum_{i=1}^m \sum_{j=1}^m f_{ij}$ 。则统计量 $\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \log \frac{p_{ij}}{p_{\cdot j}} \right|$ 以自由度为 $(m-1)^2$ 的 χ^2 分布为极限分布。其中 $p_{ij} = f_{ij} / \sum_{j=1}^m f_{ij}$ 。给定显著性水平 α ，若 $\chi^2 > \chi_{\alpha}^2((m-1)^2)$ ，则认为状态序列符合马氏性，否则其不可以被认为马尔科夫链。

基于定理 1，我们对沪深 300 标的进行检验，初始日期为 2005.01，初始 2000 个交易日作为模型的训练样本。我们对训练样本进行学习和解码，选取持仓时间为 10 天，改变隐藏状态的数量，判断解码得到的隐藏状态序列是否可以通过检验。

请仔细阅读在本报告尾部的重要法律声明

表 1：隐藏状态数量变化进行马氏性检验

隐藏变量数量	分位点	计算统计量
3	9.48	109.8
4	16.91	116.66
5	26.29	146.73
6	37.65	173.47
7	50.99	203.87
8	99.33	257.44

资料来源：华西证券研究所

从表 1 中可以发现解码后的隐藏状态都可以在显著性水平为 0.05 的水平上认为隐藏状态序列满足马氏性。接着我们固定隐藏状态数量为 6，对持仓周期进行变化，来判断隐藏状态序列是否满足马氏性。我们改变持仓周期，结果表明不同的持仓周期下隐藏状态序列都具有马氏性，并且可以发现在 10，20 天左右马氏性最高，这可能意味着当持仓周期选取 10 天，20 天左右时候会更好的和模型假设相匹配。

我们这里说明了隐马尔科夫模型对于沪深 300 标的进行建模的合理性，这说明其基本满足模型的假设。但预测的准确度还受很多因素影响，我们在后两节通过回测来说明这些变量和因素的影响。

表 2：持仓周期变化进行马氏性检验

持仓周期	分位点	计算统计量
5	37.65	101.21
10	37.65	173.47
20	37.65	184.01
30	37.65	103.19
40	37.65	85.88

资料来源：华西证券研究所

2.4.2. 不定长滚动训练分析

不定长滚动训练即我们的历史观测数据是不断增加的，每当有新的观测数据时候就把这些数据加入历史数据中，所以观测数据是在不断增长的。我们这里对沪深 300 标的进行回测，回测区间为 2005.01 至 2021.07，初始 2000 个交易日作为模型的初始训练样本。

我们选取持仓周期为 10 天，每 10 天进行是否交易的判断。首先我们对观测数据的维数进行分析，确定观测数据维数对于回测结果的影响。这里隐藏状态数量我们设

请仔细阅读在本报告尾部的重要法律声明

置为 6。我们以持仓时间长度收益率作为基础观测变量，分别添加其它观测变量来判断添加其进入观测变量组合的影响。我们将观测变量组合进行编号：1 号为持仓周期收益率；2 号为持仓周期收益率+持仓期间夏普比率；3 号为持仓周期收益率+持仓期间夏普比率+持仓半周期和持仓全周期平均成交量之比；4 号为持仓周期收益率+持仓期间夏普比率+持仓半周期和持仓全周期平均成交量之比+持仓半周期收益率。

结果如下表所示，可以发现单变量并不能实现好的分类预测效果，这说明更多的观测变量使得对于隐藏状态的分类更加准确。如语音识别领域中提到过高的观测变量维度会带来严重的过拟合现象，我们这里推荐 4 个左右的观测维度可以有着不错表现。

表 3：观测变量组合对回测结果的影响

观测变量组合编号	夏普比率	年化收益率	总收益率	最大回撤	胜率	交易次数
1	0.2	0.0734	0.5858	0.3157	0.44	9
2	0.42	0.1645	1.3116	0.42	0.43	30
3	0.54	0.2238	1.7842	0.3122	0.61	18
4	0.65	0.2788	2.223	0.2761	0.74	27
基准	0.33	0.1345	1.0723	0.467		

资料来源：华西证券研究所

接着我们来对超参数隐藏变量进行分析，我们这里固定持仓周期为 10 天，改变隐藏变量的数量来研究隐藏变量的影响。

如表 4 所示，当隐藏状态比较少时候和隐藏状态数量过多的时候都不能获得良好的回测表现，这是因为机器学习模型的一个固有即过拟合和欠拟合的问题，隐藏状态数量作为一个超参数对于调节模型对于已有数据的拟合效果具有重要影响。我们希望模型具有好的预测效果即需要模型有较强的泛化能力，我们的实验回测经验告诉我们隐藏状态数量为 6 可以取得较好的预测效果。

表 4：隐藏状态数量对回测结果的影响

隐藏变量数量	夏普比率	年化收益率	总收益率	最大回撤	胜率	交易次数
3	0.14	0.0603	0.4813	0.5258	0.47	38
4	0.03	0.0259	0.2071	0.6091	0.48	31
5	0.54	0.2308	1.8402	0.3449	0.61	13
6	0.65	0.2788	2.223	0.2761	0.74	27
7	0.47	0.1881	1.4998	0.3455	0.55	27
8	0.38	0.1466	1.1694	0.3474	0.5	30
基准	0.33	0.1345	1.0723	0.467		

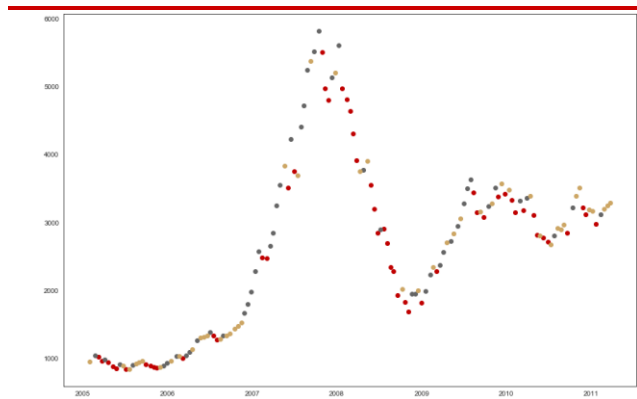
资料来源：华西证券研究所

图 6 和图 7 是我们当设定隐藏状态数量为 3 和 6 时候分别利用维特比算法对观测变量进行解码得到的效果图。对于 2000 个交易日的训练样本，基于在该期间内学习出来的模型参数使用维特比算法计算各个时间点的隐藏状态数值，将计算出来的隐藏状态特征用不同的颜色标记出来。由图 6 可以观察得到时间被分为了 3 种阶段，绿色代表上涨阶段，橘色代表下跌阶段，蓝色代表震荡阶段。由图 8 可以观察得到，不同的市场隐藏状态确实被隐马尔可夫模型解码出来，隐藏状态 0 的蓝色，隐藏状态 1 的

请仔细阅读在本报告尾部的重要法律声明

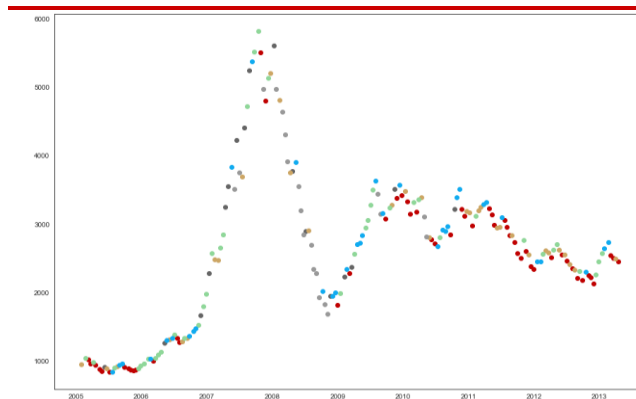
橘色，隐藏状态 2 的绿色和隐藏状态 5 的紫色为上涨阶段，且隐藏状态 0 的红色主要表述上涨的幅度较大阶段；隐藏状态 3 的红色和隐藏状态 5 的棕色对应下跌阶段；所以模型可以较好的对当前持仓周期的交易日内有着比较好的解码效果，其在牛市和熊市以及震荡市可以做到有效标注和区分。

图 6：隐藏状态数量为 3 时候解码效果



资料来源：华西证券研究所

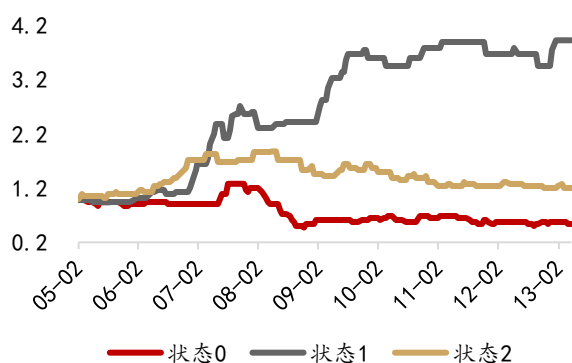
图 7：隐藏状态数量为 6 时候解码效果



资料来源：华西证券研究所

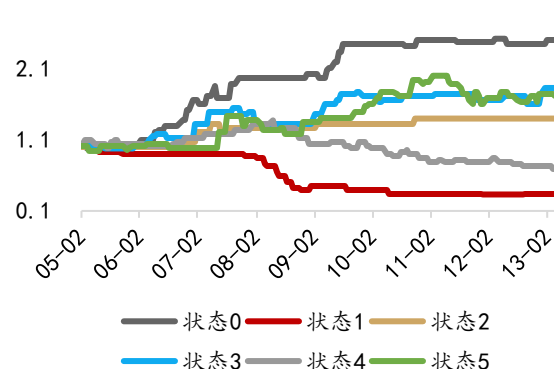
我们分别对隐藏状态为 3 以及隐藏状态为 6 的情况下对训练样本的得到的每个状态采取分时买入的多头策略，观察每种隐藏状态在截止到样本周期内最后一日的策略净值曲线，如图 8 和图 9 所示：当隐藏状态数量为 3 的时候策略净值高于初始净值 1 的隐藏状态为 0 和 2，但 0 状态对应的净值曲线存在较大回撤。而当隐藏状态数量为 6 的时候策略净值高于初始净值 1 的隐藏状态为 0, 1, 2 和 4，且这其中状态的策略净值几乎没有遇到较大回撤。通过比较可以显著发现隐藏状态更多会对预测收益有着更细致的分类，同时其买入策略净值的单调性也更好，这对于提高收益预测的胜率有着较大作用。

图 8：隐藏状态为 3 时的分时买入策略净值图



资料来源：华西证券研究所

图 9：隐藏状态为 6 时候的分时买入策略净值图



资料来源：华西证券研究所

我们还研究了持仓周期对于回测的影响，不同的持仓周期对应不同交易频率的投资者，我们来研究模型是否可以在长短不同的持仓周期下都有着超过基准的表现。

表 5 持仓周期对回测结果的影响

持仓周期	夏普比率	年化收益率	总收益率	最大回撤	胜率	交易次数
10	0.65	0.2788	2.223	0.2761	0.74	27
基准	0.33	0.1345	1.0723	0.467		
20	0.41	0.1505	1.1946	0.3911	0.61	21
基准	0.33	0.1281	1.0169	0.467		
30	0.02	0.019	0.152	0.4218	0.55	18
基准	0.33	0.1345	1.0723	0.467		
40	0.41	0.1503	1.1804	0.3125	0.85	7
基准	0.31	0.1219	0.9575	0.467		
50	0.52	0.1971	1.5402	0.3246	0.6	5
基准	0.37	0.152	1.1946	0.467		
60	0.41	0.1548	1.2163	0.47	0.625	8
基准	0.31	0.1211	0.9575	0.467		

资料来源：华西证券研究所

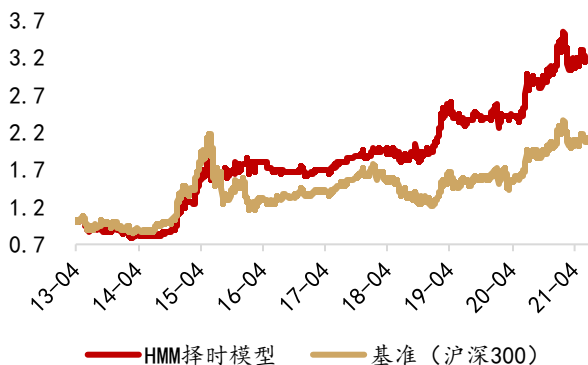
如表 5 所示，由于不同的持仓周期选取会导致初始样本日期不一致，所以每一种持仓周期对应不同的基准。我们可以发现从较短持仓周期的 10 天到较长持仓周期的 60 天即季度持仓周期都有着超过基准的表现。而例外的是持仓周期为 30 天的测试结果，其显著低于基准，尽管其胜率依然高于 50% 的水准，这由于隐马尔可夫模型的一个重要参数是初始参数的选择，由于不同的持仓周期会使得初始观测变量不一致从而推断得到的模型参数也不一致。但如果仅从胜率的角度来看，隐马尔可夫模型依然可以做到对涨跌的预判。

2.4.3. 定长滚动训练分析

请仔细阅读在本报告尾部的重要法律声明

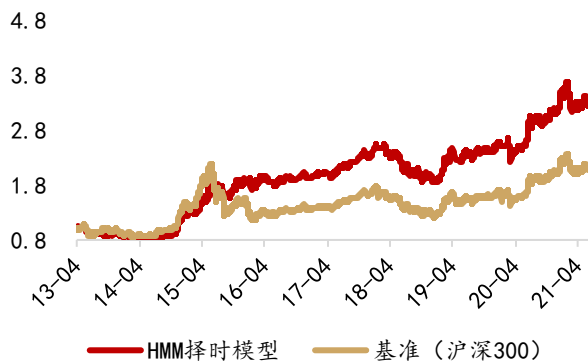
定长滚动训练即我们的历史观测数据是固定长度的，每当有新的观测数据时候就把新的数据加入到历史数据中同时将历史数据中较早的部分剔除，从而使得历史观测数据始终保持一个固定的长度。和上文一样，我们这里对沪深 300 标的进行回测，回测区间为 2005.01 至 2021.07，训练样本的时间长度选取 2000 个交易日。

图 10：定长滚动训练回测



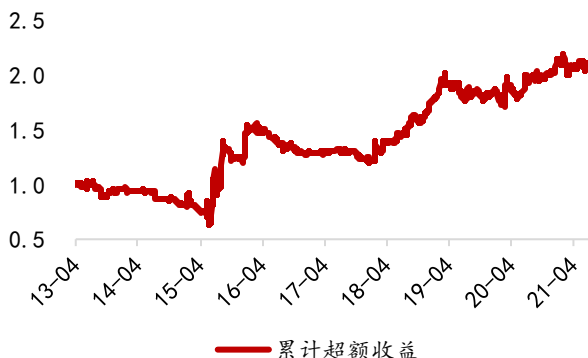
资料来源：华西证券研究所

图 11：不定长滚动训练回测



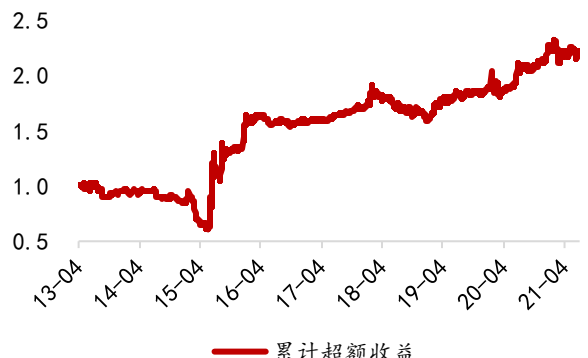
资料来源：华西证券研究所

图 12：定长滚动训练回测累积超额收益率



资料来源：华西证券研究所

图 13：不定长滚动训练回测累积超额收益率



资料来源：华西证券研究所

我们比较了持仓周期为 10 天，隐藏状态数量为 6 的定长滚动训练结果和不定长滚动训练结果，从结果来看不定长滚动训练会比定长滚动训练稍好，尤其是在胜率的表现上，这是因为前者的交易次数较少，其累积了更长时间尺度的历史信息，从机器学习的角度来看增大了样本数量从而减少过拟合。但定长滚动训练有着较小的最大回撤，因为其仅选取交易日前一段时间数据进行训练，本质上是增大了交易日较近历史数据的权重，降低了更远历史数据的权重，进而在某些状态上有着更好的识别。从超额收益率的角度来看定长训练法会获得更高的超额收益率，但从图 10 和 11 中明显发现不定长训练带来的累积超额收益率存在更大回撤相比于定长训练方法。这进一步印证了定长训练会降低较长历史数据的干扰。

请仔细阅读在本报告尾部的重要法律声明

表 6：定长滚动训练和不定长滚动训练的比较

	夏普比率	年化收益率	总收益率	最大回撤	胜率	交易次数
定长滚动训练回测	0.61	0.2403	1.9164	0.2449	0.65	47
不定长滚动训练回测	0.65	0.2788	2.223	0.2761	0.74	27
基准	0.33	0.1345	1.0723	0.467		

资料来源：华西证券研究所

2.5. 择时策略讨论

基于隐马尔可夫模型的择时策略是一种机器学习模型测量，本质上是将高维复杂的观测变量进行降维，在更低维度的隐藏状态上对隐藏状态之间的转移过程进行拟合，根据拟合出的转移概率来对未来的隐藏状态进行预测。

模型主要包括两步骤，第一是通过 EM 算法求得模型的参数，第二是通过维特比算法来对隐藏状态进行解码。之后根据隐藏状态和转移概率来对未来状态预测。作为机器学习模型其对于数据的要求比较高，由于我们这里使用了正态隐马尔科夫模型，其要求观测变量符合正态分布的假设，所以对数据需要进行合适的处理才能输入到模型。数据还需要满足隐马尔科夫模型的三大假设，如果不满足其三大假设会导致拟合预测效果变差，如我们在表 3 的持仓周期 30 天例子所示。如何判断数据是否满足假设并没有一个直接的统计检验手段，根据机器学习中的经验，我们推荐将数据分为训练集和测试集，首先在训练集上进行模型的参数和观测变量的调优选择，之后固定调优参数在测试集上进行回测。

3. 基于隐马尔可夫择时的资产配置策略

3.1. 资产配置

资产配置的研究对于金融经济行业来讲有着非常重要的意义。等权法（1/N）就是一种很古老的资产配置方法。等权法不需要对任何参数进行预测，而只是简单地对资产进行平均的分配。尽管其操作极为简单，但其却非常有效，并且在如今的资产的配置中依然有着极高的活力。假设一个资产组合 P，由 N 个子资产构成，资产组合的权重向量为 w ，对于等权配置模型来讲，模型权重为 $w_i = w_j = 1/N$ ，即所有子资产按照相同市值权重配置。

而现代的资产配置的理論起源于 Markowitz 在 1952 年提出的均值方差模型（Mean-Variance Optimization, MVO）。这种方法通过在组合预期收益率给定的情况下选择波动率最小的一种资产组合。MVO 是一种经典的资产配置方法，许多新的资产配置模型都是建立在其思想上的。它的数学表达式如下：

请仔细阅读在本报告尾部的重要法律声明

$$\begin{cases} \min \sigma_p^2 = \frac{1}{2} \mathbf{w}^T \Sigma \mathbf{w} \\ \max r_p = \mathbf{w}^T \mathbf{r} \\ s.t. \mathbf{w}^T \mathbf{1} = 1 \end{cases}$$

均值方差模型同时优化风险和收益，及最大化组合收益的同时希望最小化组合风险。这里 σ_p^2 表示整个资产组合的风险， r_p 表示资产组合的收益， r 表示子资产的收益率。其通过优化来对配置权重进行求解。

风险平价法 (Risk Parity, RP) 是由桥水基金创始人达里奥首次提出，其初衷是构建一个投资组合使得在不同的经济环境中都有文件的表现。简单来讲就是对风险进行配平，其不考虑对收益率进行优化，而是只对波动率进行优化，也就是只对风险进行优化。其可以被形式化为如下：

$$\begin{cases} \min_{\mathbf{w}} \frac{1}{2} \sum_i \left[w_i - \frac{\sigma_p^2}{N(\Sigma \mathbf{w})_i} \right]^2 \\ s.t. \mathbf{w}^T \mathbf{1} = 1, \sigma_p = \text{constant} \end{cases}$$

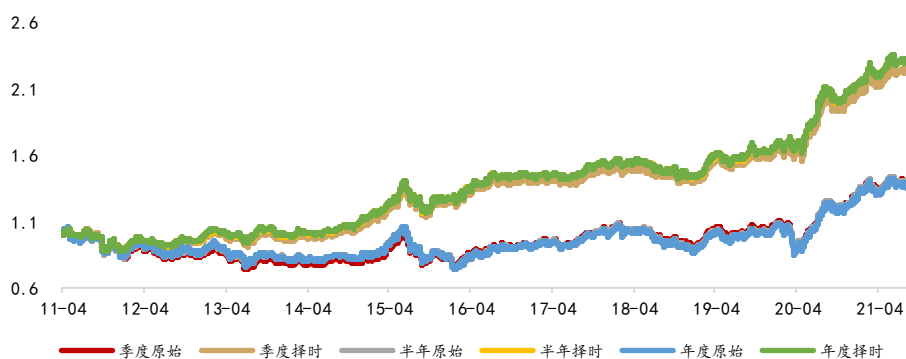
这里 σ_p^2 表示整个资产组合的波动率，基于波动率的风险平价模型只需要输入子资产的协方差，在组合波动率的约束下就可以得到资产组合的权重。

3.2. 基于隐马尔可夫择时的资产配置

如前文所述，隐马尔可夫模型可以通过对未来资产价格变动进行预测来完成较好的择时表现，我们进一步研究在其基础上的资产配置模型。这里我们选取了 8 中资产标的：沪深 300 指数，中证 500 指数，中证 1000 指数，中债企业债财富指数，COMEX 黄金，COMEX 白银，COMEX 铜，布伦特原油。囊括了股票，债券，大宗商品，黄金等资产。我们分别使用等权配置，均值方差模型配置，风险平价模型配置三种方法，来研究隐马尔可夫模型择时的资产配置方法。我们选取的回测区间是 2011.04-2021.08。

首先是等权配置下的策略表现，如表 7 所示，可以发现在等权配置下我们的策略相比于基准有着大幅的提高，这是由于隐马尔可夫在单一子资产上进行了择时增强，从而在等权下表现出来了显著增强。

图 14：等权配置模型不同调仓周期累计收益走势图



资料来源：华西证券研究所

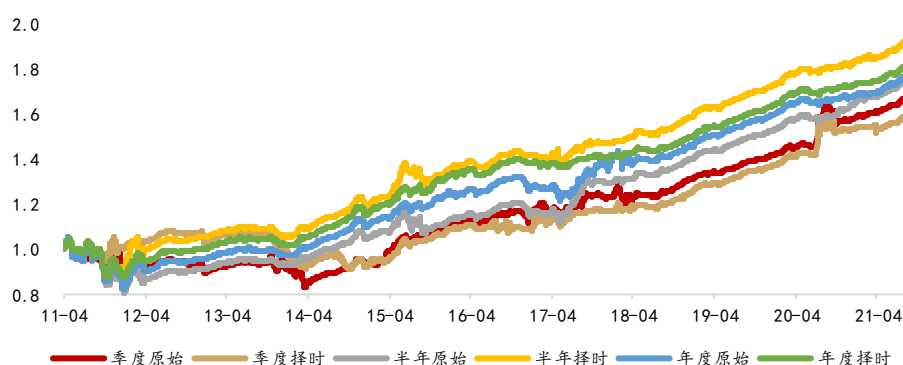
表 7：等权配置下的策略表现

调仓频率	夏普比率	年化收益率	总收益率	最大回撤	年化超额收益率	总超额收益率
季度	0.82	11.13%	112.85%	16.80%	7.70%	84.98%
基准	0.28	3.43%	37.87%	29.86%		
半年	0.85	11.78%	130.09%	17.00%	8.40%	92.70%
基准	0.28	3.38%	37.39%	29.51%		
年度	0.85	11.79%	130.15%	16.73%	8.54%	94.31%
基准	0.27	3.24%	35.83%	29.77%		

资料来源：华西证券研究所

接着我们研究均值方差模型配置下的表现，由于均值方差需要对协方差进行估计，我们是用日频数据作为输入，较小的样本点会带来估计的不稳定性。可以发现季度调仓频率收益会略低于基准，但从夏普比率角度来看策略依然相对于基准有提升。而从半年调仓频率和年度调仓频率来看策略表现依然强于基准。

图 15：均值方差配置模型不同调仓周期累计收益走势图



资料来源：华西证券研究所

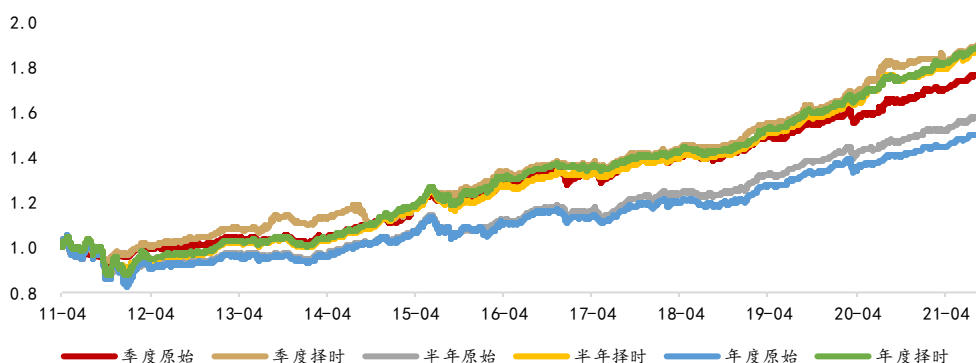
表 8：均值方差模型配置下的策略表现

调仓频率	夏普比率	年化收益率	总收益率	最大回撤	年化超额收益率	总超额收益率
季度	0.72	5.39%	59.56%	16.82%	-0.71%	-8.15%
基准	0.69	6.13%	67.71%	21.09%		
半年	1.06	8.41%	92.91%	17.84%	1.59%	17.62%
基准	0.8	6.82%	75.28%	23.44%		
年度	1.23	7.39%	81.66%	16.63%	0.43%	4.76%
基准	0.94	6.96%	76.89%	21.51%		

资料来源：华西证券研究所

最后我们来看在风险平价模型配置框架下的表现，如表 9 所示我们发现策略表现在所有的评价指标上都大幅高于基准。我们也注意到在风险平价模型配置下包括策略和基准都有着极高的夏普比率和较小最大回撤，这是由于风险平价模型主要是在优化资产组合波动率。也可以发现调仓频率越高夏普比率越高，最大回撤也越小。

图 16: 风险平价配置模型不同调仓周期累计收益走势图



资料来源: 华西证券研究所

表 9: 风险平价模型配置下的策略表现

调仓频率	夏普比率	年化收益率	总收益率	最大回撤	年化超额收益率	总超额收益率
季度	1.49	8.13%	89.79%	10.63%	1.15%	12.78%
基准	1.33	6.97%	77.01%	11.82%		
半年	1.33	7.90%	87.17%	15.73%	2.62%	28.96%
基准	0.86	5.27%	58.22%	18.26%		
年度	1.15	8.06%	88.99%	16.63%	3.48%	38.45%
基准	0.63	4.57%	50.53%	21.51%		

资料来源: 华西证券研究所

3.3. 资产配置策略讨论

我们在三种资产配置模型的框架下研究了隐马尔可夫模型择时对于资产配置策略的改进。不同配置框架下 HMM 模型对策略表现均有一定提升。由于这些资产配置模型也有择时的效果, 可以发现当调仓频率变高时候这些资产配置策略对于原始资产会有-定的提高, 但我们的策略已经通过隐马尔可夫模型进行了择时增强, 使得调仓频率对于我们的策略表现影响较小。

4. 总结

本篇报告我们提出了一个基于隐马尔可夫模型的滚动择时策略, 其通过观测变量来对隐藏状态进行推断, 之后根据推断结果进行未来预测。我们对关键参数对于模型效果的影响进行了研究, 并阐明对于隐马尔可夫模型应用到资产择时的两种合理滚动训练预测方法。我们选择沪深 300 作为回测标的, 说明了模型可以在其上取得超额收益。我们将基于隐马尔可夫模型的择时策略与资产配置模型进行结合, 提出了基于隐

请仔细阅读在本报告尾部的重要法律声明

马尔可夫模型的资产配置策略。策略在多种资产配置模型框架下相比于基准都有着明显的提高，可以获得更高的收益和付出更低的风险。

总结来看隐马尔科夫模型具有以下特点：

1. 参数量少，其仅有隐藏状态数量这一个超参数，通过调节隐藏状态这个参数即可以有效调节模型的过拟合和欠拟合现象，直观且简单；
2. 模型具有较好的鲁棒性和稳健性，对多种持仓周期都可以保持高于基准的表现；
3. 模型可以同时具有高收益，高胜率，高夏普比率，低回撤率等优点，这是由于其对隐藏状态有着较好刻画和丰富信息；
4. 对于数据要求较高，我们在研究中也发现模型需要精细的观测变量筛选才可以实现较好的表现。

5. 风险提示

模型基于对历史数据统计，仅作为投资参考。

分析师承诺

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

评级说明

公司评级标准	投资 评级	说明
以报告发布日后的 6 个月内公司股价相对上证指数的涨跌幅为基准。	买入	分析师预测在此期间股价相对强于上证指数达到或超过 15%
	增持	分析师预测在此期间股价相对强于上证指数在 5%—15%之间
	中性	分析师预测在此期间股价相对上证指数在-5%—5%之间
	减持	分析师预测在此期间股价相对弱于上证指数 5%—15%之间
	卖出	分析师预测在此期间股价相对弱于上证指数达到或超过 15%
行业评级标准		
以报告发布日后的 6 个月内行业指数的涨跌幅为基准。	推荐	分析师预测在此期间行业指数相对强于上证指数达到或超过 10%
	中性	分析师预测在此期间行业指数相对上证指数在-10%—10%之间
	回避	分析师预测在此期间行业指数相对弱于上证指数达到或超过 10%

华西证券研究所：

地址：北京市西城区太平桥大街丰汇园 11 号丰汇时代大厦南座 5 层

网址：<http://www.hx168.com.cn/hxzq/hxindex.html>

华西证券免责声明

华西证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。本报告仅供本公司签约客户使用。本公司不会因接收人收到或者经由其他渠道转发收到本报告而直接视其为本公司客户。

本报告基于本公司研究所及其研究人员认为的已经公开的资料或者研究人员的实地调研资料，但本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载资料、意见以及推测仅于本报告发布当日的判断，且这种判断受到研究方法、研究依据等多方面的制约。在不同时期，本公司可发出与本报告所载资料、意见及预测不一致的报告。本公司不保证本报告所含信息始终保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者需自行关注相应更新或修改。

在任何情况下，本报告仅提供给签约客户参考使用，任何信息或所表述的意见绝不构成对任何人的投资建议。市场有风险，投资需谨慎。投资者不应将本报告视为做出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在任何情况下，本报告均未考虑到个别客户的特殊投资目标、财务状况或需求，不能作为客户进行客户买卖、认购证券或者其他金融工具的保证或邀请。在任何情况下，本公司、本公司员工或者其他关联方均不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告而导致的任何可能损失负有任何责任。投资者因使用本公司研究报告做出的任何投资决策均是独立行为，与本公司、本公司员工及其他关联方无关。

本公司建立起信息隔离墙制度、跨墙制度来规范管理跨部门、跨关联机构之间的信息流动。务请投资者注意，在法律许可的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的前提下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。

所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，如需引用、刊发或转载本报告，需注明出处为华西证券研究所，且不得对本报告进行任何有悖原意的引用、删节和修改。

请仔细阅读在本报告尾部的重要法律声明