

隐马尔可夫模型商品期货应用初探

专题摘要：

本报告主要介绍了隐马尔可夫模型的基本概念、模型及相关问题。并简单选取了一些技术分析指标纳入模型的可观察向量中，根据模型给出的结果对螺纹钢主力合约进行价格预测。

隐马尔可夫模型（HMM）被用来描述一个含有隐含未知参数的马尔可夫过程，它的状态不能直接观察到，但能通过可观察向量序列观察到。

将适当的技术指标纳入隐马尔可夫模型可观察向量组来预测螺纹钢期货的后期走势能取得一定的效果。但由于隐马尔可夫模型每次运行所得的隐状态分类较为随机，上升隐状态和下降隐状态是根据收益曲线人工区分而来的，模型的稳定性还有待提升。

投资咨询业务资格：
证监许可【2012】669号

量化组：

研究员：
刘宾
0755-83212741
liubin@citicsf.com
从业资格号 F0231268
投资咨询号 Z0000038

王建伟
021-60812992
wangjianwei@citicsf.com
从业资格号 F3014595
投资咨询号 Z0013229

联系人：

王炳瑜
021-60812989
wangbingyu@citicsf.com
从业资格号 F3018918

邹天舒
021-60812993
zoutianshu@citicsf.com
从业资格号 F3027249

陈舜尧
0755-82723054
chenshun Yao@citicsf.com
从业资格号 F3029712

肖璋瑜
0755-82723054
xiaozhangyu@citicsf.com
从业资格号 F3034888

目录

专题摘要:	1
一、隐马尔可夫模型介绍	3
1.1 马尔可夫过程	3
1.2 隐马尔可夫模型	3
1.3 隐马尔可夫模型的三个典型问题	5
二、HMM 模型商品期货市场应用	6
2.1 技术指标以及时间段选取	6
2.2 样本内数据分析	8
2.3 样本外数据分析	9
2.4 结论与后期优化方向	10
免责声明	11

图目录

图 1: 隐马尔可夫模型状态转移图	4
图 2: 五个技术指标分布图	7
图 3: 处理后的五个技术指标分布图	8
图 4: 样本内隐状态收益曲线	8
图 5: 样本内隐状态数据标注	9
图 6: 样本外隐状态数据标注	9
图 7: 样本外策略收益曲线	10

一、隐马尔可夫模型介绍

1.1 马尔可夫过程

马尔可夫过程（Markov Process），安德烈·马尔可夫于 1906 年最早定义了这样的一个随机过程，并模拟生成了一条状态序列，该状态序列在任一时刻的预期状态都只与上一时刻的状态信息有关，这一特征被称为马尔可夫性质。在该过程中，每个状态的转移只依赖于在此之前的 n 个状态，这个过程被称为 1 个 n 阶的模型，其中 n 是影响转移状态的数目。一阶马尔可夫过程是最简单的马尔可夫过程，任何时刻的状态的转移只依赖于其前一时刻的那一个状态。值得注意的是，这一过程与确定性过程有明显差异，状态之间的转移不是确定性的，而是有一定概率的。

马尔可夫链是随机变量 $X_1, X_2, X_3, \dots, X_n$ 的一个数列。这些变量所有可能的取值被称为“状态空间”，而 X_n 的值表示的是在时间节点 n 的取值状态。如果时间阶段 $n+1$ 的取值状态 X_{n+1} 对于过去状态的条件概率分布仅仅是与 X_n 相关的一个函数，则：

$$P(X_{n+1} = x | X_0, \dots, X_n) = P(X_{n+1} = x | X_n)$$

这里 x 为过程中的某个状态。上面这个恒等式可以被看作是马尔可夫性质。例如，假设某地只存晴天、阴天和下雨天三种天气，某一天该地的天气状况的概率只与前一天的天气状况有关，而与之之前过去所有日子的天气状况无关，那么这个天气变化的随机过程就是一个马尔可夫过程。

在上面的例子中，定义了以下三个部分：状态：晴天、阴天和下雨天；初始向量为：定义系统在时间为 0 的时候的状态的概率；状态转移矩阵：不同天气之间转换的概率都能被这样描述的系统就被称为一个马尔可夫过程

但是，在某些特定的情况下，简单的马尔可夫过程并不能描述我们想要发现的模式。例如，假设某个整天在屋子里的盲人并不能观察到今天外面的天气情况，但是屋内植物的状态在某种概率上是和外面的天气情况相关的。在这样的情况下，我们有两个不同的状态序列，一个是可以被观察到的状态序列（屋内植物的状态），还有一个是观察不到的隐藏状态序列（外面的天气情况），我们可以找到一种算法来根据马尔可夫过程和屋内植物的状态来预测外面的天气情况。

在上面的这种情况，可以观察到的状态序列和隐藏的状态序列在一定程度上是概率相关的。于是我们可以将这种情况的过程描述为一个可以被观察到的序列以及和观察序列概率相关的一个隐藏状态序列的马尔可夫过程。这就是接下来要介绍的隐马尔可夫模型。

1.2 隐马尔可夫模型

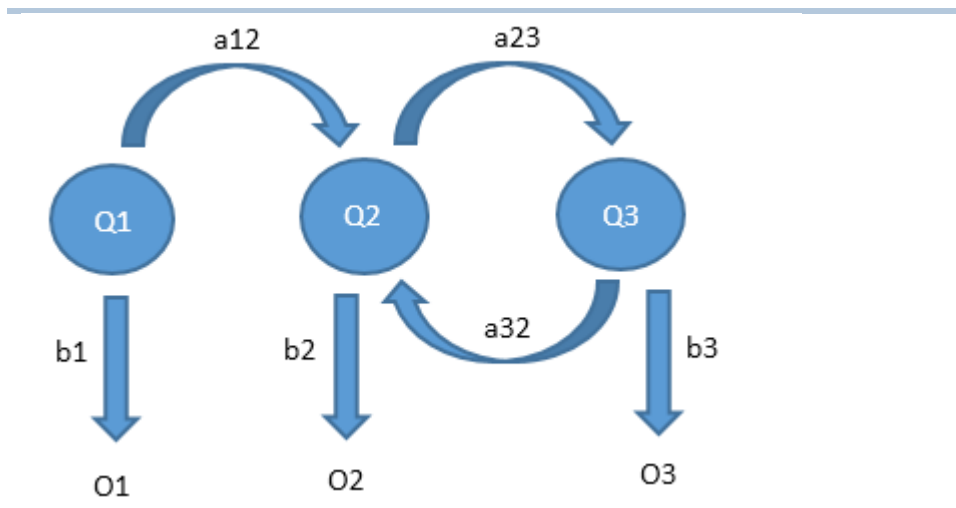
隐马尔可夫模型本质上是基于马尔可夫过程衍生出的一种概率图模型，他被用来描述一个含有隐含未知参数的马尔可夫过程，该模型假设一组可观察序列是

由一些隐藏的状态生成，这些隐状态的转移过程是一个马尔科夫过程，每一个隐状态对应一个或几个可观测到的变量。通常隐状态的转移概率矩阵未知，观测序列是由哪个隐状态序列生成也是未知的，需要通过观测值序列作为隐马尔科夫模型的输入变量来推测。

除了上面植物与天气的例子，还可以用一个比较直观的例子来表述隐马尔可夫模型。假设一个房间里面有三个不同的箱子，编号分别为 1, 2, 3，每个箱子里都装有个数不同的红黄蓝三种颜色的球，有一个盲人每次从其中的一个箱子里摸出一个球，然后将摸出的球丢出房间，房间外面的人会记录球的颜色，再让盲人把球放回原来的箱子。虽然在整个过程中，盲人并不清除他丢出去的球是什么颜色，但他通过箱子的形状可以知道摸的球来自其中哪个箱子，而且他选择箱子的行为是一个离散马尔可夫链。这样在盲人进行多次扔球后，房间外面的人知道每次扔出来的球颜色的序列，还清楚房间里面有几个箱子，但是不了解盲人选择箱子的规律（也就是被隐藏起来的马尔可夫链）和不同箱子里面不同颜色球的分布。这就是一个离散的隐马尔科夫模型。

下图是一个由三个状态组成的隐马尔可夫模型的状态转移图，其中 Q 表示隐含状态，O 表示可观察的序列，a 表示状态转换概率，b 表示输出概率。

图 1：隐马尔可夫模型状态转移图



数据来源：中信期货研究部

隐马尔可夫模型（HMM）可以简单的用五个元素来描述，包括 2 个状态集合和 3 个概率矩阵：

1. 隐含状态 Q

这些状态是马尔可夫模型中隐藏的状态信息，满足马尔可夫性质。通常情况下这些状态无法通过直接观测而得到。（例如 Q_1 、 Q_2 、 Q_3 等等）

2. 可观测状态 O

HMM 模型中与隐藏状态概率相关，通过直接观测可以得到的序列。（例如 O_1 、

O_2 、 O_3 等等，可观测状态的数目不一定要和隐含状态的数目一致。）

3. 初始状态概率矩阵 π

表示隐含状态在初始时刻 $t = 1$ 的概率矩阵，（例如 $t = 1$ 时， $P(S_1) = P_1$ 、 $P(S_2) = P_2$ 、 $P(S_3) = P_3$ ，则初始状态概率矩阵 $\pi = [P_1 \ P_2 \ P_3]$ 。

4. 隐含状态转移概率矩阵A

描述了 HMM 模型中各个隐藏状态之间的转移概率。

其中 $A_{ij} = P(Q_j|Q_i)$, $1 \leq i, j \leq N$ 。表示在 t 时刻、隐藏状态为 Q_i 的条件下，在 $t + 1$ 时刻隐藏状态是 Q_j 的概率。

5. 观测状态转移概率矩阵B

令 N 代表隐藏状态的数目， M 代表可观察状态数目，则：

$B_{ij} = P(O_i|Q_j)$, $1 \leq i \leq M, 1 \leq j \leq N$ 。表示在 t 时刻、隐藏状态是 Q_j 条件下，可观察状态为 O_i 的概率。

对 HMM 来说，有如下三个重要假设：

假设一：马尔可夫假设（各状态之间构成一阶马尔可夫链）

$$P(X_i|X_{i-1}, \dots, X_1) = P(X_i|X_{i-1})$$

假设二：不动性假设（各状态与具体的时间节点无关）

$$P(X_{i+1}|X_i) = P(X_{j+1}|X_j), \forall i, j$$

假设三：输出独立性假设（输出仅与当前状态有关）

$$P(O_1, \dots, O_T|X_1, \dots, X_T) = \prod P(O_t|X_t)$$

1.3 隐马尔可夫模型的三个典型问题

不管在什么领域，如果使用隐马尔可夫模型来进行模式识别，通常会遇到以下三个问题：

- 给定一个 HMM 模型，计算生成给定可观察序列的概率——评估问题
- 已知生成的可观察序列，如何得到最可能的隐藏状态序列——解码问题
- 已知可观察序列，如何得到一个最优的模型参数——学习问题

这就是隐马尔可夫模型的三个典型问题。

1.3.1 评估问题

假设我们已知一个 HMM 模型的各种该参数以及一组可观察序列。我们也许想知道在所有可能的隐藏状态序列下，给定的可观察状态序列的概率。当给定如下一个隐藏状态序列：

$$Q = q_1 q_2 \dots q_T$$

那么在给定 HMM 模型和隐藏状态序列的条件下，可观察状态序列的概率为：

$$P(O|Q, \lambda) = b_{q_1}(O_1) b_{q_2}(O_2) \dots b_{q_T}(O_T)$$

而隐藏状态序列在 HMM 条件下的概率为：

$$P(Q|\lambda) = \pi_{q_1} \alpha_{q_1 q_2} \alpha_{q_2 q_3} \dots \alpha_{q_{T-1} q_T}$$

因此，隐藏状态序列和可观察状态序列的联合概率为：

$$P(O, Q|\lambda) = P(O|Q, \lambda)P(Q|\lambda)$$

那么所有可能的隐藏状态序列上，可观察状态序列的概率为：

$$P(O|\lambda) = \sum_{all\ q} P(O|Q, \lambda)P(Q|\lambda)$$

1.3.2 解码问题

在大多数情况下，我们对包含一些不能被观察到的状态的隐藏序列更感问题。解码问题可以简单的描述为：根据可观察序列找到隐藏序列，即给定可观察序列 $O = O_1 O_2 \dots O_T$ 和模型的参数组 $\lambda = (A, B, \pi)$ ，找出在某种程度上最优的隐藏序列，也就是 $Q = q_1 q_2 \dots q_T$ 。

在给定可观察序列 $O = O_1 O_2 \dots O_T$ 和模型的参数组 $\lambda = (A, B, \pi)$ ，时间点 t 的隐藏状态处于 i 的概率为：

$$\gamma_i(t) = P(s_i = i|O, \lambda) = \frac{P(s_i = i, O|\lambda)}{P(O|\lambda)}$$

这个概率用前向后向算法中的前向概率 $\alpha_t(i)$ 以及后向概率 $\beta_t(i)$ 可以表示为：

$$\gamma_i(t) = \frac{\alpha_i(t)\beta_i(t)}{\sum_{t=1}^N \alpha_i(t)\beta_i(t)}$$

因此，最优隐藏状态序列 Q 可表示为：

$$Q = \arg \max_{1 \leq i \leq N} \gamma_i(t), 1 \leq t \leq T + 1, 1 \leq i \leq N$$

1.3.3 学习问题

学习问题是隐马尔可夫模型的核心，在很多实际的情况下，HMM 不能被直接的判断，这就变成了一个学习问题，对于给定的可观察状态序列 O 来说，如何精确地找到一组最优的 HMM 参数 λ 使 $P(O|\lambda)$ 最大。它通过优化模型的参数来最好地解释一个给定的可观察序列是怎样产生的。简单的来说，就是调整模型的参数，使得模型的 λ 能“学习”观察序列。

Baum-Welch 算法为较常用的 HMM 学习问题的一个近似的解决方法。由于状态序列未知，因此这可以看做是一个含有隐变量的参数估计问题，解决这一问题的经典算法就是 EM 算法。Baum-Welch 算法就是 EM 算法在隐马尔可夫模型学习中的具体体现。

二、HMM 模型商品期货市场应用

2.1 技术指标以及时间段选取

我们选取的时间段为 2010 年 1 月 1 日至 2018 年 2 月 14 日，共 1976 个交易日，数据来源于 Wind。数据选取螺纹钢期货主力合约的价格以及成交量数据，因为商品期货同股票市场有较大差别，同一时间同一商品会有多个合约进行交易。

我们只选取当前的主力合约进行分析，主力合约即当前持仓量最大的合约，随着时间的推移，主力合约会在不同月份合约之间转换。为了避免合约换月带来的价格跳跃，我们在计算收益率时，合约换月当天仍采用次主力合约的数据进行计算。

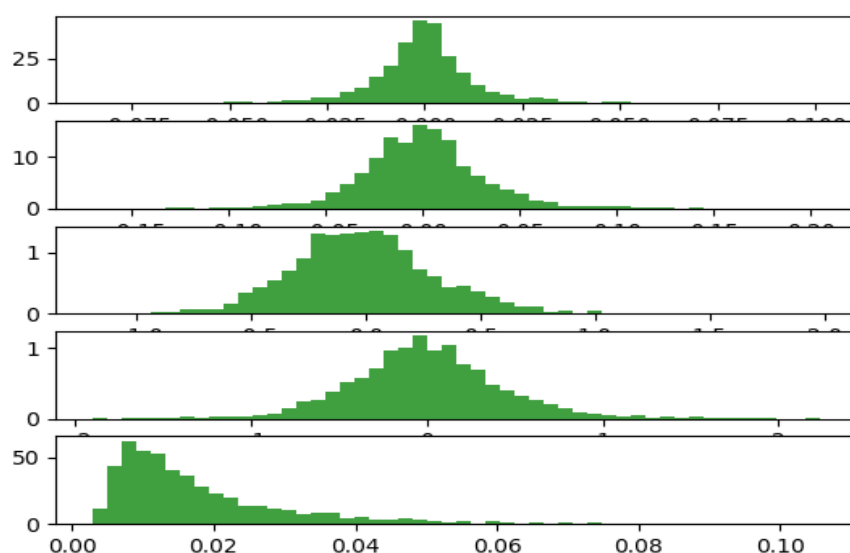
我们将数据分为两部分，进行模型外预测。使用 2010 年 1 月 1 日至 2016 年 11 月 25 日，共 1676 个交易日的的数据建立 HMM 模型。通过建立的模型对 2016 年 11 月 28 日至 2018 年 2 月 14 日共 300 个交易日的的数据进行预测分析，建立交易策略。

价格与成交量的变化能较好的反应当前金融产品的状态信息。技术指标方面，我们选取一日对数成交量差、五日对数成交量差、一日对数收盘价差、五日对数收盘价差以及当日对数高低价差等 5 个指标作为可观察序列带入到本报告的 HMM 模型中进行分析。五个指标的介绍如所示：

- 一日对数成交量差 = $\log(\text{当日成交量}) - \log(\text{上一交易日成交量})$
- 五日对数成交量差 = $\log(\text{当日成交量}) - \log(\text{五日前成交量})$
- 一日对数收盘价差 = $\log(\text{当日收盘价}) - \log(\text{前一日收盘价})$
- 五日对数收盘价差 = $\log(\text{当日收盘价}) - \log(\text{五日前收盘价})$
- 当日对数高低价差 = $\log(\text{当日最高价}) - \log(\text{当日最低价})$

2010 年 1 月 1 日至 2018 年 2 月 14 日，螺纹钢期货主力合约的上述五个技术指标的分布信息如下图所示：

图 2：五个技术指标分布图

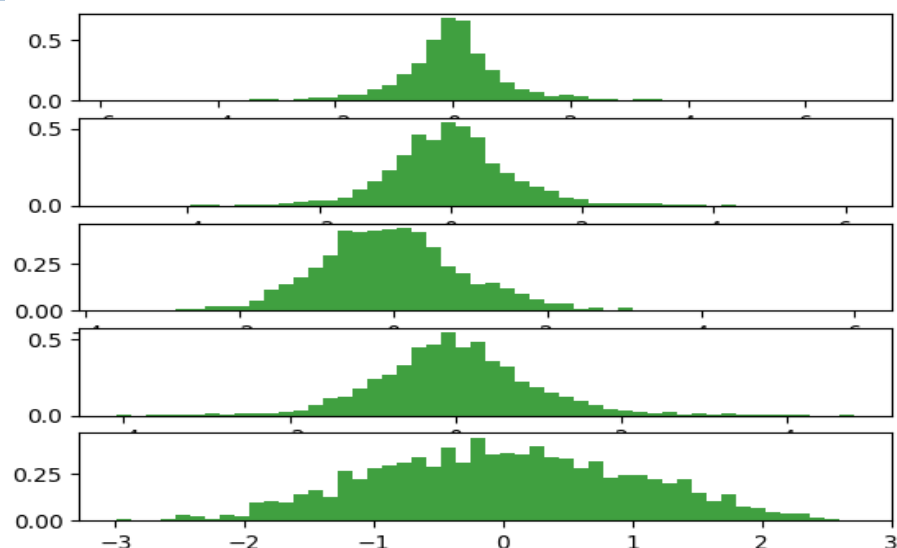


数据来源: Wind 中信期货研究部

从上图中我们可以看到，最后一个可观察序列（当日对数高低价差）较为显著的偏离正态分布。因此对其进行 Box-Cox Transformation 数据转换方法，使其更接近正态分布。此外，在经验协方差矩阵计算过程中，具有较大方差的可观察序列将具有较大的权重。如果可观察序列的单位不同，其方差的差距可能达到数个数量级。

这里为了保证所选的可观察序列具有接近的权重，我们对五个技术指标进行标准化（均值调整为 0，标准差调整为 1 个单位）。经过以上处理后，可观察序列的分布如下图所示：

图 3：处理后的五个技术指标分布图

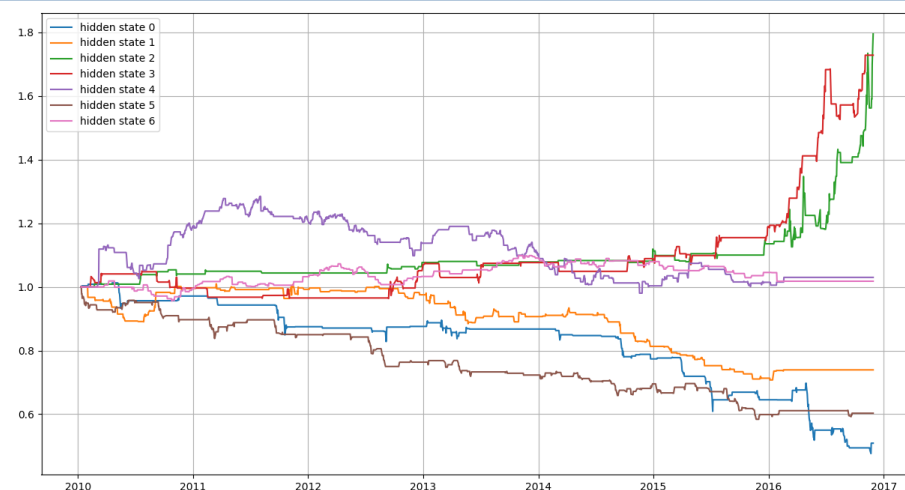


数据来源: Wind 中信期货研究部

2.2 样本内数据分析

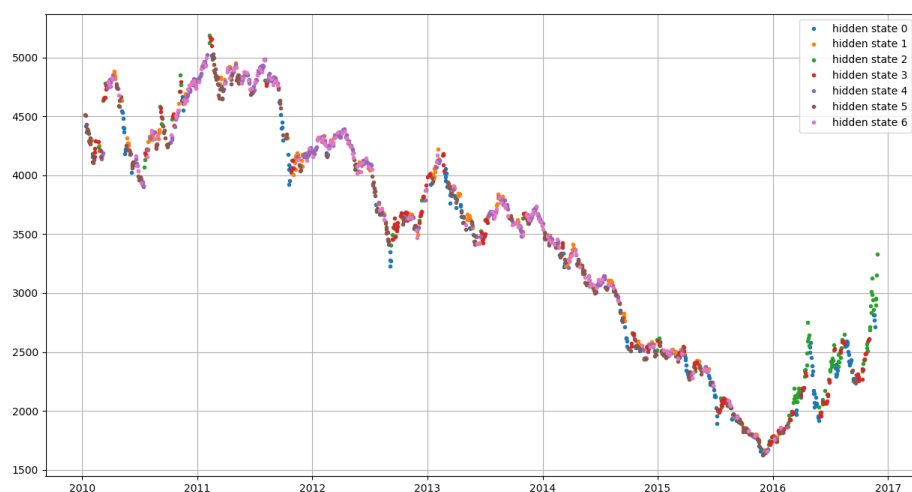
首先对样本内数据进行建模分析，输入的可观察序列为一日对数成交量差、五日对数成交量差、一日对数收盘价差、五日对数收盘价差以及当日对数高低价差等 5 个指标，隐藏状态数设置为 7 个，在出现每一个隐藏状态的第二天买入豆粕期货，分别分析各个隐藏状态的收益信息。

图 4：样本内隐状态收益曲线



数据来源: Wind 中信期货研究部

图 5：样本内隐状态数据标注



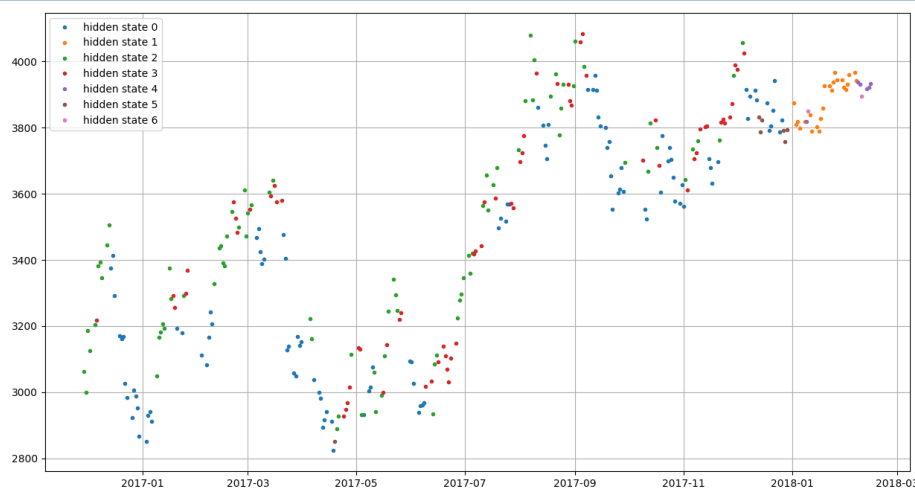
数据来源: Wind 中信期货研究部

从上面两个图可以看到，在样本内数据时间段内，状态 2 和状态 3 为较明显的正收益隐状态，状态 0，状态 1 和状态 5 为较明显的负收益隐状态，状态 6 和状态 4 为震荡隐状态。

2.3 样本外数据分析

我们在接下来的分析中，利用同样的模型对样本外数据进行分析，同样是将样本内外的数据划分为 7 个不同的隐状态。通过样本内模型分析的结果，选择收益为正的隐状态 2 和隐状态 3 在第二天对螺纹钢期货开多头，选择收益为负的隐状态 0，隐状态 1 和隐状态 5 在第二天对螺纹钢期货开空头。分析样本外时间段内的收益信息。

图 6：样本外隐状态数据标注



数据来源: Wind 中信期货研究部

图 7：样本外策略收益曲线



数据来源: Wind 中信期货研究部

2.4 结论与后期优化方向

将适当的技术指标纳入隐马尔可夫模型可观察向量组来预测螺纹钢期货的后期走势能取得一定的效果。但由于隐马尔可夫模型每次运行所得的隐状态分类较为随机，上升隐状态和下降隐状态是我们根据收益曲线人工区分的，模型的稳定性还有待提升。

目前我们只将一日对数成交量差、五日对数成交量差、一日对数收盘价差、五日对数收盘价差以及当日对数高低价差等 5 个指标纳入隐马尔可夫模型可观察向量组进行建模分析，没有考虑其他的一些技术指标。指标的选取范围较为狭窄。

在 HMM 模型的建立过程中并没有对模型的算法以及参数进行设置，只采用了初始的模型算法和参数，通过算法优化和参数设置可能得到更好的结果。

没有对不同隐状态对应的做多、做空操作进行分析，只是简单的选取了收益为正的隐状态做多，收益为负的隐状态做空。

免责声明

除非另有说明，本报告的著作权属中信期货有限公司。未经中信期货有限公司书面授权，任何人不得更改或以任何方式发送、复制或传播此报告的全部或部分材料、内容。除非另有说明，此报告中使用的所有商标、服务标记及标记均为中信期货有限公司的商标、服务标记及标记。中信期货有限公司不会故意或有针对性的将此报告提供给对研究报告传播有任何限制或有可能导致中信期货有限公司违法的任何国家、地区或其它法律管辖区域。

此报告所载的全部内容仅作参考之用。此报告的内容不构成对任何人的投资建议，且中信期货有限公司不因接收人收到此报告而视其为客户。

中信期货有限公司认为此报告所载资料的来源和观点的出处客观可靠，但中信期货有限公司不担保其准确性或完整性。中信期货有限公司不对因使用此报告及所载材料而造成的损失承担任何责任。此报告不应取代个人的独立判断。中信期货有限公司可提供与本报告所载资料不一致或有不同结论的报告。本报告和上述报告仅反映编写人的不同设想、见解及分析方法。本报告所载的观点并不代表中信期货有限公司或任何其附属或联营公司的立场。

此报告中所指的投资及服务可能不适合阁下，我们建议阁下如有任何疑问应咨询独立投资顾问。此报告不构成投资、法律、会计或税务建议，且不担保任何投资及策略适合阁下。此报告并不构成给予阁下的私人咨询建议。

中信期货有限公司2018版权所有并保留一切权利。

深圳总部

地址：深圳市福田区中心三路8号卓越时代广场（二期）北座13层1301-1305、14层

邮编：518048

电话：400-990-8826

传真：(0755)83241191

网址：<http://www.citicsf.com>