

时间序列模型中样本时间窗口的选择

华泰期货研究所 量化组

陈维嘉

量化研究员

☎ 0755-23991517

✉ chenweijia@htfc.com

从业资格号: T236848

投资咨询号: TZ012046

训练样本时间窗口简介

常见的量化模型根据训练样本的结构大致可分成两类:一类是截面模型,这类模型的输入通常是同一个时间节点上因子;另一类是时间序列模型,这类模型的训练样本通常涉及一段连续的时间序列。在训练截面模型时,通常可以随机选取一大部分数据作训练集,一小部分数据作验证集和测试集进行交叉验证,训练样本在时间上的出现顺序没有太严格的要求。而时间序列模型训练样本中的训练集和验证集在时间上往往会有先后顺序,因此在时间序列模型的实际训练通常是使用一部分时间靠前的数据进行模型训练,再去预测时间上靠后的数据,这样就不得不面对一个训练样本的时间窗口长度选择问题。如果使用了太长的数据可能包含了过时的信息,但如果使用了太短的数据就无法对历史情况做出准确的刻画,因此如何选择样本的时间窗口是所有时间序列模型训练中必须面对的一个问题。

在本报告主要研究递归时间窗口和滚动时间窗口训练方法在时间序列模型中的应用。递归时间窗口即根随着时间的流逝,训练样本数量不断增加,而滚动时间窗口则是指随着时间的流逝不断增加新样本,剔除就样本但保持训练样本数量不变。有相关学者的研究指出,如果数据结构发生了较大改变,使用滚动时间窗口的方法会比递归时间窗口要好,但实际上很难准确判断数据结构有否发生改变以及何时发生了改变,所以在实际应用中可以把这两种方法结合从而提高模型的预测效果。本报告尝试把这种方法应用到19种商品期货的周度收益率预测上,对大部分商品而言能够改善预测效果。

研究背景

常见的时间序列模型训练方法通常有三种：

- (1) 固定时间窗口，这种方法训练样本的时间窗口和数据样本量都是固定的。比如有 10 年的数据，那就使用前 8 年的数据来训练时间序列模型，然后用第 9 年的数据进行模型筛选，第 10 年的数据进行模型评价。这种训练方法跟截面模型类似，最大的优势是做模型筛选时只需要训练一次，比较适合运算量大的模型训练。
- (2) 递归时间窗口(recursive window)，这种方法训练样本的时间窗口是递增的。比如有 300 周的数据，最开始时使用前 100 周的数据训练模型，然后对第 101 周进行预测，这第 101 周的预测就相当与是一个样本外的预测结果了。接着再用前 101 周的数据训练模型，再对第 102 周进行预测，……这样依次进行下去，每过一周，训练样本的数量就增加一周，这样最后也能获得 200 个样本外的预测结果，最后根据这 200 个结果进行模型筛选或评价。
- (3) 滚动时间窗口(rolling window)，这种方法训练样本的时间窗口长度是固定的，但是训练样本会做一定调整。例如有 300 周的数据，设定时间窗口是 100 周，那最开始时使用前 100 周的数据训练模型，然后对第 101 周进行预测，接着时间窗口往前移一周，即使用第 2 周至第 101 周的数据训练模型然后预测第 102 周的数据，……这样依次进行下去，每过一周，训练样本的数量就增加最近一周的数据，同时剔除最远一周的数据使得训练样本的总数量不变。这样最后也能获得 200 个样本外的预测结果，最后根据这 200 个结果进行模型筛选或评价。

由于经典的时间序列模型通常都是线性的，训练所需的运算量并不会很大，而且模型使用者更加偏向于考察模型的实际样本外预测效果，所以第一种方法采用得较少。而第二和第三种方法能比较持续地体现出时间序列模型的样本外预测效果，所以实际上有较多的使用。递归时间窗口的方法每次都用所有样本去重新训练模型虽然不需要选择时间窗口，但是用到所有的历史数据也未必是最优的，而滚动时间窗口的方法则需要动态更新训练样本，则又出现该如何选择时间窗的问题。对于模型使用者来说，这的确是一个实际问题，因为利用不同的时间窗口校正出来的模型参数也不同，模型的预测结果也会有所差别。在实际中模型使用者可能会比较随意地选取一个时间窗口进行滚动预测，例如使用一个 2 年至 5 年的时间窗口去训练，然后对比不同模型在这个时间窗口下的预测结果。这样随意选择的结果可能会导致所选的时间窗口不是最合适的从而高估或者低估了模型的预测效果。另外如果根据模型的预测效果去选择最佳的预测窗口，那相当于增加了模型的超参数，从而造成过度拟合的问题。

对于这个训练样本时间窗口应该使用多长的问题，目前学术界的处理方法大致可以分成两类：

- (1) 这类是利用不同的时间窗口训练模型，然后对不同模型的训练结果进行加权平均。通常多模型混合后的预测结果要比使用单一时间窗口的预测结果有所提高。
- (2) Barbara Rossi 等人在 *Out-of-Sample Forecast Tests Robust to the Choice of Window Size* 中提出了一种假设检验的方法，用来衡量模型的预测结果对时间窗的选择是否敏感。这种方法的目的并非去选择最优的时间预测窗口或是去改进预测效果，而是去利用这种假设检验的方法去选择一个对时间窗口不敏感模型。

由于第二种方法的假设检验条件比较复杂，而且对于提高模型预测效果帮助有限，这篇报告侧重介绍第一种方法。这种方法的理论依据首先由 Todd E. Clark 等人在 *Improving Forecast Accuracy By Combining Recursive And Rolling Forecasts* 这篇论文中做了比较详细的分析。这篇论文首先假设在全样本的时间序列里存在，且只存在一个较大的结构性改变，然后根据这个改变的程度去计算递归窗口模型和滚动窗口模型的权重对模型输出结果进行加权平均。这种方法需要计算时间序列里结构改变的时间以及两个时间窗口的权重，这两个参数的计算实际上是很难计算准确的。所以 M. Hashem Pesaran 等人在 *Forecast Combination across Estimation Windows* 中提出使用多个时间窗口进行模型训练，然后进行平均的方法。他们的方法适用于时间序列上出现了多次结构性改变的情况，而且在使用中不需要计算时间序列上何时会出现结构性改变，也不需要计算这些改变的大小，所以这种方法在实现上是比较简单的，可是由于这种方法使用了较多的时间窗口，通常 10 个左右，对于稍复杂的模型而言，无异于大大地增加了计算量，所以这篇报告里也没有对这种方法进行深入研究。计算量比较适中的是 Todd E. Clark 等人的方法，即用递归时间窗口和滚动时间窗口进行加权平均。

时间窗口选择原理

这里主要介绍 Todd E. Clark 等人在 *Improving Forecast Accuracy By Combining Recursive And Rolling Forecasts* 中提出的方法。假设数据的生成符合以下过程(Data Generating Process, DGP):

$$y_{T,t+\tau} = x'_{T,t} \beta_{T,t}^* + u_{T,t+\tau}$$

$$\beta_{T,t}^* = \beta^* + T^{-1/2} g(t/T)$$

其中时间节点 $t = 1, \dots, T, \dots, T + P$, T 为预测的起点, 共有 P 次预测。标量 $y_{T,t+\tau}$ 为待预测值, $k \times 1$ 列向量 $x'_{T,t}$ 为可观测变量。 $u_{T,t+\tau}$ 为随机误差。 $\beta^*_{T,t}$ 为数据生成的列向量, 它的结构可能在不断变化, 其变化快慢和程度由一个函数 $g(t/T)$ 决定。这里用 $\hat{\beta}_{R,t}$ 和 $\hat{\beta}_{L,t}$ 分别表示使用递归时间窗口和滚动时间窗口进行参数估计的 $\beta^*_{T,t}$ 。对应的预测误差分别为 $\hat{u}^2_{R,t+\tau} = (y_{t+\tau} - x'_t \hat{\beta}_{R,t})^2$ 和 $\hat{u}^2_{L,t+\tau} = (y_{t+\tau} - x'_t \hat{\beta}_{L,t})^2$ 。

如果使用了递归时间窗口, 则在估计参数 $\hat{\beta}_{R,t}$ 时, 可能会包含了数据结构出现改变之前的时间序列数据, 这些数据跟预测未来的数据没有关系, 这就导致了模型参数估计和预测产生偏差。这些偏差会累积下去使得模型估计的平方误差与不使用这部分数据相比变大。但如果只使用滚动时间窗口的方法即使用了较少的样本, 这样却会造成模型估计参数的方差变大, 也同样会造成模型预测的平方误差增加。因此选择的时间窗口实际上是需要平衡使用过多或者过少的样本数据。在宏观经济时间序列的预测上, 研究人员往往偏好于使用递归时间窗口, 而金融时间序列的预测上, 研究人员使用滚动时间窗口较多。Todd E. Clark 等人认为, 造成这种现象的原因主要是金融时间序列较长, 因为他们可获得的数据通常为日频, 所以数据量比较大。

根据 Todd E. Clark 等人的研究, 由于参数估计的偏差主要来源于 $\hat{u}^2_{R,t+\tau}$, 方差来源于 $\hat{u}^2_{L,t+\tau}$, 如何平衡偏差和方差关键在于对这两种方法的差异 $\sum_{t=T}^{T+P} (\hat{u}^2_{R,t+\tau} - \hat{u}^2_{L,t+\tau})$ 的分析。Todd E. Clark 等人认为这种差异可以分解成三个随机过程:

$$\sum_{t=T}^{T+P} (\hat{u}^2_{R,t+\tau} - \hat{u}^2_{L,t+\tau}) \rightarrow_d \int_1^{1+\lambda P} \xi_W(s) = \int_1^{1+\lambda P} \xi_{W1}(s) + \int_1^{1+\lambda P} \xi_{W2}(s) + \int_1^{1+\lambda P} \xi_{W3}(s) \quad (1)$$

这三个随机过程分别代表方差的贡献、偏差的贡献以及这两者的共同作用。这三部分又同时受以下因素影响:

- (1) 滚动时间窗口的大小
- (2) 预测的次数
- (3) 参数 β 的维度
- (4) 参数 β 的改变程度, 即函数 g
- (5) 预测的时间长度
- (6) 预测因子的二阶矩

如果假设参数 β 是一直没有变化的, 也就是说无论往前回溯多长的样本数据, 这些样本数据的产生方式都是一样的, 那么使用所有的样本数据进行参数估计将不会带来偏差, 也就是说使用递归时间窗口从而减少方差, 这种方法是总是最优的。

但是如果参数 β 存在一个变化时递归时间窗口就不是最优的了。Todd E. Clark 等人的研究表明，在这种情况下存在一个理论上最优的时间窗口来平衡偏差和方差，但是这个窗口的估计涉及到参数 β 出现变化具体时间节点。如何衡量参数 β 出现变化以及计算出现这种变化的时间点 \hat{T}_B 实际上是一个相当棘手的问题。如果这两方面无法准确计算，那么得到的最优滚动时间窗口也不一定准确。所以 Todd E. Clark 等人提出了一种更加实际的操作方法就是固定一个滚动的窗口做预测，同时也使用递归时间窗口做预测，然后再对两者做加权平均，而这个最优权重 α_t 的计算就取决于滚动时间窗口和滚动时间窗口之前的数据结构。Todd E. Clark 等人认为使用这种方法去改进预测效果也不差。所以在这里就省略了对 \hat{T}_B 进行估计的介绍。新的模型参数可以表示为 $\hat{\beta}_{W,t} = \alpha_t \hat{\beta}_{R,t} + (1 - \alpha_t) \hat{\beta}_{L,t}$ 。最优权重 α_t 可以通过优化公式(1)的期望值，即 $\int_1^{1+\lambda p} E \xi_W(s)$ 得到：

$$\alpha_t = \frac{1}{1 + \left(\frac{t \Delta \hat{\beta}' \hat{B}^{-1} \Delta \hat{\beta}}{\text{tr}(\hat{B} \hat{V})} \right) \hat{\delta}_B (1 - \hat{\delta}_B)} \quad (2)$$

其中 \hat{B} 与预测因子的二阶矩有关：

$$\hat{B} = \left(t^{-1} \sum_{j=1}^t x_j x_j' \right)^{-1}$$

\hat{V} 与模型拟合的残差 \hat{u} 有关：

$$\text{tr}(\hat{B} \hat{V}) = \text{tr} \left[\left(t^{-1} \sum_{j=1}^t x_j x_j' \right)^{-1} \left(t^{-1} \sum_{j=1}^t \hat{u}_{j+\tau}^2 x_j x_j' \right) \right]$$

$\hat{\delta}_B$ 表示出现结构性改变前 \hat{T}_B 在所有拟合样本中的比例：

$$\hat{\delta}_B = \frac{\hat{T}_B}{t}$$

$\Delta \hat{\beta}$ 表示使用结构性改变后 \hat{T}_B 的数据进行参数估计的结果 $\hat{\beta}_2$ 与结果性改变前的估计结果 $\hat{\beta}_1$ 的差异：

$$\Delta \hat{\beta} = \frac{\hat{\beta}_2 - \hat{\beta}_1}{\sqrt{T}} \quad (3)$$

可以看出参数差异 $\Delta \hat{\beta}$ 越大则 $\hat{\alpha}_t^*$ 越小，即使用递归时间窗口的比重越小，滚动时间窗口的比重越大。

模型的实际应用

这里假设期货的周度收益率同时由 N 个可观测因子影响。这里使用的可观测因子包括产销量，商品库存和商品上下游产物等信息，因子个数范围一般在 20-30 之间，使用资产价格或周频回报率。由于可观测的因子数量比较多而且彼此之间的信息有重叠，所以这里使用主成分分析(Principal Component Analysis, PCA)的方法进行降维处理，把因子个数减少至 10 个以下。把期货主力合约的周度收益率作为待预测变量，利用降维后的 PCA 因子作为预测变量进行向量自回归(Vector Autoregression, VAR)。模型训练的时间窗口包含两个，一个是递归时间窗口，另一个是固定为期两年的时间窗口。两年的时间窗口的选择依据主要是为了使模型的周频样本数量能达到 100 个。而 Todd E. Clark 等人训练时间序列模型使用的是 10 年的固定时间窗口，样本采样频率是季频或者月频，所以样本数量与当前使用的较为一致。

为了研究最优权重 α_t 的分布以及变化情况这里列举了模型在螺纹钢、甲醇和豆油这三个品种中每周算出的 α_t 以及与最近 5 周年化波动率的对比，如下图 1-3 所示。由图可见最优权重的变化在每个品种都不一样，变化范围在 0.1 至 0.8 之间，具体取值随着时间变化而变化。但是值得注意的是通常 5 周年化波动率达到局部最高值时，最优权重 α_t 会下降到局部最低点。例如在图 1 中，2016 年 5 月、2016 年 12 月和 2018 年 3 月时螺纹钢的 5 周年化波动率都达到了 0.5 以上，甚至高达 0.8，在这些时间点上最优权重 α_t 会降至 0.2 左右，意味着大部分的权重都会被分配到滚动时间窗口上，递归时间窗口的权重只占 20%左右。通常市场出现较大波动时时间序列也会出现较大的结构性改变也就是公式(3)中 $\Delta\hat{\beta}$ 会比较大从而减少最优权重 α_t 的值。而当 5 周年化波动率处在较低位置时，最优权重 α_t 则相应出现局部最大值，例如在 2016 年 3 月，2016 年 9 月时，螺纹钢波动只有 0.1 左右，这时最优权重 α_t 会在 0.4 至 0.5 之间，相对较大，这也意味着这时使用递归时间窗口的权重较大，因为这时时间序列并没有出现较大的结构性改变。

图 1: 最优权重在螺纹钢上的变化

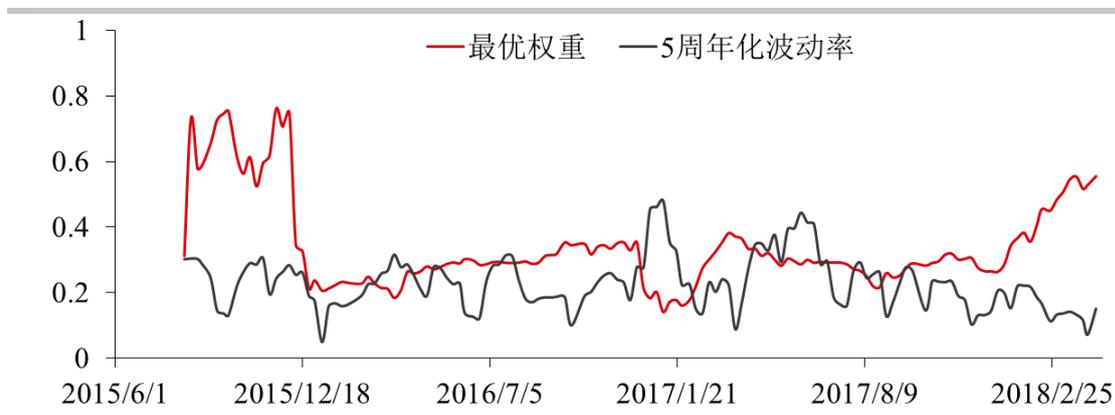


数据来源：华泰期货研究院

最优权重 α_t 的变化特点以及与波动率的关系并非只在螺纹钢上出现，在其他品种如甲醇、豆油上也有类似效果，例如图 2 中甲醇的最优权重 α_t 在 2015 年下半年和 2018 年初都处在高位，

恰好这时甲醇波动在 2015 年 10 月，2018 年 3 月都处在较低的波动区间，最优权重 α_t 能达到 0.5 以上。而其他时间则较低，最优权重 α_t 则较小。

图 2: 最优权重在甲醇上的变化



数据来源：华泰期货研究院

图 3 列举了豆油的最优权重 α_t 变化曲线从 2013 年 11 月起，最优权重 α_t 经历了先上涨到 2015 年 4 月下跌，到 2016 年 9 月再上升的过程。与之相反的豆油波动率走势，从 2013 年年底起是先下降，再上升，然后下降的过程。

图 3: 最优权重在豆油上的变化

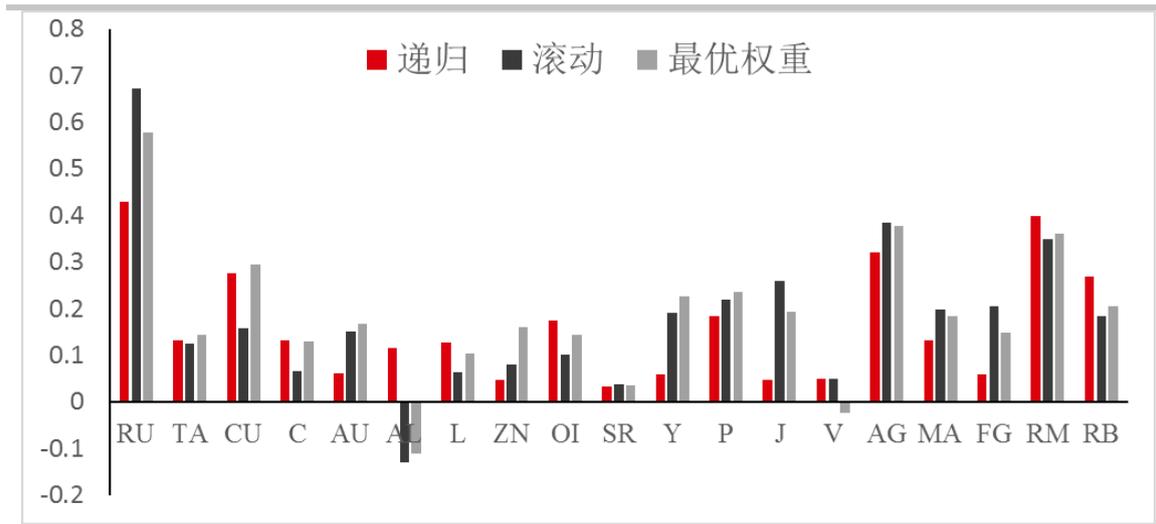


数据来源：华泰期货研究院

通过以上 3 个品种的最优权重 α_t 与波动率走势可以看出大多数时候当波动率上升到历史高位时该方法倾向于使用较短的滚动时间窗口，当波动率处在低位时该方法倾向于使用较长的递归时间窗口。虽然最优权重 α_t 的计算，即公式(2)，并非根据波动率进行的优化但两者之间会有一定的联系。通过对比最优权重 α_t 与波动率的关系，可以得到公式(2)的一个比较直观理解。另外值得注意的是，公式(2)是针对线性多因子模型这一数据生成过程进行的时间窗口选择优化，虽然真实的数据生成过程并非就是简单的线性模型，单从图 1 至图 3 的效果来看，最优权重 α_t 的选取在一定程度上能够符合直观预期。

这种时间窗口的优化方法能应用在橡胶、PTA 和沪铜等 19 个活跃的期货品种上，统计时间从 2015 年 5 月底至 2018 年 4 月初，对比使用 3 种时间窗口的收益统计如图 4 所示。由图可见绝大多数品种在这段时间内都能获得正收益。而且对大部分品种而言，在同一品种内部，使用各个时间窗口训练模型的收益差异并不大。使用递归时间窗口训练这 19 个品种模型没有一个品种会获得负收益，而使用滚动时间窗口训练沪铝会获得负收益。在 PTA、铜、黄金、沪锌、豆油和棕榈油这 6 个品种里最优权重法能获得最好收益。而在橡胶、焦炭、白银、白糖、甲醇和玻璃这 6 个品种里滚动时间窗口法能获得最好收益。在玉米、沪铝、塑料、菜油、菜粕、PVC 和螺纹钢这 7 个品种上递归时间窗口能获得最好收益。从这点上看使用各个时间窗口的收益表现在各个品种上分配比较均匀。并没有哪种方法能取得明显的领先优势。另外在 PVC 这个品种上，虽然递归时间窗口和滚动时间窗口都没有产生负收益，但是使用最优权重法却获得了负收益，所以最优权重法并非任何时候都能取得好效果。毕竟最优权重法是假定了线性的数据生成过程，真实情况与之必定存在差异。

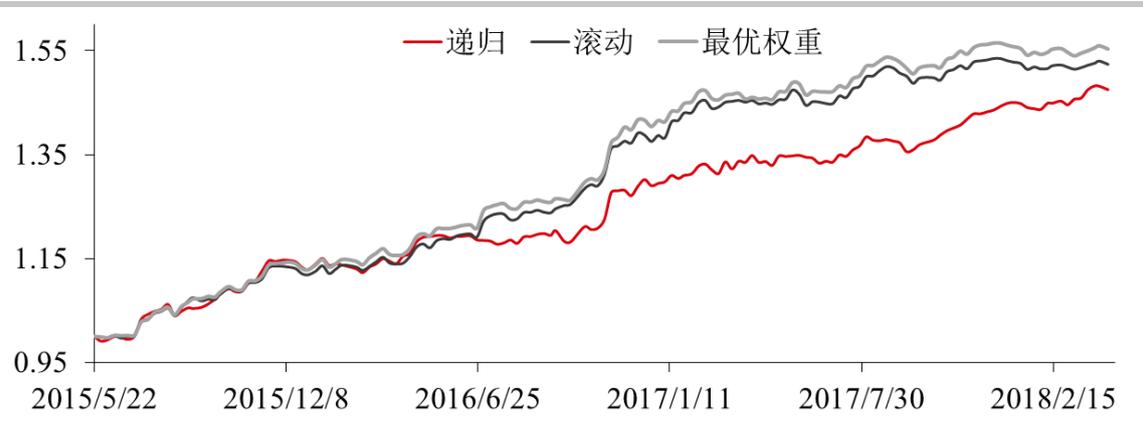
图 4：19 个品种的年化收益



数据来源：华泰期货研究院

图 5 是这 19 个品种按等权重配置的净值曲线。由图可见使用这 3 中方法训练模型效果也比较一致，基本上也是同时产生收益和回撤。这三种方法中，表现最差的是递归时间窗口，最好的是最优权重法。滚动时间窗口的净值曲线与最优权重法在曲线上的差异并不大。表 1 总结了各时间窗口的收益与风险。最优权重法的平均年化收益最高达到，18.69%，夏普率也是最高达到 2.63，而回撤最小只有 2.08%。从数据上看虽然最优权重法，并没有带来太大收益的提升，但是对 19 个品种平均来讲，这种方法无论从收益还是风险的角度来说都是最优的。

图 5: 各时间窗口的策略组合净值曲线



数据来源: 华泰期货研究院

表格 1 各时间窗口的收益与风险统计

	递归时间窗口	滚动时间窗口	最优权重
年化收益	16.05%	17.69%	18.69%
年化波动率	6.95%	7.00%	7.09%
夏普率	2.3	2.53	2.63
最大回撤	2.38%	2.11%	2.08%
最大回撤期	5 交易周	4 交易周	4 交易周

数据来源: 华泰期货研究院

结果讨论

这篇报告介绍了时间序列模型训练时, 时间窗口的选择方法, 包括递归时间窗口、滚动时间窗口和最优权重法。理论上数据结构不发生改变时使用递归时间窗口最优, 当数据结构发生足够大的改变而且时间足够长那只使用改变后的数据是最优的。因此理论上存在一个平衡点使得按一定比例结合递归时间窗口和滚动时间窗口得到一个最优的预测值。而实际上时间序列的数据结构处在不断变化中, 变化的形式也并非按照特定因子进行线性组合, 而且也没有办法准确判断什么时候数据结构发生了改变, 更没法精确衡量数据改变的程度, 所以这篇报告按照 Todd E. Clark 等人的分析方法, 对时间窗口的选择问题进行了一系列的简化处理。

从本报告 19 个品种的测试效果来看, 使用哪种时间窗口的差别并不是很大。可是如果不选择时间窗口而使用递归时间窗口的方法得到的效果是最差的。使用滚动时间窗口要比递归时间窗口好些, 但是只使用滚动时间窗口需要面对时间窗口长度的选择问题, 如果任意选

一个窗口可能会低估模型的预测效果，如果多试几个时间窗口选择最优的窗口则可能造成过度拟合的问题。所以这篇报告参照 Todd E. Clark 等人的方法把两个时间窗口按一定比例相结合，回测结果证明使用最优权重的方法都比前两种时间窗口方法要好。

● 免责声明

此报告并非针对或意图送发给或为任何就送发、发布、可得到或使用此报告而使华泰期货有限公司违反当地的法律或法规或可致使华泰期货有限公司受制于的法律或法规的任何地区、国家或其它管辖区域的公民或居民。除非另有显示，否则所有此报告中的材料的版权均属华泰期货有限公司。未经华泰期货有限公司事先书面授权下，不得更改或以任何方式发送、复印此报告的材料、内容或其复印本予任何其它人。所有于此报告中使用的商标、服务标记及标记均为华泰期货有限公司的商标、服务标记及标记。

此报告所载的资料、工具及材料只提供给阁下作查照之用。此报告的内容并不构成对任何人的投资建议，而华泰期货有限公司不会因接收人收到此报告而视他们为其客户。

此报告所载资料的来源及观点的出处皆被华泰期货有限公司认为可靠，但华泰期货有限公司不能担保其准确性或完整性，而华泰期货有限公司不对因使用此报告的材料而引致的损失而负任何责任。并不能依靠此报告以取代行使独立判断。华泰期货有限公司可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。为免生疑，本报告所载的观点并不代表华泰期货有限公司，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。此报告并不构成投资、法律、会计或税务建议或担保任何投资或策略适合或切合阁下个别情况。此报告并不构成给予阁下私人咨询建议。

华泰期货有限公司 2018 版权所有并保留一切权利。

● 公司总部

地址：广东省广州市越秀区东风东路761号丽丰大厦20层、29层04单元

电话：400-6280-888

网址：www.htfc.com