

## 华泰期货量化专题报告：

### 利用混频 MIDAS 模型预测期货周频收益率

#### MIDAS 模型简介

金融时间序列通常包含多个维度，不同维度数据的采样频率也不一致。例如螺纹钢研究员通常关心螺纹钢的因素有日频更新的现货螺纹钢价格，周频更新的螺纹钢库存，高炉开工率和线螺采购量等。如果其中某些可观测因子发生了变化，投资者对未来螺纹钢期货涨跌也会发生改变。如何及时利用更新的信息进行投资决策是投资者普遍关心的热点。

本报告介绍一种可利用高频数据预测低频数据的时间序列模型：这里我们尝试利用计量经济学里经典的混频数据处理模型 MIDAS(Mixed Data Sampling)来综合螺纹钢基本面的日对周度收益率进行预测。MIDAS 是一种线性回归模型，由于包含较长的日频数据，传统的线性回归模型要求较多的可调参数，所以 MIDAS 的研究者 Eric Ghysels 等人就尝试利用多项式来代替模型中的可调参数。这类模型在学术界主要是用来预测宏观经济指标如 GDP 增长率等，这里我们尝试利用 MIDAS 模型来预测螺纹钢期货周频收益率。

Eric Ghysels 等人提出的原始 MIDAS 模型只适用于低频周期内的预测，例如在每周结束后利用日频数据预测下周的收益率，但却无法在每周结束前根据日频数据的更新来更新对本周收益率的预测。Michelle T. Armesto 等人在此基础上进行了改进，通过在模型中添加高频领先项实现了低频周期内的预测。

在本报告里首先介绍了 MIDAS 模型的原理和结构，然后尝试利用周频和日频的螺纹钢因子数据对螺纹钢期货主力周频收益率进行预测，发现随着高频项的引入，这种模型在预测周频收益率方面均方根误差有所减少，虽然在螺纹钢这个品种上未能通过提高调仓频率来有效增加收益，但是能减少最大回撤。

华泰期货研究所 量化策略组

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

## MIDAS 模型研究背景

基本面研究员在预测期货未来走势时通常会考察不同时间频率的信息，例如反映通胀的居民消费价格指数(CPI)，商品进出口增长，周频的商品库存以及日频的现货价格、期货价格等等。然而目前的量化模型大多数都只使用了相同频率数据的信息，例如按周频调仓的模型就只是用周频数据，日频数据通常会被降为周频。从数据频率的使用角度上说，量化模型所考虑的信息要比实际可获得的信息要少，这也是低频量化模型的不足之处。从持仓周期来说，一个按周频调仓的量化模型，如果在一周内期货价格发生了不利于投资者的剧烈波动，那模型使用者则需要根据自己的风险承受能力来决定是否承担浮亏或者止损离场，又或者借助投资经验和主观判断来进行投资决策。这种决策虽然不无道理，但往往会导致投资过程变为量化加主观，投资收益也无法进行合理回测。出现以上困难的主要原因在于当高频变量出现变化时，低频量化模型无法及时更新预测。因此本报告试图利用日频数据和周频数据来实现每天更新期货周频收益率的预测。

大量学术研究表明，利用高频变量能够有效提高低频变量的预测准确性。在考虑单个低频因子 $Y_t^Q$ 和单个高频因子 $X_t^Q$ 的情况下，低频因子 $Y_t^Q$ 可以用如下多项式展开

$$Y_{t+1}^Q = \mu + \sum_{j=0}^{p_Y^Q-1} \alpha_{j+1} Y_{t-j}^Q + \sum_{j=0}^{p_X^Q-1} \beta_{j+1} X_{t-j}^Q + \varepsilon_{t+1} \quad (1)$$

其中 $p_Y^Q$ 为低频变量的延时， $p_X^Q$ 为高频变量的延时， $\mu$ 为截距， $\varepsilon_{t+1}$ 是误差项。该式子中的 $\mu$ ， $\alpha_{j+1}$ 和 $\beta_{j+1}$ 属于待调参数，所以共有 $p_Y^Q + p_X^Q + 1$ 个待调参数。如果高频延时 $p_X^Q$ 较大，例如当季频和日频数据混合时，当 $p_Y^Q=3$ ， $p_X^Q=66$ ，那将有 70 个待调参数。由于模型的输入和输出使用的主要是宏观经济数据，可获得的样本量非常有限，大量的可调参数往往导致模型出现过度拟合从而降低样本外预测效果。仔细观察公式(1)可以发现模型参数主要来自高频项系数 $\beta$ 。

Eric Ghysels 等人研究的 MIDAS(Mi(xed) Da(ta) S(ampling))模型通过引入多项式函数来减少高频项系数 $\beta$ ，如下式所示

$$Y_{t+1}^Q = \mu + \sum_{j=0}^{p_Y^Q-1} \alpha_{j+1} Y_{t-j}^Q + \beta \sum_{j=0}^{q_X^Q-1} w_j(\theta^D) X_{t-j}^D + \varepsilon_{t+1} \quad (2)$$

其中高频变量 $X_t^Q$ 前的系数使用了多项式函数 $w_{i+j*N_D}(\theta^D)$ 表示， $w_j(\theta^D)$ 有多种函数形式可以选择，一种比较常见的是指数型 Almon 多项式

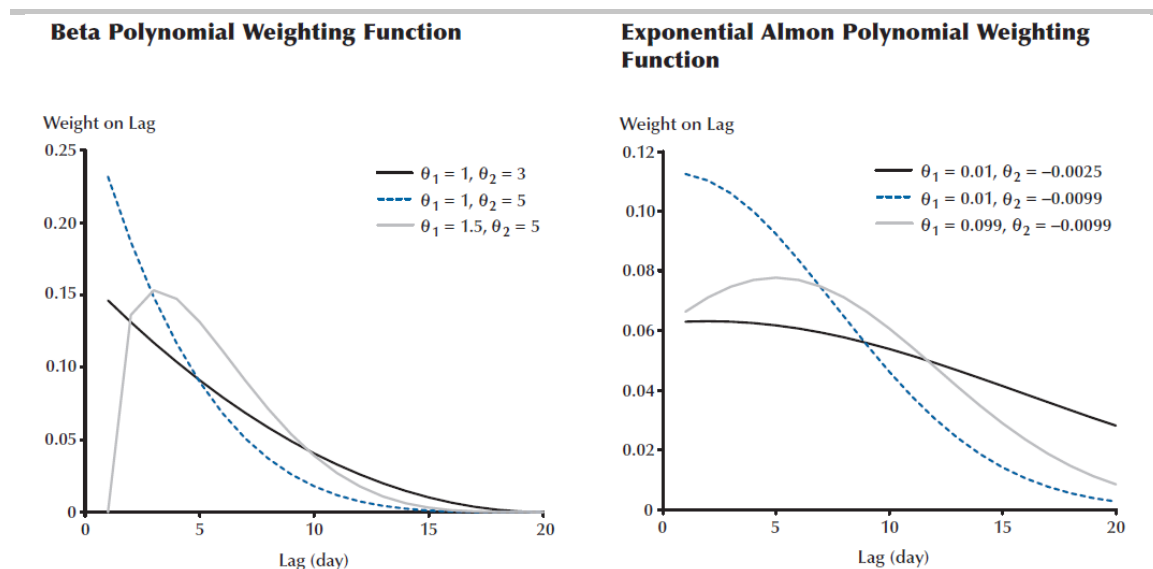
$$w_j(\theta^D) \equiv w_j(\theta_1, \theta_2) = \frac{\exp\{\theta_1 j + \theta_2 j^2\}}{\sum_{j=1}^m \exp\{\theta_1 j + \theta_2 j^2\}} \quad (3)$$

其中 $w_j(\theta^D)$ 只包含了 $\theta_1$ 和 $\theta_2$ 两个参数。这样待校正参数的数量就大大降低了，从而有效防止参数过多造成的模型过度拟合。常见的 MIDAS 多项式除了指数型 Almon 外，还有 Beta 型：

$$w_j(\theta^D) \equiv w_j(\theta_1, \theta_2) = \frac{x_i^{\theta_1-1}(1-x_i)^{\theta_2-1}}{\sum_{i=1}^N x_i^{\theta_1-1}(1-x_i)^{\theta_2-1}} \quad (4)$$

图 1 画出了这两种多项式取不同参数值时高频权重系数 $w_j(\theta^D)$ 在不同延时上的分布。由图可见，一般情况下 $w_j(\theta^D)$ 是随延时日数的增加而减少，也就是说临近的高频项影响通常较大，时间较长的高频项影响较少，MIDAS 多项式的这个特征与直观认识相符。另外也可以看到当 MIDAS 多项式参数 $\theta^D$ 取不同值时也可能出现延时 2-3 天的系数权重比延时 1 天要大，这体现了 MIDAS 多项式的灵活性，即根据具体数据去调整高频项的权重。对比 Beta 多项式和指数型 Almon 多项式也可以发现 Beta 多项式的权重随延时日数的增加下降较快，而指数型 Almon 多项式则下降较缓。

图 1: MIDAS 多项式



数据来源：Michelle T. Armesto *et al*, Forecasting with Mixed Frequencies

MIDAS 模型通常使用在利用日频/周频/月频数据预测季频 GDP 上。在混频领域预测比较成功的还有状态空间(state-space)模型，但这种模型实现比较复杂，MIDAS 在 GDP 上的预测效果通常不会比状态空间模型差很多。

Eric Ghysels 等人研究的 MIDAS 模型虽然能利用高频变量降低低频变量的预测误差，但其在期货交易中的实际应用会有一些限制，例如：

- (1) 只能利用高频变量预测低频变量，也就是说在期货上如果结合日频和周频数据，则只能利用 MIDAS 预测周频收益率。

(2) 只能在收集完所有高频变量后才能预测低频变量，也就是说在期货交易上只能在一周最后一个交易日(通常是周五)预测下周的周频收益率，而无法在当周，比如第一个交易日预测当周的收益率。

(3) MIDAS 为单因子模型只有一个待预测的低频变量和一个高频变量。

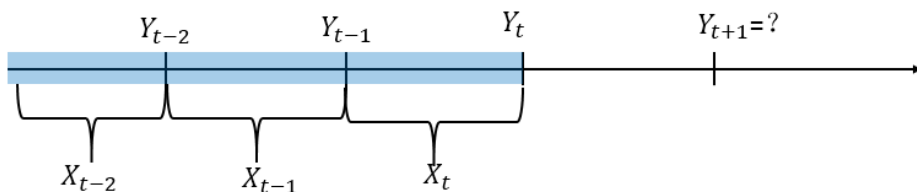
对于限制(1),目前 MIDAS 模型是无法改变的，但是如果能改变限制 (2) 即每过一个交易日都能利用日频更新的数据对当周的收益率重新预测，那也相当于提高了模型的交易频率。这里参考 Michelle T. Armesto 等人在 Forecasting with Mixed Frequencies 中给出的方法进行改进。对于限制(3)，Eric Ghysels 等人的做法是利用多个高频变量分别对低频变量进行预测，然后再对预测结果做加权平均。

## MIDAS 模型改进方法

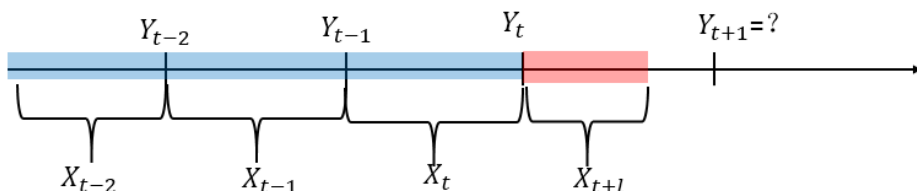
下面针对限制(2)和(3)，利用 Michelle T. Armesto 等人的方法对 MIDAS 模型进行改进。图 2 对比了原 MIDAS 模型和改进后要达到的效果，原 MIDAS 模型是利用低频变量 $Y_t, Y_{t-1}, Y_{t-2}, \dots$ 以及期间的高频变量 $X_t, X_{t-1}, X_{t-2}, \dots$ 预测 $Y_{t+1}$ ，当中并没有使用在 $Y_t$ 和 $Y_{t+1}$ 之间出现的高频数据，而改进后的 MIDAS 是要达到低频周期内的高频数据 $X_{t+l}$ 更新后重新预测 $Y_{t+1}$ 。

图 2: MIDAS 预测周期

原MIDAS: 只能做**周期外**的低频预测



改进版: 做**周期内**的低频预测



数据来源：华泰期货研究院

另外，改进版的 MIDAS 模型通过增加外生变量的方法，同时引入多个低频变量进行预测。改进后的 MIDAS 模型可用以下式子表示

$$Y_{t+1|d} = \mu + \sum_{j=0}^p \alpha_{j+1} Y_{t-j}^Q + \gamma \sum_{j=m-d+1}^m w(j; \theta) X_{t+1-j} + \gamma \sum_{k=0}^m w(k; \theta) X_{t-k} + \sum_{i=0}^n \sum_{j=0}^p \lambda_{ij} E_{i,t-j} + \varepsilon_{t+1} \quad (5)$$

其中  $\gamma \sum_{j=m-d+1}^m w(j; \theta) X_{t+1-j}$  为高频领先项表示低频周期内的高频变量，它与高频延时项  $\gamma \sum_{k=0}^m w(k; \theta) X_{t-k}$  使用相同的 MIDAS 多项式和参数  $\theta$ 、 $\gamma$ ，主要是为了限制参数数量。在低频周期内每当有高频变量  $X$  更新时  $Y_{t+1|d}$  也随之更新。 $\sum_{i=0}^n \sum_{j=0}^p \lambda_{ij} E_{i,t-j}$  为低频因子，下标  $i$  表示低频因子数量。

Michelle T. Armesto 等人利用这个改进后的 MIDAS 模型在季度内进行预测发现能降低季频 GDP 的预测误差。接下来本报告尝试利用螺纹钢的周频和日频基本面数据预测螺纹钢主力期货的周频收益率。

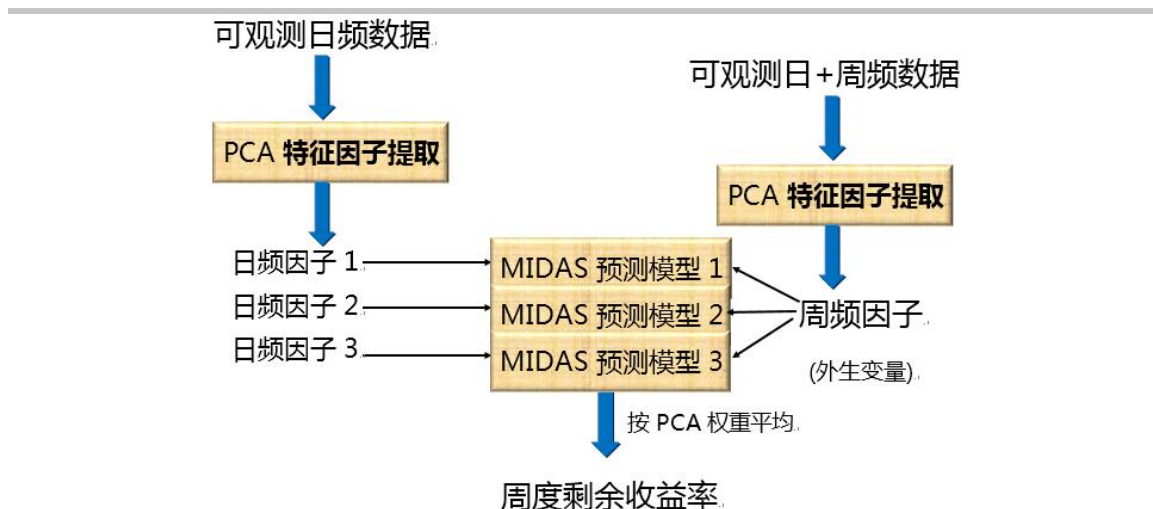
## MIDAS 应用实例

在实际应用中 MIDAS 模型只有在每周最后一个交易日(通常为周五)才进行低频周期外预测，即预测下一周的周频收益率，这时公式(5)就不包含高频领先项  $\gamma \sum_{j=m-d+1}^m w(j; \theta) X_{t+1-j}$  了，这个时候公式(5)就相当于普通的 MIDAS 模型。而在其他交易日(通常为周一至周四)则包含有该高频领先项。预测是基于每周发布的螺纹钢库存、高炉开工率和钢厂产能等因子以及日度发布的螺纹钢、铁矿石、焦煤日收益率等因子。由于可观测因子的个数较多，这里先使用主成分分析(Principle Components Analysis, PCA)进行降维处理，得到的模型使用结构可以用图 3 表示。

可观测的日频数据首先利用 PCA 进行降维处理得到多个日频因子，隔周的日频因子放入公式(5)中的延时项  $\gamma \sum_{k=0}^m w(k; \theta) X_{t-k}$ ，每周最后交易日更新，当周的日频因子放入公式(5)中的领先项  $\gamma \sum_{j=m-d+1}^m w(j; \theta) X_{t+1-j}$ ，每个交易日更新。另外，可观测的日频数据通过降低采样频率变为周频，然后与可观测的周频数据一起降维得到周频因子，放入外生项  $\sum_{i=0}^n \sum_{j=0}^p \lambda_{ij} E_{i,t-j}$ ，每周最后交易日更新。对每个日频因子使用相同的周频因子训练 MIDAS 模型预测周频收益率，也就是说有多少个日频因子就要有多少个 MIDAS 模型需要训练。在预测时再根据日频因子的对 PCA 组合方差的贡献度对预测结果进行加权平均。模型训练采用普通的最小二乘法最小化周内各天的周度收益率预测误差，模型样本采用每周递增的方法进行滚动训练，从 2013 年 5 月开始，每过一周训练样本数增加一。模型包含的超参数包括低频、高频变量的因子个数和延时等。值得注意的是 MIDAS 模型预测的是周度收益率，在实际使用上必须把每个交易日预测的周度收益率减去本周已实现的收益率从而得到剩余收益率的期望值。比如在周二收盘后预测本周的周度收益率，然后减去周一和周二这两个交易日的实现收益率从而得到本周的剩余收益率，最后根据剩余收益率的正负进行开平仓操作。周度剩余收益率为正则开多仓，反之为空仓。



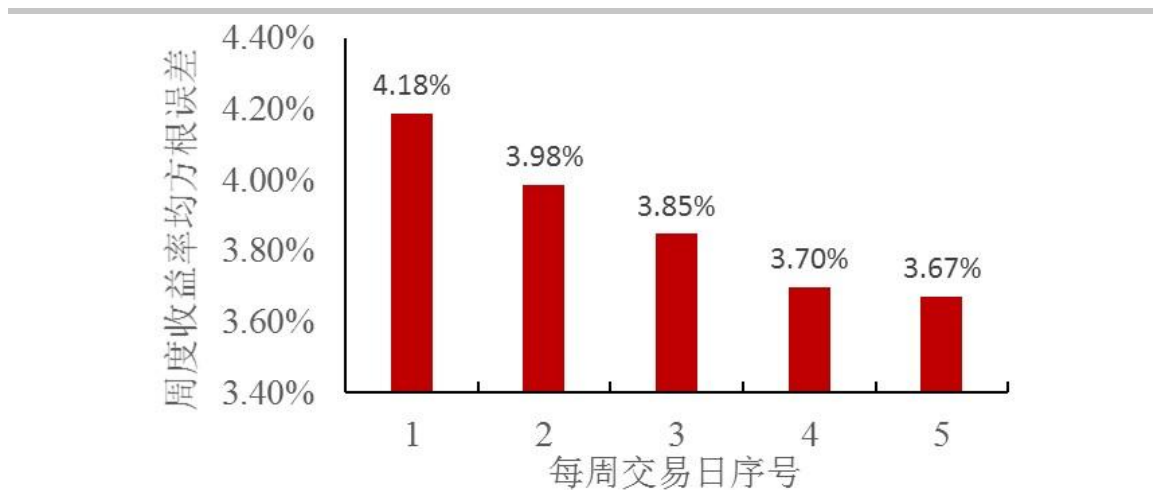
图 3: MIDAS 使用模型结构



数据来源：华泰期货研究院

由于 MIDAS 模型能实现周内预测，这里首先考察该模型在第零至第五交易日预测螺纹钢收盘价周度收益率的均方根误差。图 4 中可以看出预测周度收益率误差随着每周交易日的结束而减少，即在第一交易日，即在通常周五收盘后做预测模型的误差最大为 4.18%，之后逐渐减小，到第四交易日，即临近一周结束时预测的误差最小。从结果上看这点与直观预期是一致的，随着一周临近结束，周度收益率的可变空间也在减小，所以均方根误差也逐渐减小。

图 4: MIDAS 周频收益率均方根误差

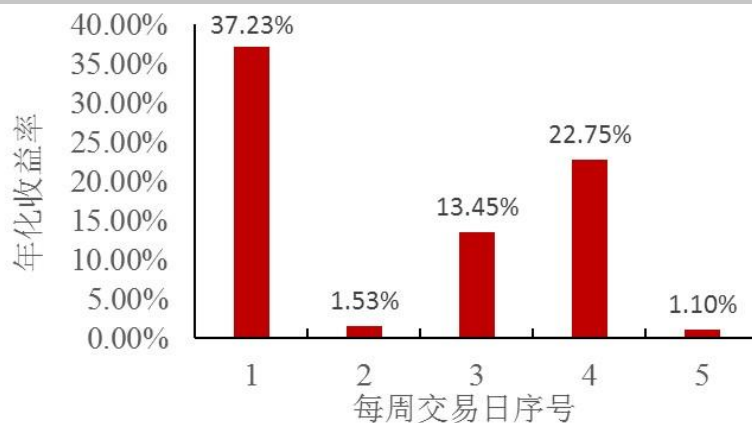


数据来源：华泰期货研究院

虽然预测误差较大，但使用 MIDAS 模型仍能获得正收益，图 5 列出了 MIDAS 模型的收益在第一至第五交易日的累积分布。值得注意的是，在图 5 中的 1 表示第一次预测，通常在上周五收盘，第一次预测通常会持仓 5 天，第二次预测通常在周一，持仓 4 天，依次类推。各次预测的收益均用年化表示，由图可见，虽然第一次预测的均方根误差最大，但是收益也是最大的，达到 37.23%。其次是第 4 次预测(通常是周三收盘后),年化收益率达

到 22.75%，再次是第三次预测，年化收益为 13.45%。其他时间的预测收益率均较低，但仍有正收益。

图 5: MIDAS 在不同交易日的年化收益



数据来源：华泰期货研究院

MIDAS 模型应的预测数据和交易数据全部使用每个交易日的收盘价，图 6 对比了 MIDAS 模型只在第一交易日调仓，即上周结束后（通常周五）预测，然后持仓一周的情况；以及把资金分 5 份，每个调仓进场的结果。图 6 中的黑线和左边纵轴为第一交易日调仓的结果，红线和右边纵轴为每日调仓的结果。由图可见只在第一交易日调仓，然后持仓到一周结束的曲线比较平稳向上，效果比频繁交易的每日调仓要好。也就是说从收益的角度上来讲，利用 MIDAS 模型频繁调仓并非最优选择。

图 6: 螺纹钢 MIDAS 策略在第一交易日调仓与每日调仓的净值曲线



数据来源：华泰期货研究院

螺纹钢 MIDAS 策略的表现统计如下表格 1 所示，如果只在第一交易日调仓并持有一周，年化收益可达 37.23%，如果每天调仓则收益下降为 11%，年化波动率方面两者相差不多，分别为 12.67%和 11.51%，由于第一交易日调仓的收益率较高，其整体夏普率也较高达到 2.93，但同时最大回撤也高达 26.23%，相比之下每日调仓策略的最大回撤只有 14.01%，所以从最大回撤的

角度上来讲利用 MIDAS 模型合理增加调仓频率在一定程度上是能够减少回撤的。但最大回撤期方面每日调仓策略达到 120 交易日，而第一交易日调仓策略则只有 50 个交易日。

表格 1 螺纹钢 MIDAS 策略收益与风险统计

	第一交易日调仓	每日调仓
年化收益	37.23%	11%
年化波动率	12.67%	11.51%
夏普率	2.93	0.955
最大回撤	26.23%	14.01%
最大回撤期	50 交易日	120 交易日

数据来源：华泰期货研究院

## 结果讨论

本报告首先对 Eric Ghysels 等人提出的 MIDAS 模型进行了基本介绍。这种模型主要是为了适应包含两个不同频率的金融时间序列，其主要原理是利用多项式代替高频延时项的权重系数从而减少模型参数。由于原模型只能在低频周期结束后进行预测，Michelle T. Armesto 等人在此基础上进行了扩展，通过加入高频领先项实现低频周期内的预测。本报告主要采用 Michelle T. Armesto 等人提出的方法利用螺纹钢周频和日频数据对周度收益率进行预测。随着一周的结束，MIDAS 模型预测周频收益率的误差也逐渐下降，说明利用高频变量预测低频变量能提高准确度，这点与大多数相关文献的结果一致。虽然根据 MIDAS 模型每日预测剩余收益率的结果进行交易却没有提升收益，但是却能够在一定程度上减少策略的最大回撤。



## ● 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