

期权策略专题（六）： 股指波动率预测：舆情分析、深度学习能否战胜传统计量模型？



报告日期：2022 年 12 月 15 日

★研究背景：

波动率的预测一直是金融市场中（尤其是期权投资领域）非常关注的课题。本报告分别从数据端和模型端两个角度，引入舆情数据和机器学习模型，优化波动率预测的准确率。

★模型选择：

多元回归模型：本报告采用 OLS、Lasso 和 Ridge 作回归分析模型，其中 Lasso 和 Ridge 是在 OLS 最小化残差平方和的基础上分别加入了 L1 和 L2 范数的惩罚项。**自回归模型：**采用 GARCH（广义条件异方差模型）对波动率时间序列进行自回归建模。**深度学习模型：**LSTM 是基于 RNN 改进的具有记忆长短期信息的能力的神经网络，更加善于处理具有较长周期时序关系依赖的数据。**舆情分析：**基于贝叶斯模型进行文本的情感分类，通过中文语义分析库 SnowNLP 完成。

★模型结果：

价量数据的多元回归周度（5 日）和日度（1 日）正确率为：**64.16%和 65.05%**；GARCH 模型周度（5 日）和日度（1 日）正确率为：**65.80%和 62.93%**；价量数据的 LSTM 模型周度（5 日）和日度（1 日）正确率为：**70.73%和 67.08%**；补充舆情数据后多元回归周度正确率为：**63.95%**；补充舆情数据后 LSTM 周度正确率为：**67.17%**。LSTM 表现最好，回归模型中 Lasso 优于 OLS 和 Ridge，GARCH 模型随参数敏感性较大；舆情数据自身具有一定预测能力，却无法为价量数据提供增益信息。上述模型预测效果均未在样本外出现明显衰减。

★风险提示：

市场风格的切换会造成特征有效性发生变化，导致模型效果下降。

王冬黎 高级分析师(金融工程)
从业资格号: F3032817
投资咨询号: Z0014348
Tel: 8621-63325888-3975
Email: dongli.wang@orientfutures.com

联系人： 谢怡伦（分析师）
从业资格号: F03091687
Tel: 8621-63325888-1585
Email: yilun.xie@orientfutures.com

相关报告

《期权策略专题（五）——基于择时的波动率做空以及 Delta 动态对冲》

《衍生品量化对冲系列（三）：期权对冲策略组合实证研究》

目录

1、研究背景	5
2、数据	5
3、回测框架	7
4、多元回归	8
4.1、多元回归模型	9
4.1.1、OLS	9
4.1.2、Lasso	9
4.1.3、Ridge	10
4.1.4、样本数量不足情况下的讨论	10
4.2、多元回归结果分析	10
4.3、样本外	13
5、GARCH	14
5.1、模型介绍	14
5.2、模型结果分析	15
5.3、样本外	17
6、LSTM	18
6.1、模型介绍	18
6.2、模型结果分析	19
6.3、样本外	21
7、舆情分析	22
7.1、模型介绍	22
7.2、数据	22
7.3、结果分析	25
8、舆情数据数据扩充	25
8.1、多元回归	27
8.2、多元回归样本外	27
8.3、LSTM	28
8.4、LSTM 样本外	29
9、总结	29

图表目录

图表 1：价量因子	5
图表 2：上证 50 历史波动率	7
图表 3：滚动回归框架	8
图表 4：多元回归预测	11
图表 5：多元回归预测结果（5 日）	11
图表 6：不同回看周期下多元回归结果（5 日）	11
图表 7：不同回看周期下多元回归结果（1 日）	12
图表 8：不同预测周期结果（固定回看周期为 120 天）	12
图表 9：多元回归样本外表现	13
图表 10：多元回归样本外预测结果（5 日）	13
图表 11：多元回归样本外预测结果（1 日）	14
图表 12：GARCH 模型表现	15
图表 13：GARCH 模型回测结果（5 日）	15
图表 14：GARCH 模型回测结果（1 日）	16
图表 15：GARCH 模型样本外表现	17
图表 16：GARCH 模型样本外回测结果（5 日）	17
图表 17：GARCH 模型样本外回测结果（1 日）	18
图表 18：RNN 模型原理	18
图表 19：LSTM 模型回测结果（5 日）	19
图表 20：LSTM 模型回测表现	20
图表 21：LSTM 模型回测结果（1 日）	20
图表 22：LSTM 模型样本外表现	21
图表 23：LSTM 模型样本外回测结果（5 日）	21
图表 24：LSTM 模型样本外回测结果（1 日）	21
图表 25：舆情数据举例	22
图表 26：舆情数据情感评分	23
图表 27：舆情数据梳理	23
图表 28：波动率 VS 情绪指标（MA5）	24
图表 29：舆情数据回归结果	25
图表 30：舆情数据样本外回归结果	25
图表 31：因子相关性	26
图表 32：价量+舆情数据多元回归回测结果	27
图表 33：价量+舆情数据多元回归样本外回测结果	27

图表 34：价量+舆情数据 LSTM 表现.....	28
图表 35：价量+舆情数据 LSTM 回测结果.....	28
图表 36：价量+舆情数据 LSTM 样本外表现.....	29
图表 37：价量+舆情数据 LSTM 回测结果.....	29

1、研究背景

波动率的预测一直是金融市场中非常关注的课题。无论是资产定价、风险管理还是在构建投资组合的过程中，对市场波动率的估计占据着十分重要的位置。在期权市场上，一种被广泛采纳的波动率估计方式是利用期权价格倒推出来的隐含波动率作为对未来市场波动率的估计，然而受限于国内期权市场起步较晚，市场上容易出现了很多非理性的交易行为，造成期权价格的非理性偏差，故由此所推出的隐含波动率无法公允地反应市场的未来波动率。

本报告采用的方式是基于波动率的历史数据，通过对历史价量数据以及舆情数据进行建模以预测未来的市场波动率走势，当然这种波动率估计方式也存在一定弊端：一方面容易出现模型过拟合的风险出现；另一方面，基于历史数据建模的先验假设是历史会不断重演。针对过拟合的问题，本报告通过区分样本内和样本外数据进行优化：基于样本内数据建模，再利用样本外数据验证，以此来降低模型过拟合的风险；针对历史重演的质疑，本报告会对波动率数据进行平稳性检验，相较于纯粹的市场价格数据，波动率数据更趋于平稳，故可以认为其“历史重演”的概率更高。

2、数据

在多元回归以及LSTM部分，选取上证50ETF的价量数据作为模型的输入因子，包括开盘价、最高价、收盘价、成交量等共计20个价量因子。在基础价量数据的基础上，计算其多日滚动均值以扩充其时序上的信息，同时计算其偏度和峰度以加入数据的统计特征。

图表1：价量因子

high	close	vol	ma60	skew_20
num_trades	low	return	vol_ma5	kurt_20
open	total_turnover	ma5	vol_ma20	skew_60
volume	iopv	ma20	vol_ma60	kurt_60

资料来源：东证衍生品研究院

而在舆情分析部分，梳理财经软件上股民的评论数据作为输入因子，本报告统计了每一天的发帖数量、阅读数量、帖子情绪指标等数据作为模型的情绪因子。

图表 2：情感因子

发帖数量	阅读数量	评论数量	总情感得分
平均阅读数	平均评论数	平均情感得分	

资料来源：东证衍生品研究院

本报告基于上证 50ETF 的 5min 数据的对数价格，利用无模型方法计算出 t-1 到 t 时刻的已实现波动率，以此来作为标的历史波动率的估计。首先需要计算每天的对数收益率：

$$r(t+\Delta, \Delta) = \log[s(t) + \Delta] - \log[s(t-1)], \Delta > 0$$

其中 Δ 为时间间隔，当 t 取整数时，则表示日间价格对数收益率有：

$$r(t, 1) = \log[s(t)] - \log[s(t-1)]$$

而日内价格对数收益率向量可以表示为：

$$r\left(t + \frac{i}{n}, \frac{1}{n}\right) = \log\left[s\left(t + \frac{i}{n}\right)\right] - \log\left[s\left(t + \frac{i-1}{n}\right)\right], i = 1, 2, \dots, n$$

日内对数收益率与日间对数收益率存在以下关系：

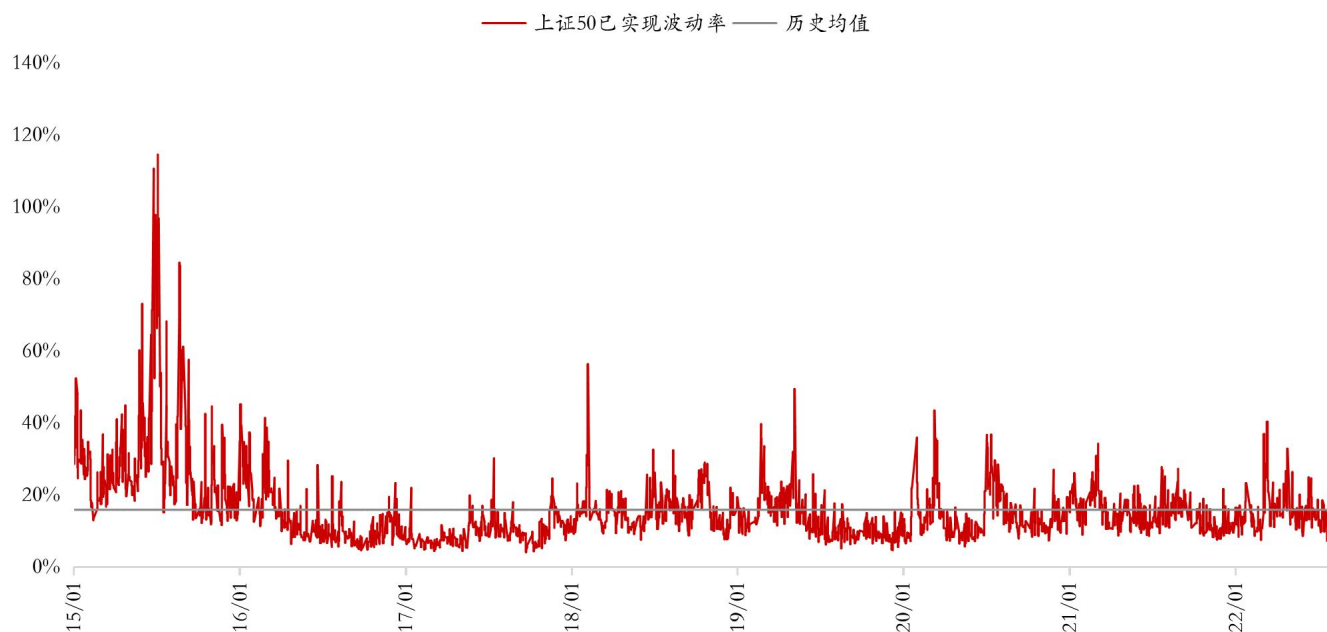
$$r(t-1) = \sum_{i=1}^n r\left(t + \frac{i}{n}, \frac{1}{n}\right)$$

由于本报告取 5min 级别价格数据，每天共有 48 的数据，由上述公式以及已实现波动率的定义可以得到：

$$RV_t = \sum_{i=1}^{48} r_{t,i}^2$$

由上式可以计算得到第 t 日的已实现波动率。

图表 3：上证 50 已实现波动率



资料来源：东证衍生品研究院

上图考察了从 2015 年初至 2022 年 8 月的上证 50 的波动率水平，其历史中枢为 15.90%。可以看到长期来看，波动率数据围绕着中枢进行震荡，而当极端行情时，波动率则会迅速攀升至较高水平，而后快速回落。通过上图可以较为直观地发现，波动率序列相对平稳，为验证这一猜想，下面对其进行单位根检验：

其 ADF 检验统计量为 -4.544 (t 值)，P 值为 0.00016，在 99%，95%，90% 不同置信区间下临界的 ADF 检验值分别为 -3.4339、-2.8631、-2.5676，不难发现 t 统计量明显小于各个置信区间下的 ADF 检验值，且 P 值明显小于给定的显著水平，十分接近于 0，故拒绝原假设，认为该波动率序列为平稳序列。

3、回测框架

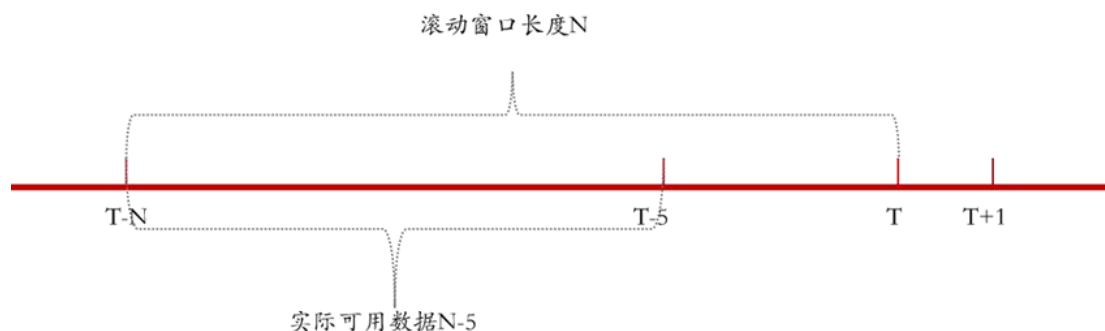
本报告首先依据多元回归进行滚动回归预测。OLS（普通最小二乘法）多元回归的原理为，最优拟合曲线使得各点到直线的距离的平方和（残差平方和 RSS）最小：

$$RSS = \sum_{t=1}^T (y_t - \hat{y}_t)^2 = \sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2$$

而 Lasso 和 Ridge 回归则是在 OLS 的基础上加入了 L1 和 L2 正则惩罚项。

本报告采用滚动回归的方式进行回测，以尽可能模拟真实预测环境。以周度预测为例，首先设置相应的滚动回归窗口长度N，对每一天T都截取T-N到T时间段的基本面数据，由于为周度预测，需要对价量数据进行5天的移动处理，所以实际可用数据点为N-5天内的数据，随后根据训练模型得到一系列预测值，再根据预测值与真实值的比较去构造回测模型。LSTM模型和舆情分析部分的回归方式与多元回归一致，均采用滚动回归的方式。

图表 4：滚动回归框架



资料来源：东证衍生品研究院

4、多元回归

在这一部分，通过多元回归检验股指的价量数据对波动率的预测能力。此处所采用的OLS、Lasso、Ridge模型均是传统的多元线性回归模型，其基础形式如下：

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

其中 $\{x_{i1}, x_{i2}, \dots, x_{ik}\}$ 表示个体i的k个自变量， y_i 为个体i的因变量，而 $\{\beta_1, \beta_2, \dots, \beta_k\}$

为系数矩阵，多元回归的逻辑是通过约束条件对 $\{\beta_1, \beta_2, \dots, \beta_k\}$ 进行参数估计，由此即可以得到回归方程。无论是OLS，还是Lasso和Ridge回归，其参数估计的核心思想都是一致的，即残差平方和最小化： $\min \sum (y_i - \hat{y}_i)^2 = \min \sum e_i^2$ ，稍有差异的是Lasso回归和Ridge回归在OLS的基础上分别加入了L1和L2范式的惩罚项，所以这三个回归模型的参数估计的约束条件分别为：

$$\text{OLS: } \min \sum e_i^2$$

$$\text{Lasso: } \min \sum e_i^2 + \lambda \|\hat{\beta}\|_1$$

$$\text{Ridge: } \min \sum e_i^2 + \lambda \|\hat{\beta}\|_2^2$$

4.1、多元回归模型

4.1.1、OLS

前文提到 OLS 模型的约束方程为 $\min \sum (y_i - \hat{y}_i)^2 = \min \sum e_i^2$ ，其中 $e_i = y_i - \hat{y}_i$ 表示实际值和估计值之间的差值，即估计误差，故该约束方程的意义是求得一组系数矩阵 $\{\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k\}$ ，使得所有点的误差平方之和最小，得到的参数估计方程为 $\hat{\beta} = (X'X)^{-1}X'y$ 。

OLS 模型可以有效解决基础的多元线性回归问题，然而在实际建模过程中，自变量 $\{x_{i1}, x_{i2}, \dots, x_{ik}\}$ 的数据质量不一定能够满足 OLS 回归的条件。通过 OLS 的参数估计方程可以得到，若 $X'X$ 不是满秩的，则不存在该项的逆矩阵。对应到实际问题当中，若样本个数小于待估参数数量，则无法进行 OLS 回归。此外，若自变量 $\{x_{i1}, x_{i2}, \dots, x_{ik}\}$ 存在多重共线性的问题，同样会导致 $X'X$ 几乎不可逆。为应对上述问题，在 OLS 的基础上引入相应的惩罚项，即 Lasso 回归和 Ridge 回归。

4.1.2、Lasso

Lasso 回归在残差平方和最小化的基础上加入了 L1 范数的惩罚项，使得约束方程变为：

$$\min \sum e_i^2 + \lambda \|\hat{\beta}\|_1 = \min \sum (y_i - \hat{y}_i)^2 + \lambda \sum_{u=1}^k |\hat{\beta}_u|$$

不难发现 $\lambda \sum_{u=1}^k |\hat{\beta}_u|$ 是绝对值的形式，在零点除不可导，故不具有解析解，但可以使用梯度下降法求解。

Lasso 回归的最大特点是其可以产出稀疏的系数矩阵，即使得一些不重要的回归系数缩减为 0，起到删除部分变量的作用。

4.1.3、Ridge

Ridge 回归在残差平方和最小化的基础上加入了 L2 范数的惩罚项，使得约束方程变为：

$$\min \sum e_i^2 + \lambda \|\hat{\beta}\|_2^2 = \min \sum (y_i - \hat{y}_i)^2 + \lambda \sum_{u=1}^k (\hat{\beta}_u)^2$$

该方程解析解为：

$$\hat{\beta} = (X'X + \lambda I)^{-1} X'y$$

可以发现 L2 范数惩罚项使得 $X'X$ 项变成了 $(X'X + \lambda I)$ ，保证了该项满秩，并且可逆，但由于惩罚项的引入，使得回归系数的估计不再无偏。故 Ridge 通过放弃参数估计的无偏性、降低精度的方式解决病态矩阵的回归问题。

4.1.4、样本数量不足情况下的讨论

当样本长度小于自变量数量时，多元回归的参数估计方程 $\hat{\beta} = (X'X)^{-1} X'y$ 是无法进行求解的，原因在于 $X'X$ 不满秩，故无法求可逆矩阵。此时可以采用求解伪逆矩阵的方式进行参数估计方程的求解，其形式如下：

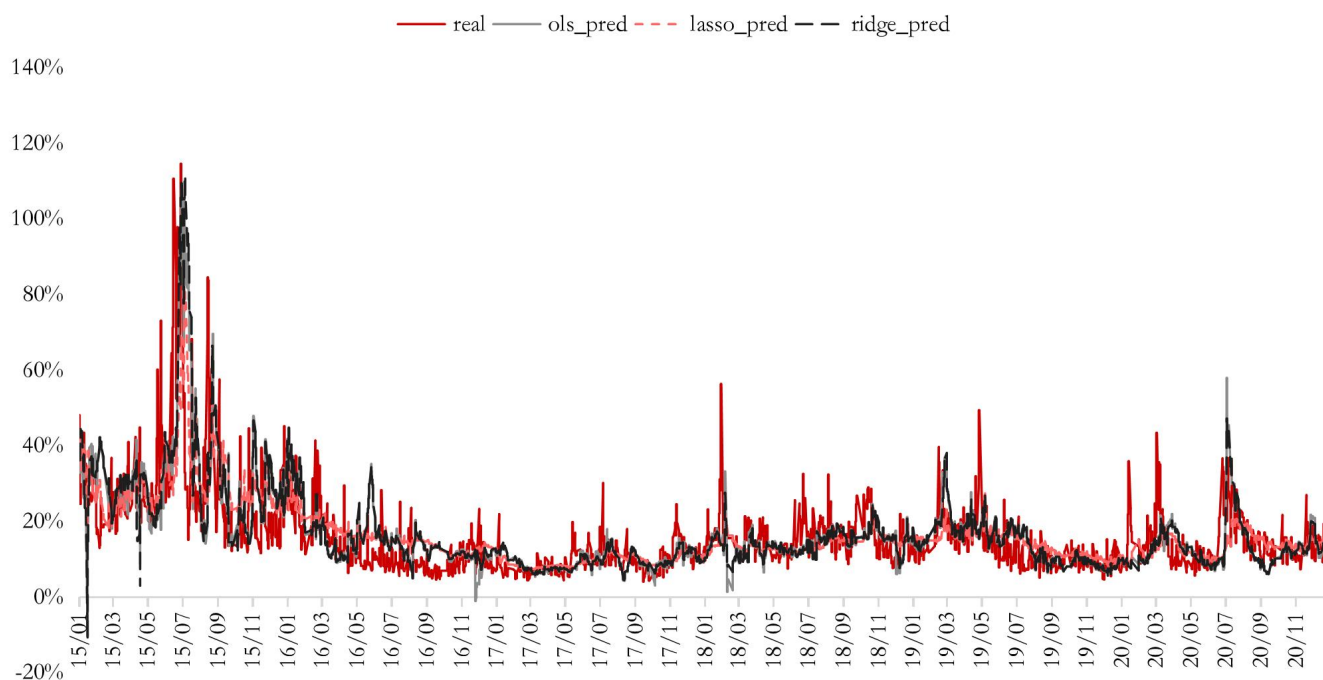
$$\text{pinv}(X) = (X'X)^{-1} X'$$

本报告在建模过程中若遇到样本长度不够的情况，均采用此方法进行求解。

4.2、多元回归结果分析

下图展示的是利用 250 个交易日作为回看周期时，OLS、Lasso 和 Ridge 回归的结果。大体的趋势上，三个模型均能够很好地预测波动率的走势，然而 OLS 和 Ridge 会在某些时段出现异常值（小于 0 的情况出现），潜在原因可能是自变量数量较大，导致出现一定程度的多重共线性，而 OLS 和 Ridge 在模型层面并不会剔除冗余因子。因此当共线性因子系数大批量出现负数情况下，会导致最终拟合结果出现波动率为负的情况。相较之下 Lasso 回归由于产出结果为稀疏系数矩阵，通过赋予 0 系数的方式可以有效剔除部分变量，故 Lasso 回归的结果没有出现异常现象。

图表 5：多元回归预测



资料来源：东证衍生品研究院

图表 6：多元回归预测结果（5 日）

	OLS	LASSO	RIDGE
MAE	0.0608	0.0527	0.0618
MSE	0.0104	0.0074	0.0131
RMSE	0.1019	0.0858	0.1143
胜率	0.6142	0.6306	0.6149

资料来源：东证衍生品研究院

从整体误差的角度，Lasso 回归的预测能力更强；从胜率方面来看，Lasso 回归对波动率的预测可以达到 63%，可以认为具有较为显著的预测能力，相比较之下，OLS 和 Ridge 的正确率略低，为 61%。

图表 7：不同回看周期下多元回归结果（5 日）

	OLS	LASSO	RIDGE
10 天	0.7107	0.5486	0.4925
20 天	0.5691	0.5787	0.4733
60 天	0.6135	0.6313	0.5527
120 天	0.5739	0.6416	0.5739
250 天	0.6142	0.6306	0.6149

资料来源：东证衍生品研究院

考虑到不同回看周期设置下，预测结果可能出现较大差异，本文同样回测了在不同尺度的回看周期下各个线性回归模型的预测能力，此处以胜率作为主要的参考指标。不难发现，OLS 对参数的敏感性较强，在回看周期为 10 天时，胜率最高达到 71.07%，而随着回看周期的拉长，OLS 模型的胜率呈现并无明显规律的变化，总体而言胜率在 60% 附近出现较大幅度震荡。相较之下，Lasso 和 Ridge 模型随着回看周期的拉长，模型胜率逐渐趋于稳定，这其中，Lasso 回归的效果明显好于 Ridge。当回看周期较短时（小于等于 20 天时），Ridge 回归的胜率不足 50%，而随着回看周期不断变长，Ridge 的胜率可以达到超过 60% 的水平；相较之下，Lasso 在回看周期较短的情况下依然可以达到约 55% 的胜率，而随着回看周期进一步拉长，其胜率维持在 63% 左右。

上文的回测过程均对波动率进行周度级别的回测，考虑到模型数据样本均为日度数据，同样也可以进行日度级别的波动率预测，下图考虑在日度预测框架下，各多元回归模型的正确率如何。

图表 8：不同回看周期下多元回归结果（1 日）

	OLS	LASSO	RIDGE
10 天	0.7025	0.6368	0.5971
20 天	0.6026	0.6635	0.6101
60 天	0.6156	0.6539	0.6252
120 天	0.6300	0.6505	0.6430
250 天	0.6532	0.6450	0.6607

资料来源：东证衍生品研究院

相较于周度级别的预测，日度级别波动率预测准确率并无显著变化。其中 OLS 和 Ridge 模型的能力略有提升，日度级别预测下，OLS 和 Ridge 在不同回看周期下均能够给出超过 60% 的预测正确率。

进一步地，固定回看周期为 120 天，考察不同预测周期下模型的预测能力如何变化。

图表 9：不同预测周期结果（固定回看周期为 120 天）

	1 日	2 日	3 日	4 日	5 日	10 日	20 日
mae	0.0413	0.0456	0.0484	0.0513	0.0540	0.0613	0.0711
mse	0.0044	0.0058	0.0065	0.0073	0.0083	0.0102	0.0137
rmse	0.0663	0.0763	0.0803	0.0857	0.0909	0.1011	0.1172
胜率	0.6505	0.6464	0.6347	0.6477	0.6423	0.6450	0.6354

资料来源：东证衍生品研究院

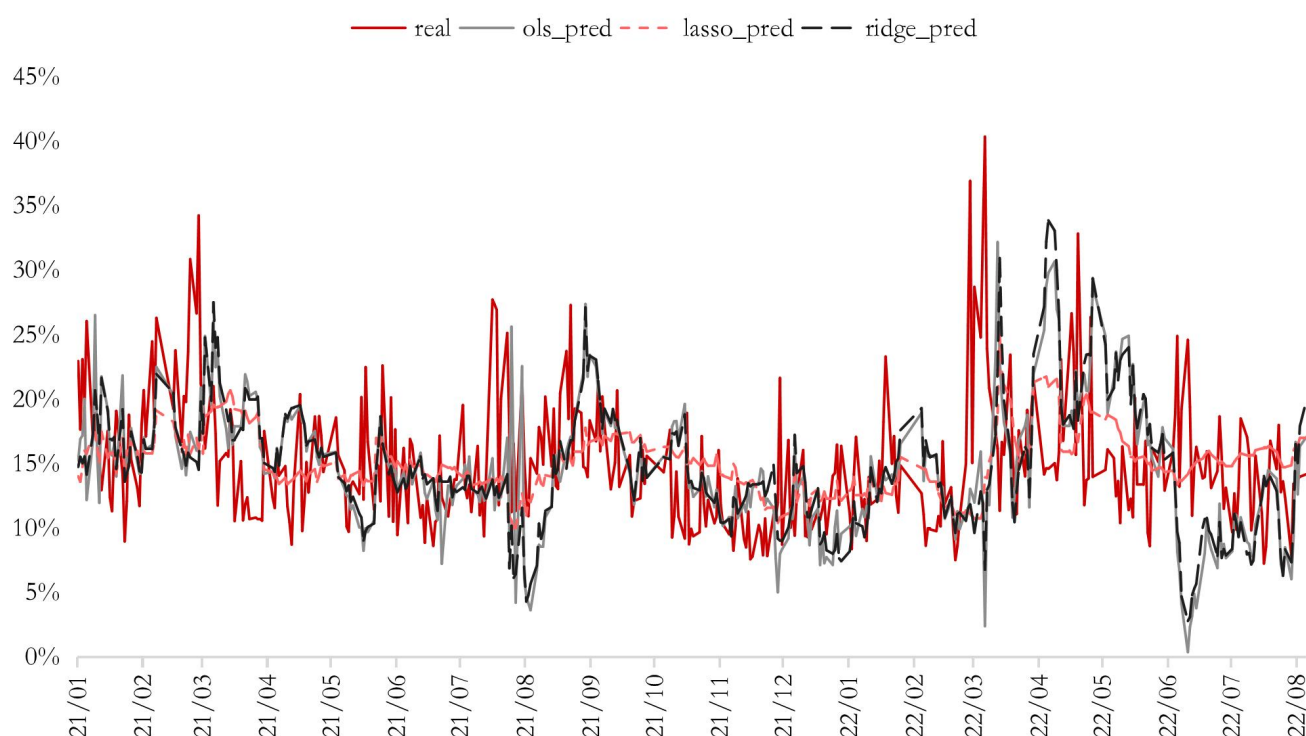
数据显示，各项指标随预测周期变长而不断下降，也就意味着模型对远期波动率的预测能力小于对近期波动率的预测能力。这一点也是符合直观预期的，对于跨度周期越长的指标预测，模型预测能力难免预测衰减。

4.3、样本外

本报告预留了2021年1月至2022年8月的数据作为样本外数据，对模型进行进一步验证。选取回看周期为120天的数据进行比对，整体效果相较于样本内数据非但没有下降，反而存有一定程度的增长，说明模型在样本外依旧有效。

需要说明的是，根据前述分析，在不同回看周期下，当参数设置为120天时，模型整体的预测能力较强，故在样本外检验部分统一以回看周期为120天作为参数设置，并且在后续的分析过程中，若无特殊说明，均以120天作为回看周期。

图表 10：多元回归样本外表现



资料来源：东证衍生品研究院

图表 11：多元回归样本外预测结果（5日）

	OLS	LASSO	RIDGE
MAE	0.0510	0.0409	0.0517
MSE	0.0047	0.0031	0.0048
RMSE	0.0688	0.0560	0.0696
胜率	0.6305	0.6848	0.6202

资料来源：东证衍生品研究院

上表展示了样本外模型的表现情况，与样本内的表现对比，模型整体预测能力非但没有显著衰减，反而有一定提升。首先对于样本内模型的参数设置，并没有过多调参的操作，唯一的参数就是回看周期的设置，故在一定程度上避免了过拟合的情况出现。其次，模型在市场波动加剧时的预测能力显著下降，而2021初至2022年8月这一时间段，市场整体的波动率较为平稳，所以造成了样本外表现好于样本内的现象。最后，观察图像可以发现，与样本内一样，OLS和Ridge在某些时段会出现异常值，比如在2022年6月初，OLS和Ridge模型均给出了接近于0的预测值，明显与真实数据不符，相较之下，Lasso的预测值更为准确。

图表 12: 多元回归样本外预测结果 (1 日)

	OLS	LASSO	RIDGE
MAE	0.0391	0.0344	0.0383
MSE	0.0025	0.0021	0.0025
RMSE	0.0504	0.0461	0.0497
胜率	0.6796	0.6770	0.6667

资料来源：东证衍生品研究院

相似地，上图考察了日度级别预测再样本外的准确率，把回看周期同样设置为120天。数据显示，多元回归模型在样本外表现并未出现衰减，甚至相较于样本内存在一定提升，这一现象的潜在原因在前一段已进行分析。此外，相较于周度预测的正确率，在样本外日度预测正确率并无显著区别，三个回归模型的正确率均落在67%左右。

5、GARCH

5.1、模型介绍

GARCH模型作为ARCH模型的推广，被广泛用于金融时间序列数据的建模过程中。

ARCH模型的实质是使用残差平方序列的 q 阶移动平均拟合当期异方差函数值，由于移动平均模型具有自相关系数 q 阶截尾性，所以ARCH模型实际上只适用于异方差函数短期自相关系数。

但是在实践中，有些残差序列的异方差函数是具有长期自相关性，这时使用ARCH模型拟合异方差函数，将会产生很高的移动平均阶数，增加参数估计的难度并最终影响ARCH模型的拟合精度。

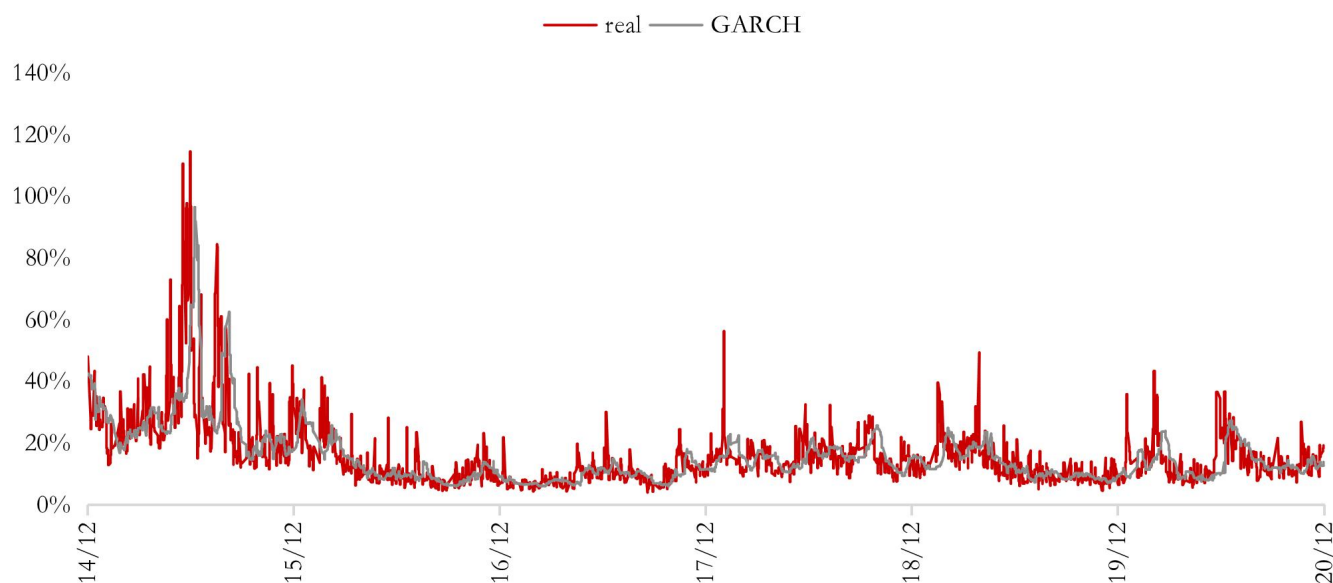
为了修正这个问题，提出了广义自回归条件异方差模型，这个模型简记为GARCH(p, q)。GARCH模型实际上就是在ARCH的基础上，增加考虑异方差函数的 p 阶自回归性而形成，它可以有效的拟合具有长期记忆性的异方差函数。相较于利用权益市

场的价量数据对波动率进行多元回归，GARCH 模型是金融界更为常用的对金融时序数据进行建模的模型。

$$GARCH : \sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

5.2、模型结果分析

图表 13: GARCH 模型表现



资料来源：东证衍生品研究院

图表 14: GARCH 模型回测结果 (5 日)

	10 天	20 天	60 天	120 天	250 天
mae	0.0311	0.0383	0.0498	0.0560	0.0684
mse	0.0027	0.0049	0.0082	0.0096	0.0132
rmse	0.0518	0.0703	0.0903	0.0979	0.1148
胜率	0.8010	0.7565	0.6785	0.6580	0.6231

资料来源：东证衍生品研究院

GARCH 模型的预测效果整体优于多元回归，且由于该模型的预测周期为 5 天，故随着回看周期逐渐拉长，预测准确率逐步降低，回看时间越长，预测值越接近于长期均

值。当回看周期为 10 天时，模型预测准确率达到惊人的 80% 以上，随着回看周期逐步拉长到 250 天，模型准确率逐步稳定在 62%。

图表 15: GARCH 模型回测结果 (1 日)

	10 天	20 天	60 天	120 天	250 天
mae	0.0352	0.0437	0.0508	0.0562	0.0685
mse	0.0038	0.0065	0.0082	0.0095	0.0132
rmse	0.0619	0.0804	0.0904	0.0972	0.1150
胜率	0.7312	0.6936	0.6525	0.6293	0.5985

资料来源：东证衍生品研究院

为了进一步验证 GARCH 在不同周期下的预测能力，同样对未来一天的波动率进行预测，比较可得，日度预测的准确率相较于周度预测在不同回看天数下全面下滑。无论是 10 日回看周期，还是 250 日回看周期，可以发现日度准确率下降 3%-7% 不等。这一点与之前在多元回归模型中观察到的现象是不一样的。

为进一步检验该现象，固定回看天数为 120 天，讨论在不同预测周期下模型的效果。数据表明，对于各个误差指标，随着预测周期逐渐拉长，误差逐渐变大，这一现象与多元回归结果是一致的。但是对于胜率这一指标而言，并未呈现出明显的规律性变化。

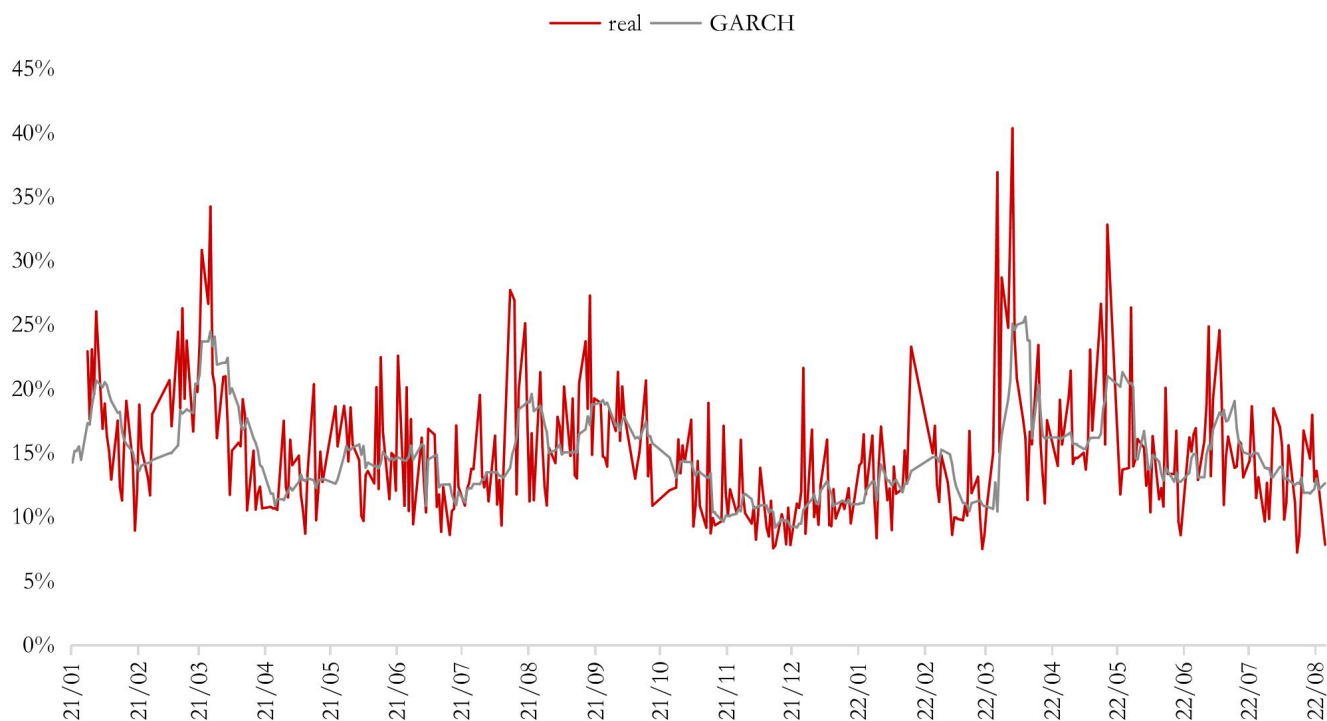
图表 16: GARCH 不同预测周期结果 (固定回看周期为 120 天)

	1 日	2 日	3 日	4 日	5 日	10 日	20 日
mae	0.0562	0.0561	0.0560	0.0560	0.0560	0.0560	0.0574
mse	0.0095	0.0095	0.0095	0.0096	0.0096	0.0099	0.0112
rmse	0.0972	0.0974	0.0975	0.0977	0.0979	0.0995	0.1059
胜率	0.6293	0.6662	0.6436	0.6505	0.6580	0.6970	0.7038

资料来源：东证衍生品研究院

5.3、样本外

图表 17: GARCH 模型样本外表现



资料来源：东证衍生品研究院

图表 18: GARCH 模型样本外回测结果 (5 日)

	10 天	20 天	60 天	120 天	250 天
mae	0.0267	0.0312	0.0362	0.0390	0.0391
mse	0.0013	0.0018	0.0026	0.0029	0.0029
rmse	0.0360	0.0427	0.0506	0.0536	0.0536
胜率	0.8217	0.7881	0.7132	0.6925	0.6848

资料来源：东证衍生品研究院

同样选取 2021 年 1 月至 2022 年 8 月的数据作为样本外数据对模型进行进一步检验。由于在使用样本内数据建模时并未过度调参，故在样本外模型表现并未出现衰减，且与样本内情况一样：回看天数越短，模型胜率越高，随着回看周期逐步拉长到 250 天，模型准确率逐步稳定在 68%。

图表 19: GARCH 模型样本外回测结果 (1 日)

	10 天	20 天	60 天	120 天	250 天
mae	0.0299	0.0339	0.0372	0.0393	0.0393
mse	0.0016	0.0023	0.0027	0.0029	0.0029
rmse	0.0404	0.0481	0.0519	0.0539	0.0538
胜率	0.7545	0.7132	0.7003	0.6899	0.6822

资料来源: 东证衍生品研究院

对于日度预测而言, 其样本外表现同样未出现衰减, 且其与周度预测准确率的差距在样本外得到缩小。虽然在回看周期较小时 (10 天和 20 天), 日度准确率仍低于周度准确率 7% 左右, 而在回看周期较长时 (60 天、120 天和 250 天), 日度预测准确率与周度预测准确率相近。

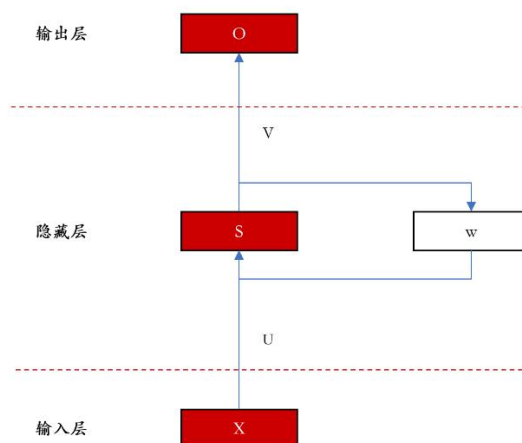
6、LSTM

6.1、模型介绍

LSTM (长短期记忆网络) 是在传统循环神经网络 RNN 的基础上, 为解决梯度爆炸问题而进行的优化。

RNN (Recurrent Neural Network), 又称循环神经网络, 是一类用于处理序列数据的神经网络。基础的神经网络包括输入层, 隐藏层, 输出层, 并通过激活函数控制输出, 层与层之间通过权值连接, 而 RNN 在此基础上, 对层之间的神经元也建立了权连接, 基于此, RNN 便具有了记忆功能。

图表 20: RNN 模型原理



资料来源: 东证衍生品研究院

上图显示， U 是输入层到隐藏层的权重矩阵， V 是隐藏层到输出层的权重矩阵。循环神经网络中隐藏层的值 S 不仅仅取决于当前这次的输入 X ，还取决于上一次隐藏层的值 S ，权重矩阵 W 就是隐藏层上一次的值作为这一次的输入的权重，因此 RNN 网络对上一时刻的隐藏层具有记忆功能。

期权方面，帮助大型金融机构优化期权对冲方案，针对市场主流的期权对冲策略进行实证研究；同时研发期权相关套利策略，以协助私募基金提升期权策略表现。

然而传统的 RNN 会遇到的一个问题是后面时间的节点对于前面十点节点的感知力下降。于是在 RNN 记忆模型中再加入遗忘门，输入门和输出门，使得可以记忆更加长期的信息，这就是 LSTM（长短期记忆）模型，LSTM 是一种特殊的 RNN 模型。遗忘门（forget gate）在 LSTM 中一定的概率控制是否遗忘上一层的隐藏细胞状态；输入门（input gate）负责处理当前序列位置的输入，确定什么样的新信息被存放在细胞状态中；输出门（output gate）则对更新后的细胞状态进行输出。

6.2、模型结果分析

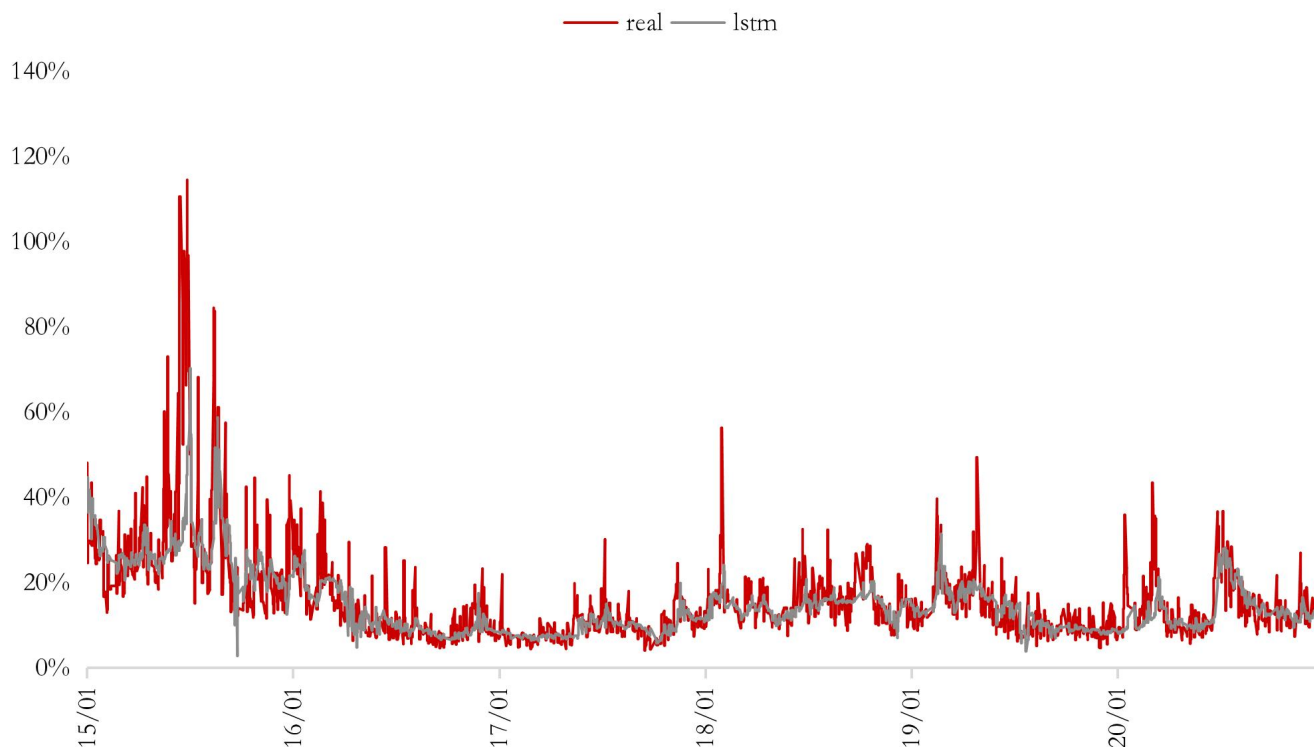
图表 21：LSTM 模型回测结果（5 日）

	LSTM	线性回归 (Lasso)	效果提升
MAE	0.0455	0.0527	13.64%
MSE	0.0067	0.0074	8.95%
RMSE	0.0818	0.0858	4.58%
胜率	0.7073	0.6306	12.15%

资料来源：东证衍生品研究院

本部分使用回看周期为 120 天的价量数据作为 LSTM 模型的输入数据，得到的结果显著好于线性回归模型。以表现最好的线性模型 Lasso 回归作为参考，在各个评价指标下 LSTM 均明显好于 Lasso 回归模型。胜率方面，周度胜率达到了 70.73%，具有较强的预测能力。

图表 22: LSTM 模型回测表现



资料来源：东证衍生品研究院

图表 23: LSTM 模型回测结果 (1 日)

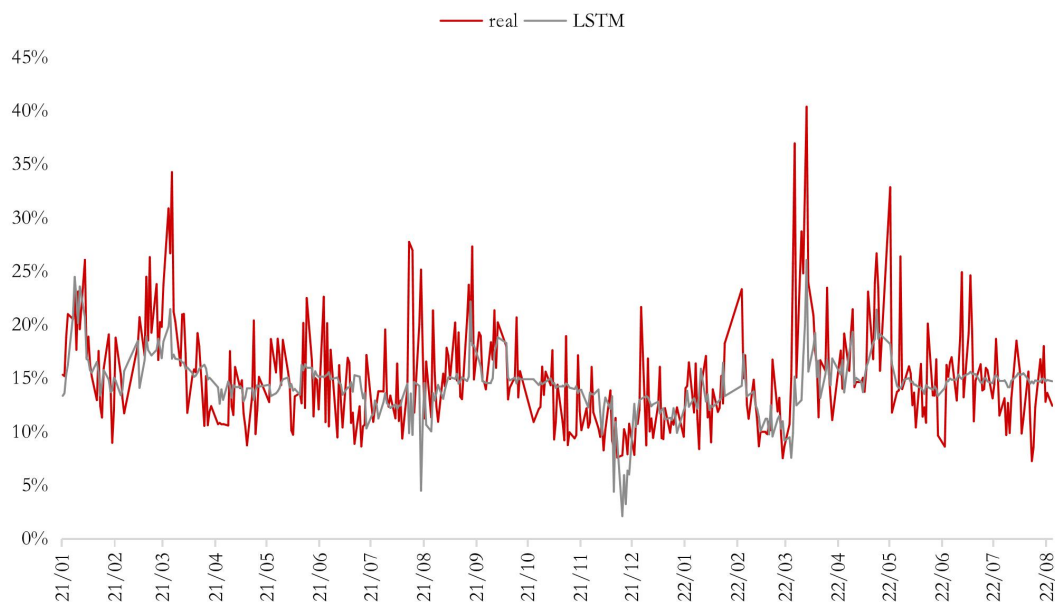
	LSTM	线性回归 (Lasso)	效果提升
MAE	0.0433	0.0413	-5.06%
MSE	0.0062	0.0044	-40.89%
RMSE	0.0787	0.0663	-18.70%
胜率	0.6708	0.6505	3.12%

资料来源：东证衍生品研究院

在日度级别的预测上，LSTM 准确率出现一定程度提升，然而却依然小幅由于线性回归下的日度预测，这一现象与前文的观察一致。

6.3、样本外

图表 24: LSTM 模型样本外表现



资料来源：东证衍生品研究院

图表 25: LSTM 模型样本外回测结果 (5 日)

	LSTM_样本外
mae	0.0397
mse	0.0030
rmse	0.0552
胜率	0.6770

资料来源：东证衍生品研究院

根据回测结果发现，在样本外 LSTM 的表现较其样本内表现略低，原因在于此类机器学习模型可解释性较差，一旦切换训练样本，其表现难免出现一定程度下滑。根据对比可得，模型在样本外表现的正确率下滑约 5% 左右，在可以接受的范围内，故认为模型在样本外依旧有效。

图表 26: LSTM 模型样本外回测结果 (1 日)

	LSTM_样本外
mae	0.0346
mse	0.0022
rmse	0.0469
胜率	0.7098

资料来源：东证衍生品研究院

在样本外，LSTM 的日度预测相较于周度预测能力表现出一定提升。误差指标进一步减小，而胜率升高，表明在 LSTM 样本内，模型对于近期波动率的预测能力依然好于远期波动率。

7、舆情分析

7.1、模型介绍

舆情数据能够反应市场上投资者的情绪，不难想见，当市场出现较大幅度的波动时，市场的舆情数据必然会发酵，具体体现在舆情数据数量的增多，以及情感指标的分化。当然，相关性不能代表因果性，故在这部分，将舆情数据作为模型的数据，考察舆情数据对波动率的预测能力。

本报告通过某财经网站的股民评论作为舆情数据，并且采用 SnowNLP 库对舆情数据进行情感分析。

本报告舆情数据选取股吧对于上证 50 指数的评论，舆情数据包括了帖子阅读数量（read），评论数量（mark）以及帖子标题（title）。通过对评论内容进行情感分析，可以获取每一条评论的积极程度\消极程度。

7.2、数据

图表 27：舆情数据举例

	read	mark	title
1	27	0	早就成为一个笑话了，信心满满[献花][摊手][赞]
2	640	2	大盘股掩护，反弹结束，指数剑指两千五
3	88	3	起风了，各位，我先上了
4	60	1	好牛逼啊，继续拉啊，怎么不拉了
5	55	3	4000 点不是梦
6	33	0	2400 ↓ ↓ 。 。 。 。 ↓ ↓ 。 。 。 。
7	43	0	一说就大跌没一次不是的，都在用脚投票，该醒了
8	69	1	人为拉指数，就怕恰得其反
9	99	2	反手做多就在今天！中午收盘或下午收盘的时候。
10	48	0	现在散户聪明了，不会上当了！
11	110	3	中字头不拉还好，拉一下，个股都开始跌了。
12	46	0	要拉指数，不拉更惨
13	1357	2	11.22 走势预测
14	222	2	老乡别走啊，牛来了
...			

资料来源：东证衍生品研究院

图表 28: 舆情数据情感评分

内容	评分
早就成为一个笑话了，信心满满[献花][摊手][赞]	0.79
大盘股掩护，反弹结束，指数剑指两千五	0.43
起风了，各位，我先上了	0.34
好牛逼啊，继续拉啊，怎么不拉了	0.66
4000点不是梦	0.89
2400↓↓。。。↓↓。。。	0.50
一说就大跌没一次不是的，都在用脚投票，该醒了	0.02
人为拉指数，就怕恰得其反	0.18
反手做多就在今天！中午收盘或下午收盘的时候。	0.70
现在散户聪明了，不会上当了！	0.62
中字头不拉还好，拉一下，个股都开始跌了。	0.17
要拉指数，不拉更惨	0.92
11.22 走势预测	0.87
老乡别走啊，牛来了	0.22

资料来源：东证衍生品研究院

上表展示了针对部分内容的情感评分，该平价模型通过对评论进行分词，再根据内置的训练集对评论内容进行打分，评分范围从0到1：0代表最为消极的情绪，而1则代表最为积极的情绪。举个例子，若评论当中大量出现“涨”、“信心”等词汇，该评论的情感评分就越趋近于1，若评论当中大量出现“跌”、“惨”、“做空”等词汇，则该评论情感评分越接近于0。需要注意的是，模型内置的训练集并不是针对金融市场的，故针对一些金融市场专有的“俚语”可能出现识别误差，比如股民们经常通过调侃自己“吃面”，“吃肉”来说明自己的盈利或亏损情况，而对于模型而言，所谓的“吃面”或是“吃肉”只是吃一种食物的状态，无法识别出其深层含义，不过，这样的问题可以通过更新训练集来进行优化。

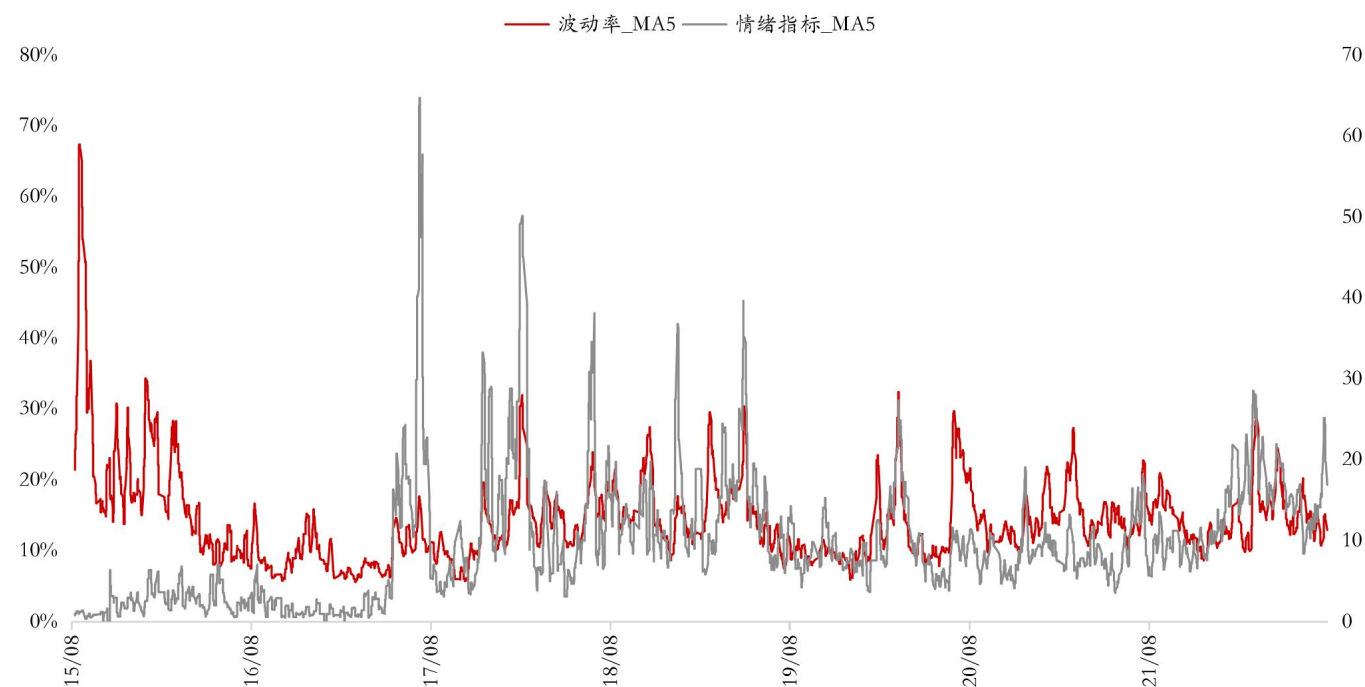
图表 29：輿情数据梳理

	发帖数量	阅读数量	评论数量	总情感得分	平均阅读数	平均评论数	平均情感得分
2022/11/8	35	53338	223	17.21	1523.94	6.37	0.49
2022/11/9	33	31396	198	15.07	951.39	6.00	0.46
2022/11/10	45	22782	122	22.74	506.27	2.71	0.51
2022/11/11	27	13287	117	15.71	492.11	4.33	0.58
2022/11/12	7	18379	122	4.36	2625.57	17.43	0.62
2022/11/13	10	14698	120	5.63	1469.80	12.00	0.56
2022/11/14	29	25287	113	14.10	871.97	3.90	0.49
2022/11/15	26	17268	182	14.54	664.15	7.00	0.56
2022/11/16	24	6212	38	12.88	258.83	1.58	0.54
2022/11/17	19	9247	72	10.51	486.68	3.79	0.55

资料来源：东证衍生品研究院

根据每一个帖子的发帖时间，可以计算出每一天总的发帖量、评论量、阅读量、平均情感得分等数据。下面基于假设，考察市场波动率数据与情绪指标之间的关系，此处情绪指标定义为：每日所有发帖内容的情绪评分求和。

图表 30: 波动率 VS 情绪指标 (MA5)



资料来源：东证衍生品研究院

从图中可以明显看出，除了在2017年之前由于发帖量较少导致情绪指标数值偏低，在2017年之后情绪指标和市场波动率呈现出较为明显的正相关性，尤其是在市场波动较为剧烈的行情当中，情绪指标的波动相较于标的波动率更为强烈，这一点也是符合我们预期的：在市场出现较大波动的情况下，往往投资者的情绪会随之产生更为强烈的波动，也就是所谓“过度反应（over-reaction）”，正是由于这样的原因，会造成部分不理智的投资者出现“追涨杀跌”的操作。而一旦可以通过模型，利用投资者的情绪反应对后市的波动率作出先一步判断，则可以在衍生品市场上进行相应获利。

为获得较好的预测能力，基于舆情数据共数理7个情绪相关因子，分别是：发帖数量、阅读数量、评论数量、总情绪评分、平均阅读量、平均评论量、平均情绪评分。

7.3、结果分析

图表 31：舆情数据回归结果

	10天	20天	60天	120天	250天
mae	0.0570	0.0554	0.0571	0.0592	0.0686
mse	0.0098	0.0088	0.0087	0.0090	0.0112
rmse	0.0988	0.0937	0.0932	0.0948	0.1060
胜率	0.6283	0.6338	0.6407	0.6345	0.6085

资料来源：东证衍生品研究院

在不同回看周期下，舆情数据的表现并无显著差异，以回看周期为 120 天为例，其周度预测正确率略高于 63%，该表现相比价量数据的预测能力并无明显提升。但不难发现，其预测正确率远高于 50%，表明舆情数据本身具有一定的预测能力。很容易想到，可以通过考察舆情数据和价量数据的相关性，若相关性较低，则可以将舆情数据作为价量数据的补充加入到模型中以提升模型整体的预测能力，在下一部分（第八节），尝试该操作。

图表 32：舆情数据样本外回归结果

	10 天	20 天	60 天	120 天	250 天
mae	0.0482	0.0414	0.0406	0.0409	0.0403
mse	0.0045	0.0032	0.0030	0.0031	0.0029
rmse	0.0674	0.0568	0.0551	0.0557	0.0541
胜率	0.6088	0.6917	0.6606	0.6813	0.6865

资料来源：东证衍生品研究院

由于针对舆情数据同样采用前述的 Lasso 模型进行多元回归，故由于样本外数据较为平稳，则其样本外表现反而略好样本内。在回看周期为 120 天和 250 天时，样本外的预测正确率均能达到 70% 以上。此外，舆情数据的丰富程度受时间关系影响较大，譬如在样本内的 2015 年，由于仍处在该财经网站早期阶段，其评论数量偏少，而在样本外的 2021 年，评论数量已经达到较高数量。数据完备性的提升也是造成模型样本外表现较好的原因之一。

8、舆情数据数据扩充

下图绘制了舆情数据和价量数据的相关性图。通过考量，发现舆情数据与价量数据之间的相关性偏低，大部分价量数据相对于舆情数据的相关性小于 0.3，考虑到舆情数据自身拥有与价量数据相同的预测能力，不难联想到可以讲两组数据进行结合，以舆情

数据作为价量数据的扩充以提升模型整体的预测能力。

除了发现舆情数据与价量数据整体相关性较低之外，还不难发现无论是舆情数据还是价量数据，其内部均存在一定数量的高相关性因子。在舆情数据中，“count”（评论数量）和“sentiment_sum”（总情绪指标）存在着高达0.99的相关性；而在价量数据当中，部分因子间的相关性也达到了0.99以上，表明数据当中难免存在多重共线性的风险。这也解释了在多元回归模型当中，Lasso回归的表现要好于OLS和Ridge，Lasso通过产出稀疏矩阵的方式将部分冗余变量的稀疏压缩到0，以此方式来避免数据多重共线性所造成的影响。

接下来分别在多元回归模型中（考虑到多重共线性的原因，此处采用Lasso回归），以及在LSTM模型中，将舆情数据作为价量数据的补充输入进模型，以考察能够提升预测效果。

图表 33：因子相关性

iopv	1	0.99	0.99	0.99	-0.12	0.99	0.18	0.99	0.99	0.3	0	0.03	0.99	0.97	0.92	0.01	0.01	-0.05	0.27	0.03	0.27	-0.09
high	0.99	1	1	1	-0.17	1	0.16	1	1	0.28	-0.01	-0.01	1	0.98	0.94	-0.01	-0.03	-0.09	0.28	0.02	0.28	-0.08
limit_down	0.99	1	1	1	-0.19	1	0.14	1	1	0.26	-0.03	-0.05	1	0.98	0.94	-0.04	-0.04	-0.1	0.27	0.01	0.27	-0.08
limit_up	0.99	1	1	1	-0.19	1	0.14	1	1	0.26	-0.03	-0.05	1	0.98	0.94	-0.04	-0.04	-0.1	0.27	0.01	0.27	-0.08
num_trades	-0.12	-0.17	-0.19	-0.19	1	-0.19	0.73	-0.19	-0.21	0.69	0.72	0.04	-0.19	-0.17	-0.2	0.75	0.68	0.59	0.12	-0.03	0.12	0.01
open	0.99	1	1	1	-0.19	1	0.14	1	1	0.26	-0.04	-0.04	1	0.98	0.94	-0.04	-0.05	-0.1	0.27	0.01	0.27	-0.08
volume	0.18	0.16	0.14	0.14	0.73	0.14	1	0.14	0.11	0.99	0.76	0.05	0.14	0.16	0.13	0.73	0.61	0.48	0.28	-0.04	0.28	0.04
close	0.99	1	1	1	-0.19	1	0.14	1	1	0.26	-0.04	0.03	1	0.98	0.93	-0.04	-0.05	-0.1	0.27	0.02	0.27	-0.08
low	0.99	1	1	1	-0.21	1	0.11	1	1	0.23	-0.07	-0.01	1	0.98	0.93	-0.07	-0.07	-0.12	0.26	0.02	0.26	-0.09
total_turnover	0.3	0.28	0.26	0.26	0.69	0.26	0.99	0.26	0.23	1	0.71	0.05	0.26	0.27	0.23	0.69	0.58	0.44	0.28	-0.03	0.28	0.02
vol	0	-0.01	-0.03	-0.03	0.72	-0.04	0.76	-0.04	-0.07	0.71	1	-0.01	-0.03	-0.01	-0.02	0.89	0.73	0.62	0.33	-0.1	0.33	0.12
return	0.03	-0.01	-0.05	-0.05	0.04	-0.04	0.05	0.03	-0.01	0.05	-0.01	1	-0.04	-0.05	-0.05	0.01	-0.03	-0.04	0	0.18	0	-0.02
ma5	0.99	1	1	1	-0.19	1	0.14	1	1	0.26	-0.03	-0.04	1	0.99	0.94	-0.03	-0.04	-0.1	0.27	-0.01	0.27	-0.08
ma20	0.97	0.98	0.98	0.98	-0.17	0.98	0.16	0.98	0.98	0.27	-0.01	-0.05	0.99	1	0.97	-0.01	-0.02	-0.07	0.26	-0.06	0.26	-0.07
ma60	0.92	0.94	0.94	0.94	-0.2	0.94	0.13	0.93	0.93	0.23	-0.02	-0.05	0.94	0.97	1	-0.02	-0.01	-0.02	0.26	-0.04	0.26	-0.05
vol_ma5	0.01	-0.01	-0.04	-0.04	0.75	-0.04	0.73	-0.04	-0.07	0.69	0.89	0.01	-0.03	-0.01	-0.02	1	0.86	0.71	0.37	-0.12	0.37	0.02
vol_ma20	0.01	-0.03	-0.04	-0.04	0.68	-0.05	0.61	-0.05	-0.07	0.58	0.73	-0.03	-0.04	-0.02	-0.01	0.86	1	0.86	0.38	-0.16	0.38	0.04
vol_ma60	-0.05	-0.09	-0.1	-0.1	0.59	-0.1	0.48	-0.1	-0.12	0.44	0.62	-0.04	-0.1	-0.07	-0.02	0.71	0.86	1	0.35	-0.15	0.35	0.06
corr	0.27	0.28	0.27	0.27	0.12	0.27	0.28	0.27	0.26	0.28	0.33	0	0.27	0.26	0.26	0.37	0.38	0.35	1	-0.08	1	-0.06
skew	0.03	0.02	0.01	0.01	-0.03	0.01	-0.04	0.02	0.02	-0.03	-0.1	0.18	-0.01	-0.06	-0.04	-0.12	-0.16	-0.15	-0.08	1	-0.08	-0.04
cov	0.27	0.28	0.27	0.27	0.12	0.27	0.28	0.27	0.26	0.28	0.33	0	0.27	0.26	0.26	0.37	0.38	0.35	1	-0.08	1	-0.06
kurt	-0.09	-0.08	-0.08	-0.08	0.01	-0.08	0.04	-0.08	-0.09	0.02	0.12	-0.02	-0.08	-0.07	-0.05	0.02	0.04	0.06	-0.06	-0.04	-0.06	1
	iopv	high	limit_down	limit_up	num_trades	open	volume	close	low	total_turnover	vol	return	ma5	ma20	ma60	vol_ma5	vol_ma20	vol_ma60	corr	skew	cov	kurt

资料来源：东证衍生品研究院

8.1、多元回归

图表 34：价量+舆情数据多元回归回测结果

	10 天	20 天	60 天	120 天	250 天
mae	0.0633	0.0615	0.0580	0.0530	0.0528
mse	0.0116	0.0100	0.0099	0.0080	0.0074
rmse	0.1077	0.1000	0.0993	0.0893	0.0859
胜率	0.5800	0.5889	0.6245	0.6395	0.6279

资料来源：东证衍生品研究院

相较于只利用价量数据进行回归，加入了舆情数据之后，整体模型的效果并没有显著提升。在回看周期较小时，模型表现略好于数据补充前，而当回看周期变长时，其表现与数据补充前基本持平。

8.2、多元回归样本外

图表 35：价量+舆情数据多元回归样本外回测结果

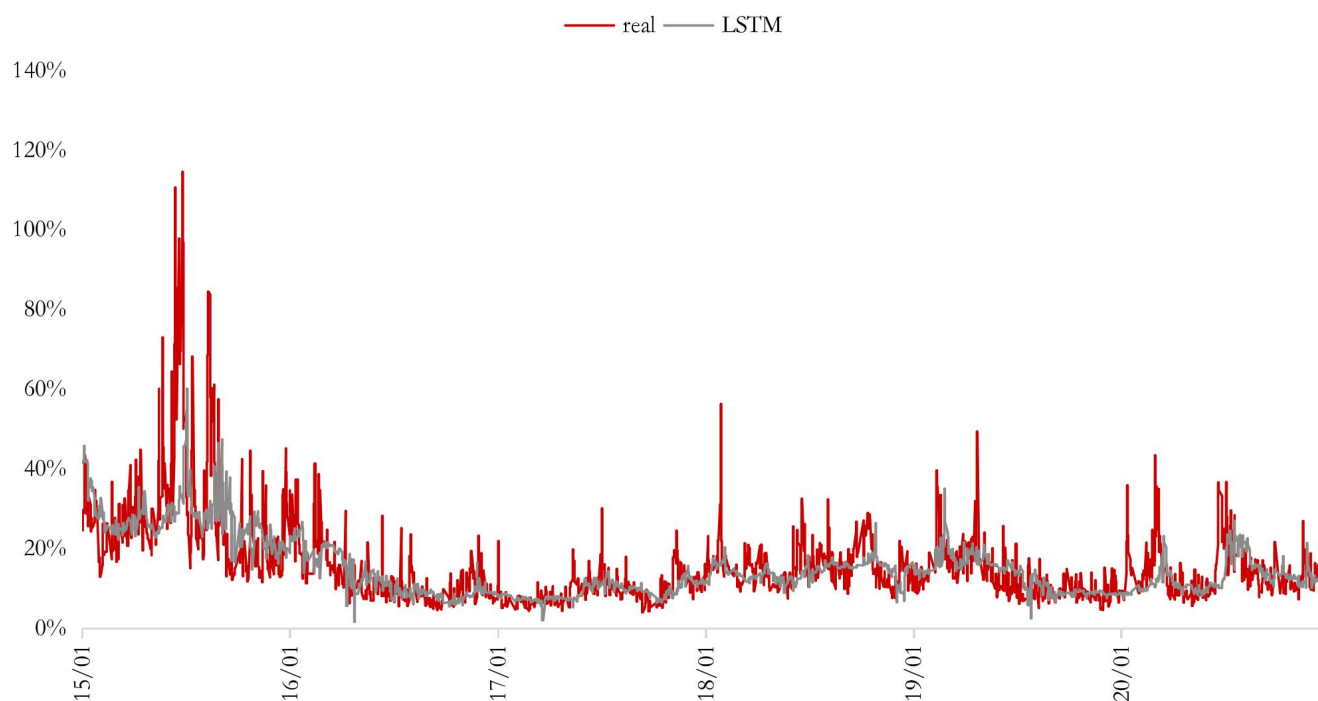
	10 天	20 天	60 天	120 天	250 天
mae	0.0522	0.0475	0.0425	0.0384	0.0408
mse	0.0049	0.0037	0.0032	0.0028	0.0031
rmse	0.0697	0.0607	0.0565	0.0526	0.0559
胜率	0.5736	0.6408	0.6667	0.7028	0.6848

资料来源：东证衍生品研究院

扩充数据后，模型的样本外表现并无出现衰减，且当回看周期为 120 天时模型达到最优效果，其周度正确率可以达到 70% 以上。可以认为在补充了舆情数据之后，模型的表现更为稳健。

8.3、LSTM

图表 36：价量+舆情数据 LSTM 表现



资料来源：东证衍生品研究院

图表 37：价量+舆情数据 LSTM 回测结果

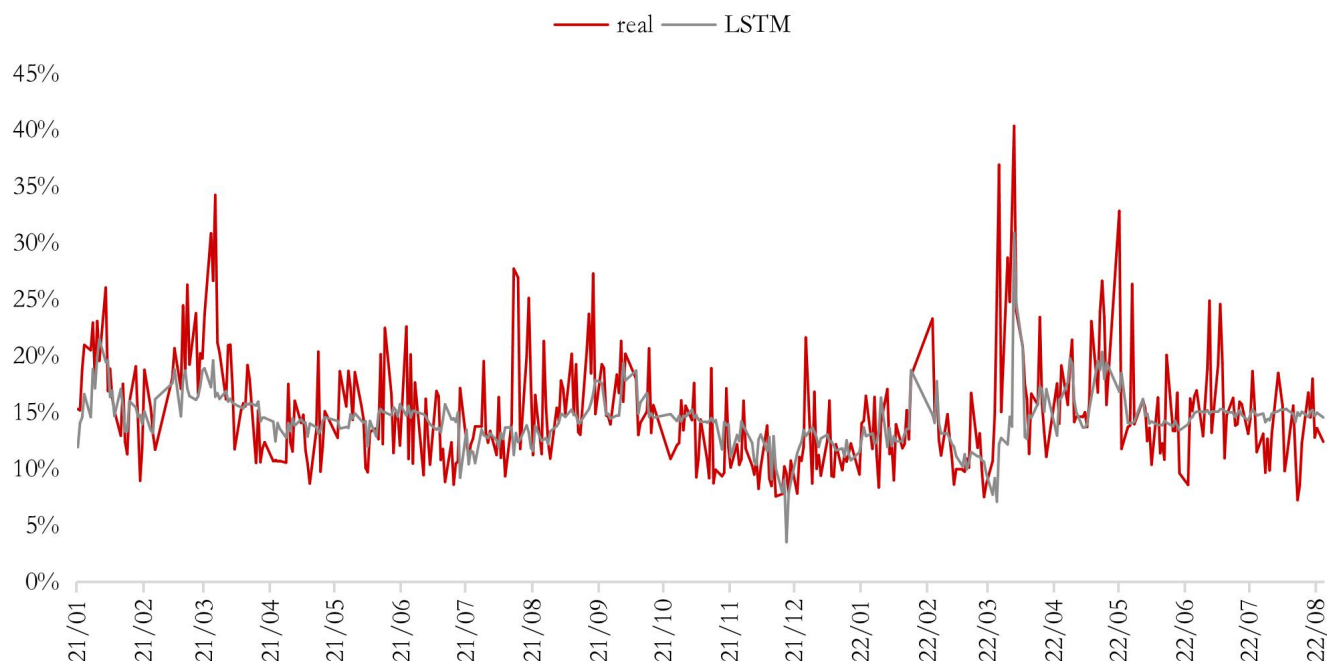
	LSTM
mae	0.0474
mse	0.0072
rmse	0.0848
胜率	0.6717

资料来源：东证衍生品研究院

相较于价量数据的 LSTM 模型，加入了舆情数据之后，模型整体表现出现约 4% 的下滑，但相较回归模型仍有较大幅度的提升。

8.4、LSTM 样本外

图表 38：价量+舆情数据 LSTM 样本外表现



资料来源：东证衍生品研究院

图表 39：价量+舆情数据 LSTM 回测结果

	LSTM_样本外
mae	0.0398
mse	0.0032
rmse	0.0566
胜率	0.6891

资料来源：东证衍生品研究院

对于样本外预测，加入舆情数据之后，模型的预测比起价量数据的 LSTM 模型的样本外表现有 2%左右的提升；此外，相较于其自身样本内表现也有 2.5%左右的提升，表明 LSTM 模型在加入舆情数据之后表现更为稳健。

9、总结

对本报告中各模型效果进行总结：价量数据的多元回归正确率为：63.06%；GARCH 模型正确率为：65.80%；价量数据的 LSTM 模型正确率为：70.73%；补充舆情数据后多元回归正确率为：63.95%；补充舆情数据后 LSTM 正确率为：67.17%。其中 LSTM 模型

正确较高，在相同回看周期下表现好于回归模型和 GRACH 模型；在多元回归模型中，由于数据方面存在一定程度的多重共线性，故 Lasso 模型的结果优于 OLS 和 Ridge；GARCH 模型随回看周期的缩短，其预测能力反而增强；舆情数据本身具有一定预测能力，但无法给价量数据的预测能力提供增量信息。

此外，根据样本内外的回测结果显示，各模型在样本外的表现并未出现明显衰减，且由于样本外波动相对平缓，甚至出现样本外表现优于样本内的情况。

同时，基于对不同预测周期的研究发现，在相同回看天数情况下，对较短周期预测的准确率要高于对较长周期的预测准确率。结论表明，各个模型对于远期波动率的预测均会随预测周期拉长而逐渐衰减。

最后回到本报告的标题：“股指波动率预测：舆情分析、深度学习能否战胜传统计量模型？”。结论表明答案是肯定的，LSTM 模型的表现样本内明显好于多元回归和 GARCH 模型，且样本外并未出现明显衰减。另一方面，舆情数据对模型的提升是有限的，在样本内并无明显改善，而在样本外有一定提升，可以认为加入舆情数据的补充后模型更加稳健。

期货走势评级体系（以收盘价的变动幅度为判断标准）

走势评级	短期（1-3个月）	中期（3-6个月）	长期（6-12个月）
强烈看涨	上涨 15%以上	上涨 15%以上	上涨 15%以上
看涨	上涨 5-15%	上涨 5-15%	上涨 5-15%
震荡	振幅-5%-+5%	振幅-5%-+5%	振幅-5%-+5%
看跌	下跌 5-15%	下跌 5-15%	下跌 5-15%
强烈看跌	下跌 15%以上	下跌 15%以上	下跌 15%以上

上海东证期货有限公司

上海东证期货有限公司成立于2008年，是一家经中国证券监督管理委员会批准的经营期货业务的综合性公司。东证期货是东方证券股份有限公司全资子公司，注册资本金38亿元人民币，员工逾800人。公司主要从事商品期货经纪、金融期货经纪、期货投资咨询、资产管理、基金销售等业务，拥有上海期货交易所、大连商品交易所、郑州商品交易所、上海国际能源交易中心和广州期货交易所会员资格，是中国金融期货交易所全面结算会员。公司拥有东证润和资本管理有限公司，上海东祺投资管理有限公司和东证期货国际（新加坡）私人有限公司三家全资子公司。

东证期货以上海为总部所在地，在大连、长沙、北京、上海、郑州、太原、常州、广州、青岛、宁波、深圳、杭州、西安、厦门、成都、东营、天津、哈尔滨、重庆、苏州、南通、泉州、汕头、沈阳、无锡、济南等地共设有36家分支机构，并在北京、上海、广州、深圳多个经济发达地区拥有149个证券IB分支网点，未来东证期货将形成立足上海、辐射全国的经营网络。

自成立以来，东证期货秉承稳健经营、创新发展的宗旨，坚持以金融科技助力衍生品发展为主线，通过大数据、云计算、人工智能、区块链等金融科技手段打造研究和技术两大核心竞争力，坚持市场化、国际化、集团化发展方向，朝着建设一流衍生品服务商的目标继续前行。

免责声明

本报告由上海东证期货有限公司（以下简称“本公司”）制作及发布。

本公司已取得期货投资咨询业务资格，投资咨询业务资格：证监许可【2011】1454号。

本研究报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本研究报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的报告之外，绝大多数研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买投资标的的邀请或向人作出邀请。

在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者需自行承担风险。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，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，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

如需引用、刊发或转载本报告，需注明出处为东证衍生品研究院，且不得对本报告进行任何有悖原意的引用、删节和修改。

东证衍生品研究院

地址：上海市中山南路318号东方国际金融广场2号楼21楼

联系人：梁爽

电话：8621-63325888-1592

传真：8621-33315862

网址：www.orientfutures.com

Email：research@orientfutures.com