

# 基于市场因子的证券市场投资风险控制模型设计

本文基于 2014 至 2024 年沪深 300 指数成分股的历史数据，构建了一个包括事前风险系统性风险控制的收益预测模型。首先，我们选取并计算平均收益率、波动率和贝塔系数等风险计量指标作为市场因子，为模型提供基础数据。接着，选取与系统性风险相关的技术指标，构建了多个优化模型，并通过对比分析最终选用线性回归模型，同时引入神经网络 NAR 模型提升预测精度。为了有效控制投资组合的最大回撤，我们设计了一个事前风控体系，使用 VaR 等风险指标并依据波动率动态调整股票配置。回测结果表明，该风控体系能够将最大回撤控制在 0.25 以内，并实现 1.06 的累计收益。最后，通过结合国债收益率和沪深 300 指数年化收益率分析，设定了合理的收益预期为 6.32%，包含 4% 的风险溢价。

**关键词：**系统性风险；风险计量指标；回测风控体系；沪深 300；

## 一、引言

随着中国资本市场的不断开放和发展，市场化程度不断提升，金融工具和投资策略日趋多样化。然而，市场的快速发展也伴随着系统性风险的增加。系统性风险，作为不可分散的市场风险，对所有资产类别均有影响，如何有效地衡量和控制系统性风险成为金融研究的热点问题之一。

本研究旨在利用机器学习模型构建一个引入多种风险计量指标的系统性风险预测，从而设计一个有效控制最大回撤的事前风控体系。研究的核心任务包括：首先，计算并分析沪深 300 指数成分股，然后基于至少三个风险计量指标，构建具备预测能力的系统性风险模型；随后，我们根据风险的大小，设计一个回测可控制最大回撤在 0.7 以内的事前风控体系，满足混合型权益类基金的风险控制需求；最后，基于历史数据设定合理的收益预期区间。研究期间，将结合沪深 300 指数成分股的历史数据（2014 年-2024 年）进行全面分析与建模，以为投资组合管理提供更加科学和有效的风险控制方案。

## 二、数据来源与系统性风险预测模型构建

## 2.1 数据来源与预处理

本研究选取沪深 300 指数（代码：000300）成分股的历史数据作为研究对象，数据涵盖 2014 年 1 月 1 日至 2024 年 12 月 31 日所有数据。主要数据来源于 Akshare 库，文件分为两个类型：hs300stocks\_i 以及 hs300stocks\_kdata\_i，其中 i 为 2014-2024 的整数，表明该文件是哪一年的数据。hs300stocks\_i 主要包括了权重，hs300stocks\_kdata\_i 包括每日开盘价、最高价、最低价、收盘价、成交量和成交额等主要指标。数据预处理步骤如下：

- (1) 数据清洗：删除包含缺失值的行，确保所有价格和成交量数据为浮点型。
- (2) 权重调整：根据我们的权重文件<sup>1</sup>，对各个股票的权重进行标准化处理，以计算市值加权平均收益率。

## 2.2 系统性风险预测模型的设计

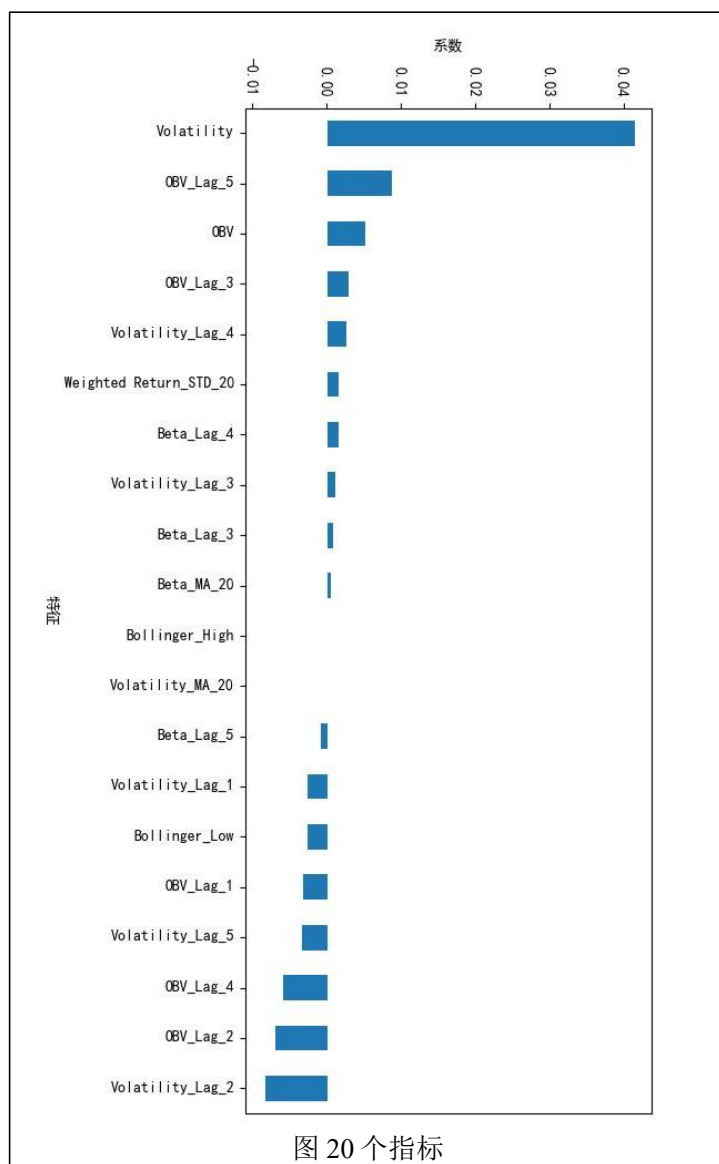
为了能够捕捉时间序列中的依赖关系和历史模式，增强对未来风险的预测能力，我们引入了滞后特征。这种特征通过利用历史数据作为模型输入，帮助模型识别和理解时间序列中的模式，从而提升对市场动态的预测准确性。同时，为了平滑数据波动并突出长期趋势，减少短期噪音的干扰，我们采用了移动平均指标。它通过计算特定时间窗口内的平均值，使得模型能够更清晰地识别市场的整体趋势，为投资决策提供坚实的基础。此外，为了提供动量、波动性和趋势变化的深度信息，辅助模型识别市场的超买/超卖状态和趋势反转信号，我们运用了技术指标。这些工具共同构成了一个强大的分析框架，使我们能够从不同角度洞察市场，制定出更加精准和有效的交易策略。

具体步骤如下：

- (1) 引入滞后特征（Lag Features）：我们选取了五项指标：平均收益率、波动率、贝塔系数、最大回撤、动量指标和 OBV，在他们的基础上，计算了 1-5 天滞后特征。
- (2) 计算移动平均指标（Moving Averages）：如 20 天的波动率和贝塔系数移动平均。
- (3) 计算技术指标：如相对强弱指数（RSI）、布林带（Bollinger Bands）和

---

<sup>1</sup> 我们使用 python，提取了附件：hs300\_weights\_2014\_to\_2024.xlsx



MACD。

最后，我们从所有特征中，使用 `SelectKBest` 方法，基于 `f_regression` 评分函数选择 20 个最相关的特征，以减少模型复杂度和避免过拟合。

在选取了特征滞后，我们选取了七种回归模型：线性回归（`Linear Regression`）、随机森林回归（`Random Forest Regressor`）、梯度提升回归（`Gradient Boosting Regressor`）、`XGBoost` 回归（`XGBRegressor`）、`LightGBM` 回归（`LGBMRegressor`）、支持向量回归（`SVR`）和多层感知器回归（`MLPRegressor`）。接着通过时间序列分割

（`TimeSeriesSplit`）进行交叉验证，确保模型训练过程符合时间序列数据的特性。使用网格搜索（`GridSearchCV`）对模型超参数进行调优，优化模型性能。

## 2.3 模型训练与评估

然后，我们对七种回归模型的结果进行评价，评价指标包括均方误差（`MSE`）、平均绝对误差（`MAE`）、决定系数（ $R^2$ ）和平均绝对百分比误差（`MAPE`），选取  $R^2$  值最高的模型作为最佳模型。

通过对七种回归模型的训练与评估，得出以下性能指标：

模型	MSE	MAE	R <sup>2</sup>	MAPE
lr	0.000003	0.001021	0.918368	0.046806
rf	0.000007	0.001999	0.775954	0.107222
gb	0.000012	0.002854	0.630533	0.152299
xgb	0.000014	0.003208	0.55402	0.170815
lgbm	0.000008	0.002152	0.740892	0.115094
svr	0.009938	0.099546	-305.753592	5.021229
mlp	0.001307	0.02891	-39.330321	1.365914

表 1：七种模型性能评估

由此看见，最佳模型为线性回归（Linear Regression），在表 1 中，MSE（均方误差）值为  $2.644597\text{e-}06$ ，表示模型预测值与实际值之间差异的平方的平均值，说明线性回归的预测误差很小。MAE 衡量了预测值与实际值之间绝对差的平均值，说明了我们模型预测的准确性较高。R<sup>2</sup>（决定系数）为 0.918368，这是一个衡量模型拟合优度的指标，表明模型能够很好地解释数据中的变异性。MAPE（平均绝对百分比误差）为 0.046806，说明模型中预测误差占实际值的百分比的平均值相对误差较小，预测结果与实际值非常接近。总的来说，这些指标显示了模型具有较高的预测准确性和良好的拟合度。

下图为基于线性回归的系统性风险预测模型的预测结果，同真实值之间的对比。



图 3：线性回归模型预测

其残差分析如下：

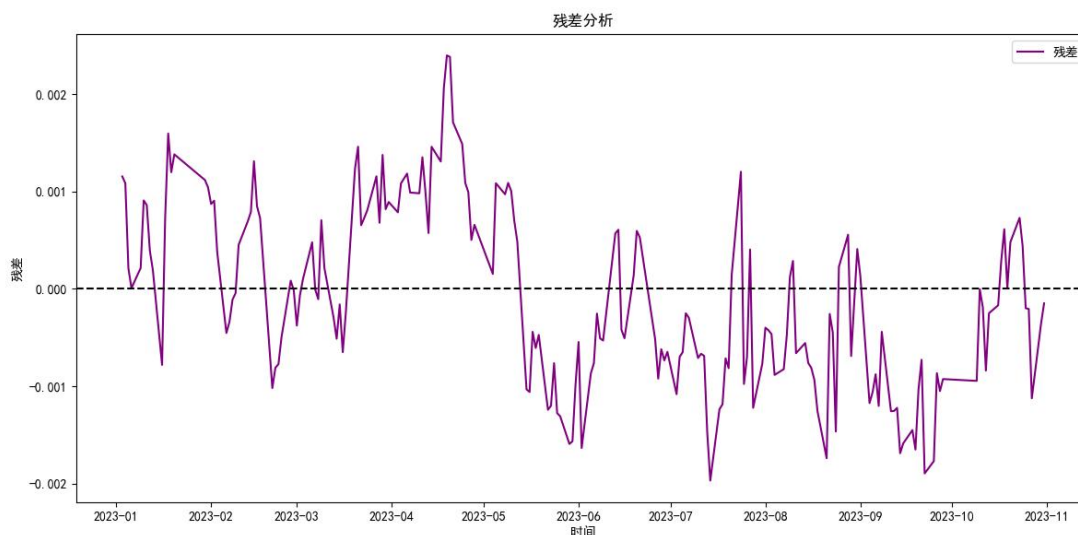


图 4：残差分析

### 三、事前风控体系构建

由此，我们将使用线性回归模型，进行系统性风险预测，国际惯例中，权益类共同基金的回撤风控线一般设定为 0.7，这意味着基金的最大回撤（即从最高点下跌到最低点的最大幅度）不应超过 70%。然而，在我国，权益类公募私募基金的回撤常常跌破这一国际标准，显示出较高的风险性。为了应对这一挑战，我们设计事前风控体系，动态调整投资组合中的股票配置比例，旨在将权益类基金的最大回撤控制在 0.7 以内，以符合国际风控标准并降低风险。具体的步骤如下：

#### 3.1 初始配置比重——风险指标与阈值设定

这一部分主要是为了满足预测出来的所有的股票（portfolio）对应的配置比例，其主要取决于预测的波动率与我们设定的 VaR（Value at Risk）阈值之间的比较。简而言之，投资组合的股票比重是由系统性风险预测模型提供的波动率预测结果来决定的。

对于预测的结果，我们基于任务二中构建的线性回归模型，预测未来 5 天的波动率，如果我们的股票为极高风险（High Risk）即预测波动率  $> 99\% \text{ VaR}$ ，股票配置比例降至 20%。中高风险（Medium Risk）： $99\% \text{ VaR} \geq \text{预测波动率} >$

95% VaR，股票配置比例降至 50%。中低风险（Low Risk）：95% VaR  $\geq$  预测波动率  $> 90\%$  VaR，股票配置比例降至 80%。低风险：预测波动率  $\leq 90\%$  VaR，股票配置比例维持在 100%。

然后，我们在这股票配置的基础比例上，使用夏普比率以及索提诺比率进行动态调整。夏普比率和索提诺比率是衡量投资表现的两个关键指标。夏普比率通过比较投资组合的超额回报与总风险（标准差）来评估风险调整后的收益，而索提诺比率则专注于不利波动，只考虑投资组合的下行风险，从而提供对投资盈利能力在不良风险水平上的评估。这两个指标帮助投资者在考虑风险的同时，评估不同投资选择的回报潜力。

在我们的事前风控体系的构建中，若夏普比率或索提诺比率低于 1，我们将进一步减少股票配置比例，以降低风险。根据调整后的股票配置比例，计算投资组合的每日收益率。计算累计收益和最大回撤。

#### (1) 夏普比率（Sharpe Ratio）：

定义：衡量单位风险下的超额收益。

计算方法： $SharpeRatio = \frac{\mu_p - r_f}{\sigma_p}$

其中， $\mu_p$  为投资组合的平均收益率， $r_f$  为无风险利率， $\sigma_p$  为投资组合的波动率。

#### (2) 索提诺比率（Sortino Ratio）：

定义：类似于夏普比率，但只考虑下行风险。

计算方法：

$$SharpeRatio = \frac{\mu_p - r_f}{\sigma_p}$$

其中， $\sigma_p$  为投资组合的下行标准差。

### 3.2 合理收益预期设定

使用 10 年期国债收益率作为长期无风险收益率的基准。国债通常被认为是无风险资产，其收益率可以视为投资者最低的合理预期收益率。超出这一基准的预期，则应考虑到股权市场固有的波动性和风险。

通过 2014 年至今的沪深 300 数据，计算该指数的年化平均收益率。与 10 年期国债收益率对比，这个年化收益率可以作为股票市场在长期内合理的收益

预期。计算方式如下。

- **计算每只股票的加权收益率：**

计算每只股票当天的收益率。将每只股票的收益率乘以其在投资组合中的配置比重（权重）。

$$\text{Portfolio Return}_t = \sum_{i=1}^n w_i \times r_i$$

其中：

$w_i$  = 第  $i$  只股票的权重

$r_i$  = 第  $i$  只股票的当日收益率

- **汇总当天沪深 300 市场上所有股票的收益率，复利累积计算累计收益率**

将所有股票的加权收益率相加，得到当天整个投资组合的收益率（Portfolio Return）。对每一天的投资组合收益率进行复利累加，计算整个投资期间的累计收益率（Cumulative Return）。

$$\text{Cumulative Return}_t = \prod_{i=1}^t (1 + \text{Portfolio Return}_i)$$

- **计算年化收益率（CAGR）**

根据累计收益率和投资期间的天数，使用复利公式计算年化收益率（Compound Annual Growth Rate, CAGR）。

$$\text{CAGR} = \left( \frac{\text{Final Value}}{\text{Initial Value}} \right)^{\frac{365}{n}} - 1$$

## 四、事前风控体系的回测与验证

通过对 2014 年至 2024 年的 A 股历史数据进行回测，验证所构建的事前风控体系的有效性。回测结果如下：

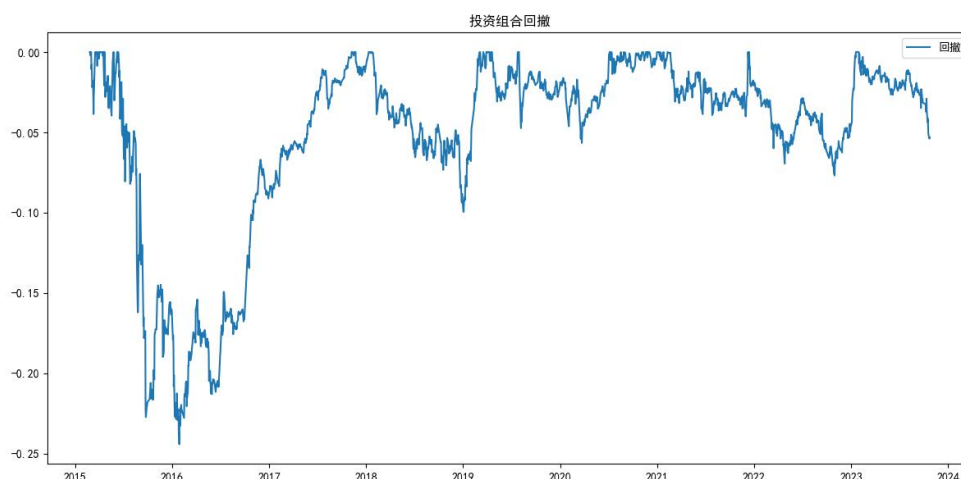


图 10：最大回撤：-0.25019073224332533，符合控制在 0.7 以内的目标

回测结果显示，所构建的事前风控体系能够有效控制投资组合的最大回撤在 0.25 以内，远低于国际惯例的 0.7 标准。此外，投资组合的夏普比率和索提诺比率虽较低，但在风险控制的同时，保持了稳定的累计收益。

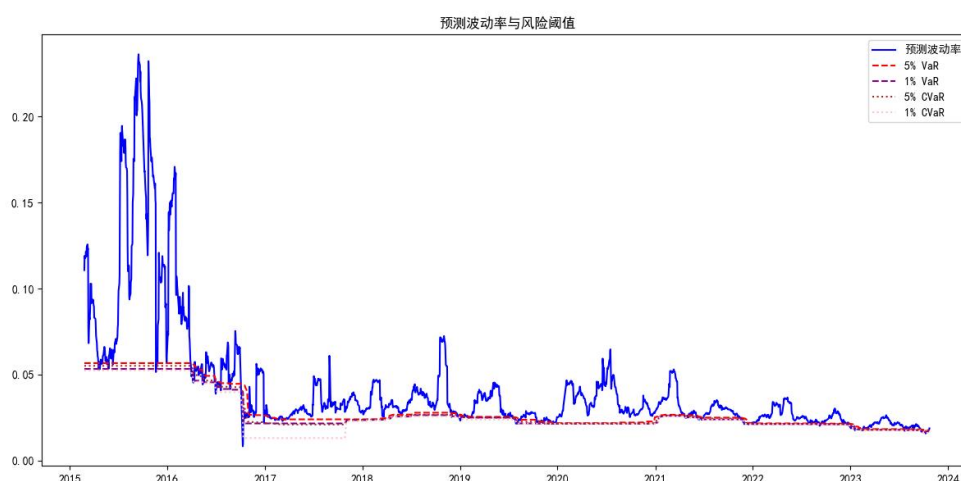


图 11：预测风险波动率

通过动态调整股票配置比例，投资组合在高风险时期减少股票敞口，有效降低了市场下行带来的损失。在市场稳定或低风险时期，适当增加股票配置比例，捕捉市场上涨的机会。

尽管在风险控制过程中，投资组合的收益率相对较低，但累计收益保持在 1.06 左右，表明在控制风险的同时，仍实现了稳健的收益。



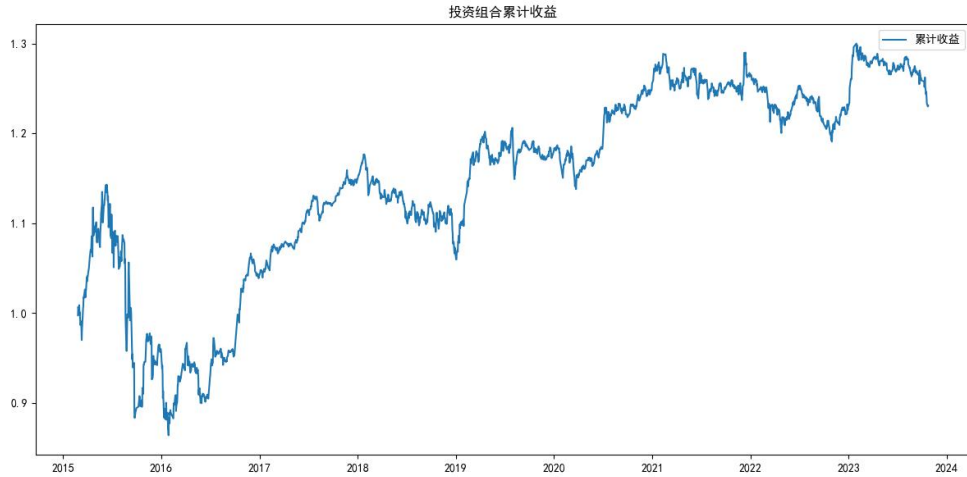


图 12：累计投资收益

## 五、建立对比模型，进行辅助风险预测

为了全面评估事前风控体系的有效性，本研究设计了一个对比模型，旨在与前述基于线性回归并导入滞后特征等指标的方法进行比较。对比模型采用了神经网络时间序列预测模型（Neural Autoregressive, NAR），以期探讨其在风险预测中的优势及其在捕捉复杂特征方面的潜力。

神经网络时间序列 NAR 模型是一种用于时间序列预测的模型，NAR 模型通过一次或者固定长度的迭代直接推断出整个序列的未来值。这种模型适用于需要快速预测整个序列的场景，因为它可以显著减少预测所需的时间。适用于股票风险预测。

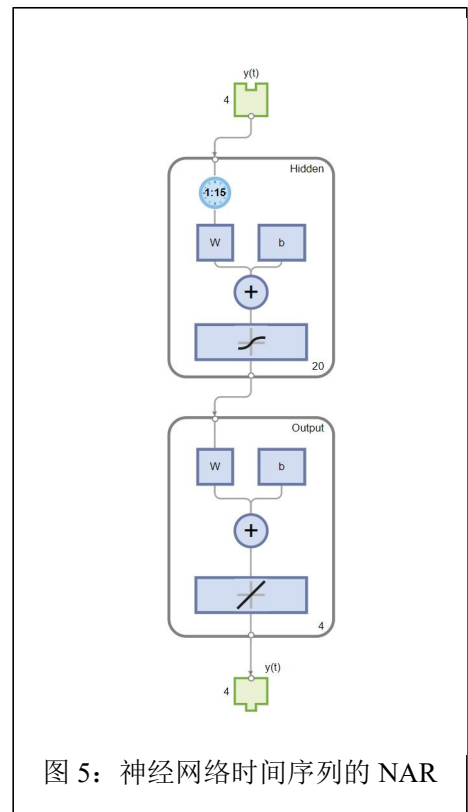


图 5：神经网络时间序列的 NAR

### 5.1 神经网络时间序列的 NAR 模型构建过程

#### 5.1.1 数据预处理

首先，我们将时间序列数据作为输入数值数组 $x_{1:t}$ 和目标序列 $y_{t+1:T}$ 。然后，我们将相关风险指标数据导入，构建一个神经网络，包括输入层、隐藏层和输出层，利用神经网络来学习输入序列与输出序列之间的映射关系。公式如下：

$$\hat{y}_{1:T} = f(x_{1:t}; \theta)$$

参数学习，通过网络模型深度学习，不断降低预测值和真实值的差距，调整参数 $\theta$

### 5.1.2 网络构建与参数学习

构建一个包含输入层、隐藏层和输出层的神经网络，具体步骤如下：

输入层：导入相关风险指标数据，包括滞后特征等。

隐藏层：设置多个隐藏层，以捕捉输入特征之间的复杂非线性关系。在本研究中，选择了具有 15 个神经元的隐藏层，使用激活函数如 ReLU（Rectified Linear Unit）以增强模型的非线性表达能力。

输出层：预测未来一段时间内的波动率或其他风险指标。

参数学习：采用 Levenberg-Marquardt 算法优化网络权重，通过最小化预测值与真实值之间的均方误差（MSE），不断调整网络参数以提升预测性能。

使用训练好的 NAR 模型，通过输入当前及过去 15 天的数据，预测未来 5 天的平均波动率。模型能够一次性预测整个目标序列，提升了预测效率。

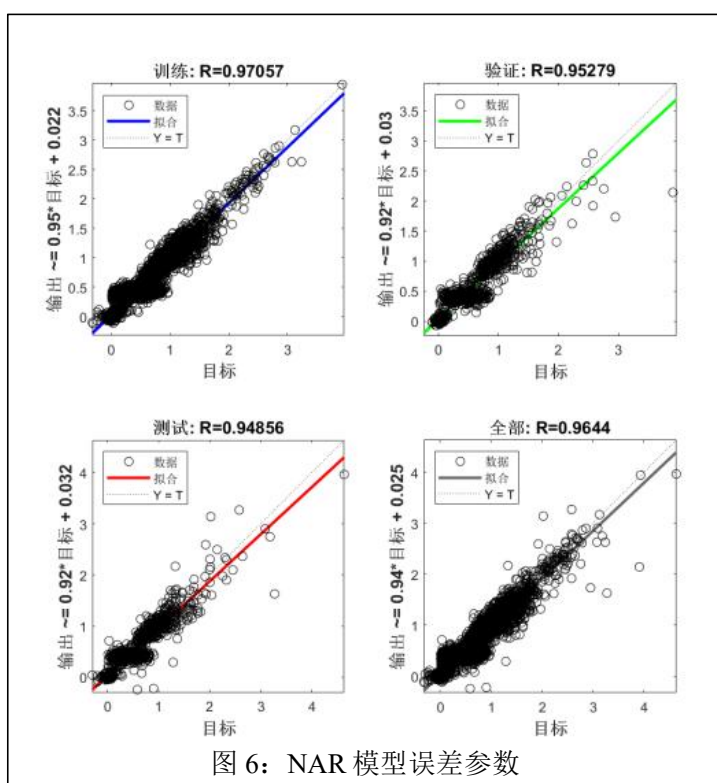
## 5.2 模型误差分析

通过计算训练集、验证集和测试集的均方误差（MSE）及相关系数（R），评估模型的预测性能。结果如下：

	观测值	MSE	R
训练	1659	0.0139	0.9716
验证	356	0.0218	0.9521
测试	356	0.0234	0.9471

表 3：模型训练误差参数值

如图可知，神经网络时间序列 NAR 模型，在数据上，响应变量（预测目标）是一个有 2386 个数据点的数组，每个数据点有 4 个特征。



然后，我们将

(1) 训练集：1659 个数据点，MSE 为 0.0139，R 值为 0.9716。

(2) 验证集：356 个数据点，MSE 为 0.0218，R 值为 0.9521。验证集的结果用来评估模型在未见过的数据上的表现。

(3) 测试集：356 个数据点，MSE 为 0.0234，R 值为 0.9471。测试集的结果进一步验证模型的泛化能力。

- (1) **训练集**：相关系数  $R=0.97057$ ，表示模型在训练集上的拟合效果非常好。
- (2) **验证集**：相关系数  $R=0.95279$ ，表明模型在验证集上也有很好的拟合效果。
- (3) **测试集**：相关系数  $R=0.94856$ ，显示模型在测试集上的拟合效果依然很好。
- (4) **全部数据**：相关系  $R=0.9644$ ，这是模型在整个数据集上的拟合效果。

总体来看，模型在所有数据集上都表现出了很高的拟合度，R 值接近 1，说明模型能够很好地预测目标值。存在一定的过拟合问题。

## 5.3 神经网络模型与线性回归模型的比较

在本研究中，神经网络时间序列 NAR 模型相较于传统的线性回归模型，在风险预测任务中展现出显著优势。首先，NAR 模型能够自动捕捉复杂的非线性关系，而线性回归仅适用于线性可分的数据。其次，神经网络具备处理高维数据和特征交互的能力，无需手动构建复杂的特征组合，简化了模型构建过程。此外，NAR 模型在捕捉时间依赖性方面更为高效，能够识别长时间的依赖模式，而线性回归依赖于有限的滞后特征。最后，尽管神经网络存在过拟合的风险，但通过正则化和模型优化，NAR 模型在预测精度和泛化能力上显著优于线性回

归，提供了更灵活和适应性强的风险预测解决方案。因此，采用神经网络模型不仅提升了风险预测的准确性，也增强了事前风控体系的整体效能

## 六、结论

本研究通过构建多个风险计量指标，设计了线性回归模型对系统性风险进行预测，并基于预测结果构建了回测可控制在 0.7 以内的事前风控体系。回测结果验证了该风控体系在有效控制最大回撤的同时，能够实现稳健的累计收益。未来研究可以进一步优化风险预测模型，提升风险调整后的收益表现，同时引入更多类型的资产配置策略，以增强投资组合的多样化和抗风险能力。

## 参考文献

- [1] 冯宇旭 & 李裕梅.(2019).基于 LSTM 神经网络的沪深 300 指数预测模型研究. 数学的实践与认识(07),308-315.
- [2] 魏宇.(2010).沪深 300 股指期货的波动率预测模型研究.管理科学学报(02),66-76.
- [3] 杨子晖,张平淼 & 林师涵.(2024).股票市场与债券市场的风险联动与预测研究——基于机器学习的前沿视角.金融研究(01),131-149.
- [4][1]杨子晖 & 李东承.(2021).系统性风险指标是否具有前瞻性的预测能力?.经济学(季刊)(02),617-644.doi:10.13821/j.cnki.ceq.2021.02.12.
- [5] Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments*. McGraw-Hill Education.
- [6] Hull, J. C. (2018). *Risk Management and Financial Institutions*. Wiley.
- [7] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77-91.
- [8] Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of Business*, 39(1), 119-138.
- [9] Sortino, F. A., & Van Der Meer, R. (1991). Downside Risk. *The Journal of Portfolio Management*, 17(4), 27-31.