工作汇报

1. 学习实证资产定价书本内容 1~5、7、8 章。
2. Tidy Finance，内容包括：金融数据使用、贝塔估计、单变量组合排序、规模排序 p 值黑客。
3. 主要阅读因子投资的文献：

(1) 机器学习与因子投资：Missing Data in Asset Pricing Panels, Missing values handling for Machine Learning protfolios, Empirical asset pricing via Machine Learning, Estimating stock market beta; Empirical Asset Pricing with Probability Forecasts.

(2) 动量效应：Factor Timing with Portfolio Characteristics, Principal portfolios, Cross-stock and Factor momentum.

(3) 其他文献：I-S O-S Sharpe ratios of multi-factor asset pricing models, Statistical Analysis with Missing Data, Open Source Cross-Sectional Asset Pricing, Managerial Risk-Taking Incentive and Firm Innovation, Lucky factors.

Empirical Asset Pricing with Probability Forecasts

Songrun He, Linying Lyu, and Guofu Zhou

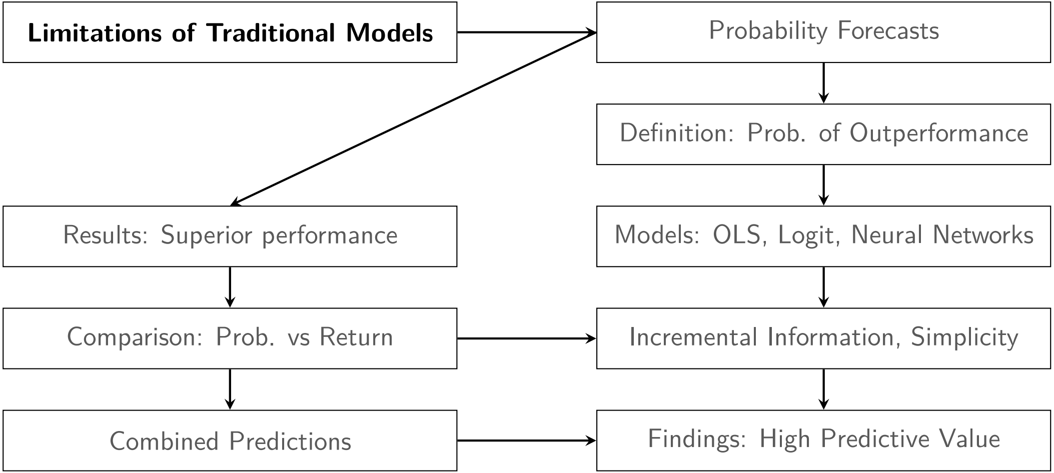
Olin Business School Center for Finance & Accounting Research Paper, 2024

Interpreter: Weihao Zhao

School of Economics and management WHU

2025-01-15

Framework


Limitations of Traditional Models
Probability Forecasts
Definition: Prob. of Outperformance
Models: OLS, Logit, Neural Networks
Incremental Information, Simplicity
Findings: High Predictive Value
Results: Superior performance
Comparison: Prob. vs Return
Combined Predictions

Motivation

Limitation of existing methods:

- Traditional models/Machine learning models mainly focus on return forecasts.
- Expected returns cannot fully reflect risk adjustment info of stocks in actual performance.

Why studying Probability forecasts

(1) Not only considering returns, but also capturing risk adjusted information.

- Providing richer investment information.
- Improving the signal-to-noise ratio in return forecasting problems.

(2) Sorting stocks by combining information ratios rather than just returns.

- Probabilistic sorting is practical for asset management.

Research Questions

Key Questions:

1. **Can probability forecasts provide accurate predictions/superb portfolio performance?**
   - Any incremental information beyond return forecasting?
2. **Will a simple probability model achieve performance similar to complex models?**
3. **Can combining return and probability forecasting methods enhances the results?**

Contributions

1. Literature on application of probability forecasts

Prior: Predominantly focused on time-series analysis.(Pesaran and Timmerman 2004)

- Focused on time-series sign forecasting of stock returns.(Papailias et al.2021)

Extend: Considers probability forecasts in cross-section of stock returns and the predictive power of charateristics, which provides a new aspect(Return + Volatity Forecasting).

2. Literature on Machine Learning in asset pricing

Prior: Employed autoencoders for factor modeling.(Gu et al. 2020)

- Proposed an E-LASSO method for equity risk premium forecasting.(Han, He, 2023)

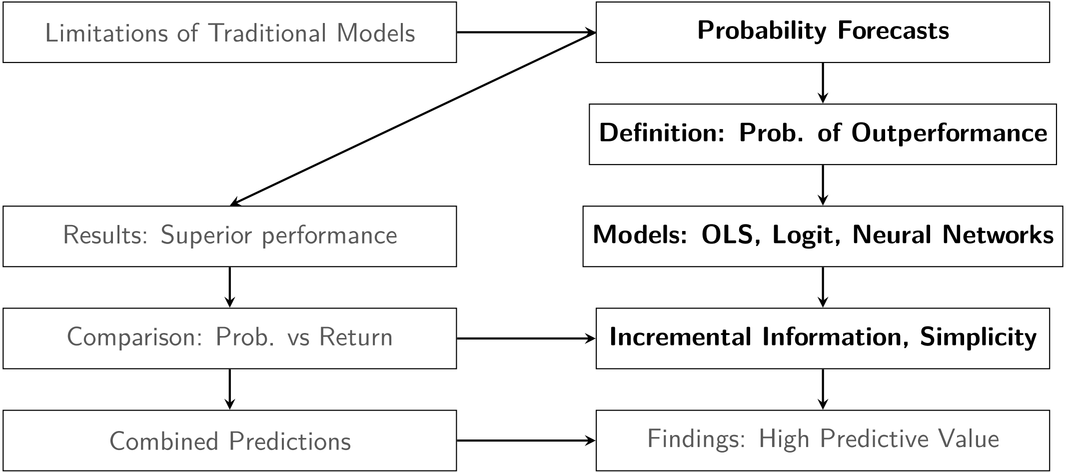
Extend: Pioneers the use of machine learning for probability forecasting in asset pricing.

- Leverage ML to predict probability of stock outperformance relative to a benchmark.
- Enhance the predictive accuracy and theoretical understanding of asset returns.

Hypothesis

1. Probability forecasts provide the distinctive information other than retun forecasts.
2. Simple probability forecasts models achieve performance similar to complex ones.
3. Return + Porbaility forecasts = better results than using each one alone.

Design


Limitations of Traditional Models
Probability Forecasts
Definition: Prob. of Outperformance
Models: OLS, Logit, Neural Networks
Incremental Information, Simplicity
Findings: High Predictive Value
Results: Superior performance
Comparison: Prob. vs Return
Combined Predictions

Probability forecasts- Definition

Information Ratio: $IR = \frac{\mathbb{E}[R_{i,t+1} - R_{mkt,t+1}]}{\sigma(R_{i,t+1} - R_{mkt,t+1}}$

Definition: Prob. of stock's return $R_{i,t+1}$ outperforming market return($R_{mkt,t+1}$)

$\text{Prob}_t \left(R_{i,t+1} - R_{mkt,t+1} > 0\right)$

$= 1 - \text{Prob}_t \left(R_{i,t+1} - R_{mkt,t+1} < 0\right)$

$= \Phi \left(\frac{\mu_i - \mu_{mkt}}{\sigma_{t+1|t}} \right)$

where $\Phi(\cdot)$ denotes the CDF of standard normal distribution $N(0, 1)$, and $\sigma_{t+1|t}$ represents the conditional volatility of excess return, varying with Ω_t .

Models 1. **MSE loss w. identity linking function**

The loss function is the standard for evaluating the predictive performance of a model.

MSE is suitable for predicting continuous variables;

Cross-Entropy Loss is suitable for probabilistic prediction.

$$\mathcal{L}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{i,t+1} - g(z_{i,t}; \theta) \right)$$

(1) **OLS**: $g(z_{i,t}; \theta) = z'_{i,t}\theta$

The final prediction may be outside the probability bounds of 0 and 1.

(2) **PLS**: $g(z_{i,t}; \theta) = (z'_{i,t}\Omega)'\theta$

Reduce overfitting through dimensionality reduction techniques.

Models 2. **Cross-Entropy Loss w. Logit linking function(sigmoid)**

$$CE\mathcal{L}(\theta) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t+1}\log(g(z_{i,t};\theta)) + (1-y_{i,t+1})\log(1-g(z_{i,t};\theta))\right)$$

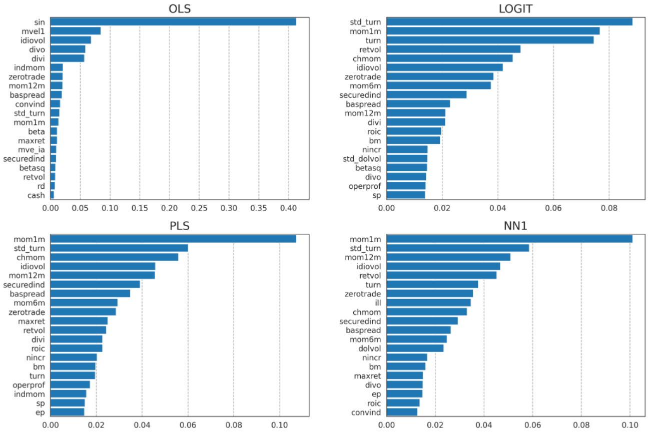
(1) **Logit Regression: Linear binary classification problem of output between [0, 1].**

$g(z_{i,t};\theta) = \frac{1}{1+e^{(-z'_{i,t}\theta)}}$, **For** $f(z_{i,t};\theta) = z'_{i,t}\theta$

(2) **Neural Networks: Nonlinear extension of basic logit regression.**

$g(z_{i,t};\theta) = \frac{1}{1+e^{-f(z_{i,t};\theta)}}$, **For** $f(z_{i,t};\theta) = W_3 \times \sigma\left(W_2 \times \sigma\left(W_1 z_{i,t} + b_1\right) + b_2\right) + b_3$

Variable importance- distinctive info of probability forecasts



Compared with Gu et al. (2020), the volatility measures have larger weights.

Variance Adjusted Decile Portfolios

Original decile portfolio: Each decile contains stocks with similar exceedance probabilities.

The necessity: Ensuring volatility of portfolios with different deciles is comparable.

Variance adjustment: Volatility are crucial for probability that a stock outperforms.

Using 36-month rolling window to calculate real-time volatility of each decile.

$$\sigma_{j,t} = \sqrt{\frac{1}{36} \sum_{\tau=t-36}^{t-1} \left(r_{j,\tau} - \bar{r}_j\right)^2}, \quad \text{for } j = 1, \cdots, 10,$$

At each t, scaling each portfolio return to have same volatility as decile 1 portfolio.

$$r_{j,t}^{adj} = r_{j,t} \frac{\sigma_{j,t}}{\sigma_{1,t}}, \quad \text{for } j = 1, \cdots, 10.$$

Rationale behind the variance adjustment strategy

Variance adjusted long short portfolio is formed as $r_{10,t}^{adj} - r_{1,t}^{adj}$

$r_{1,t} = \beta_1' F_t + \varepsilon_{1,t}$, Excess return of decile 1 portfolio.

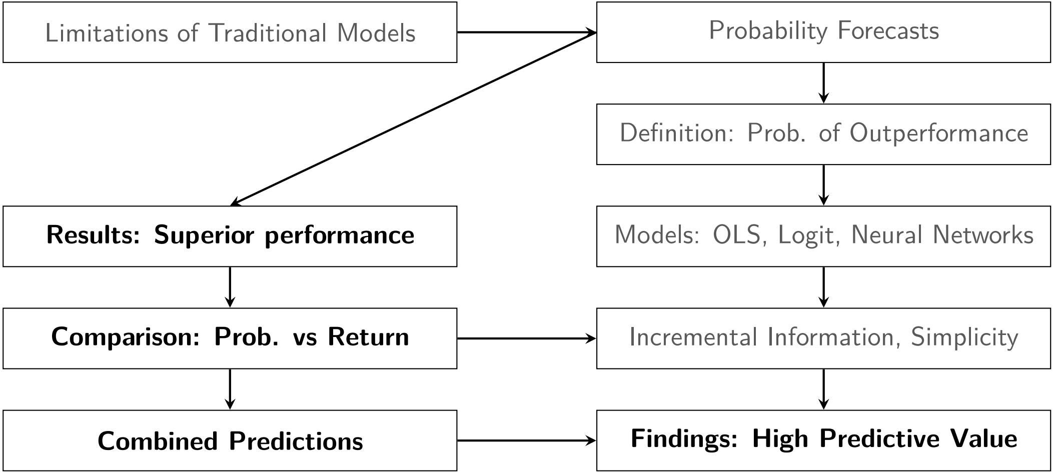
$r_{10,t} = \beta_{10}' F_t + \varepsilon_{10,t}.$, Excess return of decile 10 portfolio.

Ensure that the risk (volatility) levels of different decile portfolios are consistent.

- Making it more fair and meaningful to compare the performance of different portfolios.

- If not adjusted, low volatility portfolios may be overinvested due to appearing ‘more stable’.
- May mislead investors(these portfolios did not actually provide higher risk adjusted returns).

Results for Research questions



Q1: Superior performance of Probability Forecasts Portfolios

	OLS					Logit					PLS					NN1				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	-0.02	0.45	0.74	5.01	0.51	0.38	0.39	-0.69	7.77	-0.31	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.75	8.11	-0.32
2	0.14	0.44	0.31	5.48	0.19	0.41	0.43	0.14	7.73	0.06	0.41	0.43	0.00	7.77	0.00	0.41	0.43	-0.05	7.95	-0.02
3	0.29	0.44	0.10	5.20	0.07	0.43	0.44	0.48	7.72	0.21	0.43	0.44	0.59	7.76	0.26	0.43	0.44	0.49	8.02	0.21
4	0.38	0.44	0.34	5.25	0.23	0.45	0.45	0.62	7.64	0.28	0.45	0.45	0.63	7.69	0.28	0.45	0.46	0.51	8.08	0.22
5	0.42	0.45	0.53	5.10	0.36	0.46	0.46	0.79	7.63	0.36	0.46	0.46	0.71	7.67	0.32	0.46	0.46	0.76	7.99	0.33
6	0.44	0.46	0.51	5.15	0.35	0.47	0.47	0.93	7.71	0.42	0.47	0.47	0.95	7.62	0.43	0.47	0.47	0.94	7.87	0.42
7	0.47	0.47	0.63	5.04	0.43	0.48	0.48	1.23	7.49	0.57	0.48	0.48	1.12	7.55	0.52	0.48	0.48	1.22	7.91	0.54
8	0.49	0.48	0.73	4.98	0.51	0.49	0.48	1.46	7.40	0.69	0.49	0.48	1.30	7.41	0.61	0.49	0.48	1.35	7.78	0.60
9	0.52	0.48	0.89	4.87	0.63	0.5	0.49	1.57	7.36	0.74	0.5	0.49	1.55	7.42	0.72	0.51	0.49	1.51	7.62	0.69
High (H)	0.58	0.49	1.17	4.86	0.84	0.53	0.51	1.92	7.45	0.89	0.53	0.51	1.91	7.48	0.89	0.53	0.51	2.00	7.70	0.90
H-L	-	-	0.43	3.66	0.41	-	-	2.61	6.19	1.46	-	-	2.54	6.44	1.36	-	-	2.75	6.32	1.51

Except for OLS, returns of portfolios generally increase in line with probability forecasts.

Avg Prob is closely related to *Realized Prob*, emphasizing effectiveness of the model.

Q2: Simple Models worked well as Neural Networks

	OLS					Logit					PLS					NN1				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	-0.02	0.45	0.74	5.01	0.51	0.38	0.39	-0.69	7.77	-0.31	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.75	8.11	-0.32
2	0.14	0.44	0.31	5.48	0.19	0.41	0.43	0.14	7.73	0.06	0.41	0.43	0.00	7.77	0.00	0.41	0.43	-0.05	7.95	-0.02
3	0.29	0.44	0.10	5.20	0.07	0.43	0.44	0.48	7.72	0.21	0.43	0.44	0.59	7.76	0.26	0.43	0.44	0.49	8.02	0.21
4	0.38	0.44	0.34	5.25	0.23	0.45	0.45	0.62	7.64	0.28	0.45	0.45	0.63	7.69	0.28	0.45	0.46	0.51	8.08	0.22
5	0.42	0.45	0.53	5.10	0.36	0.46	0.46	0.79	7.63	0.36	0.46	0.46	0.71	7.67	0.32	0.46	0.46	0.76	7.99	0.33
6	0.44	0.46	0.51	5.15	0.35	0.47	0.47	0.93	7.71	0.42	0.47	0.47	0.95	7.62	0.43	0.47	0.47	0.94	7.87	0.42
7	0.47	0.47	0.63	5.04	0.43	0.48	0.48	1.23	7.49	0.57	0.48	0.48	1.12	7.55	0.52	0.48	0.48	1.22	7.91	0.54
8	0.49	0.48	0.73	4.98	0.51	0.49	0.48	1.46	7.40	0.69	0.49	0.48	1.30	7.41	0.61	0.49	0.48	1.35	7.78	0.60
9	0.52	0.48	0.89	4.87	0.63	0.5	0.49	1.57	7.36	0.74	0.5	0.49	1.55	7.42	0.72	0.51	0.49	1.51	7.62	0.69
High (H)	0.58	0.49	1.17	4.86	0.84	0.53	0.51	1.92	7.45	0.89	0.53	0.51	1.91	7.48	0.89	0.53	0.51	2.00	7.70	0.90
H-L	-	-	0.43	3.66	0.41	-	-	2.61	6.19	1.46	-	-	2.54	6.44	1.36	-	-	2.75	6.32	1.51

	NN2					NN3					NN4					NN5				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	0.37	0.39	-0.63	8.04	-0.27	0.37	0.39	-0.75	8.16	-0.32	0.37	0.39	-0.64	8.03	-0.27	0.37	0.39	-0.62	8.27	-0.26
2	0.41	0.43	0.24	7.99	0.10	0.4	0.43	0.24	8.14	0.10	0.4	0.43	0.12	7.87	0.05	0.4	0.43	0.32	8.10	0.14
3	0.43	0.45	0.55	8.02	0.24	0.42	0.45	0.52	8.30	0.22	0.42	0.45	0.57	7.98	0.25	0.42	0.45	0.49	8.36	0.20
4	0.44	0.46	0.51	8.06	0.22	0.44	0.46	0.84	8.14	0.36	0.43	0.46	0.48	7.89	0.21	0.43	0.46	0.95	8.19	0.40
5	0.46	0.46	0.78	7.87	0.34	0.45	0.46	0.58	8.10	0.25	0.44	0.46	0.83	7.80	0.37	0.44	0.46	0.66	8.18	0.28
6	0.47	0.47	0.91	7.85	0.40	0.46	0.47	0.82	7.92	0.36	0.46	0.47	0.95	7.89	0.42	0.45	0.47	1.04	8.04	0.45
7	0.48	0.48	1.21	7.80	0.54	0.47	0.47	1.16	8.02	0.50	0.47	0.48	1.02	7.83	0.45	0.46	0.47	1.16	8.02	0.50
8	0.49	0.48	1.27	7.76	0.57	0.48	0.48	1.27	8.02	0.55	0.48	0.48	1.34	7.70	0.60	0.47	0.48	1.23	7.87	0.54
9	0.5	0.49	1.59	7.61	0.72	0.49	0.49	1.51	7.94	0.66	0.49	0.49	1.56	7.66	0.71	0.49	0.49	1.57	7.89	0.69
High (H)	0.52	0.5	1.93	7.76	0.86	0.51	0.5	2.10	7.95	0.91	0.5	0.5	1.95	7.70	0.88	0.5	0.5	2.01	7.92	0.88
H-L	-	-	2.56	6.57	1.35	-	-	2.84	6.86	1.44	-	-	2.59	6.74	1.33	-	-	2.63	7.15	1.27

Single layer of neural network or Logit model defeats a neural network with more layers.

Q3: Probability Forecasts + Return Forecast = Better Forecasts

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
Corr	0.25	0.34	0.34	0.27	0.32	0.37	0.33	0.30
Panel A: Probability Forecast								
Mean	0.43	2.61	2.54	2.75	2.56	2.84	2.59	2.63
SR	0.41	1.46	1.36	1.51	1.35	1.44	1.33	1.27
α	0.42	1.81	1.73	1.88	1.64	1.85	1.69	1.55
t_α	2.15	5.72	5.60	6.62	5.35	5.49	5.52	5.18
Panel B: Expected Return Forecast								
Mean	3.00	3.00	3.00	2.24	2.68	2.87	3.00	2.89
SR	1.43	1.43	1.43	1.18	1.23	1.30	1.43	1.34
α	2.72	2.72	2.72	1.76	2.29	2.52	2.72	2.54
t_α	4.45	4.45	4.45	3.52	3.41	3.70	4.45	4.06
Panel C: 1/N Combination of Probability and Expected Return Forecasts								
Mean	1.71	2.80	2.77	2.50	2.62	2.85	2.79	2.76
SR	1.33	1.76	1.71	1.68	1.58	1.65	1.69	1.62
α	1.57	2.27	2.22	1.82	1.96	2.19	2.21	2.04
t_α	4.36	5.65	5.58	5.99	4.84	4.98	5.73	5.27
Panel D: Mean-variance Combination of Probability and Expected Return Forecasts								
Mean	6.62	9.62	9.36	9.02	9.21	9.17	9.24	9.09
SR	1.38	1.73	1.68	1.68	1.63	1.65	1.65	1.63
α	6.53	8.29	8.18	6.89	7.61	7.57	7.91	7.53
t_α	4.56	5.37	5.19	5.63	5.00	5.00	5.15	4.94

Conclusions

1. **Empirical value of probability prediction**:

 - Probability forecasts generate significant and comparable performance to return forecasts.
 - Probability forecasts provide additional information compared to return forecasts, combining them generates long-short portfolios with better Sharpe ratios.

2. **Simplified model for probability prediction**:

 - Simple model such as logit regression achieves similar forecasts and portfolio performance as complex neural network models.
 - This suggests that probability of outperforming a benchmark is a simpler object to learn.

探讨 & 未来方向

1. **概率预测的假设限制**:
   - 比较大的限制是本文的概率预测建模要求有强的正态分布假设，现实中资产回报可能不符合正态分布（可能具有厚尾特性）。
2. **文中与 Gu2020 的对比**:
   作者在展示概率预测模型表现时，并没有提供与 Gu2020 的直观对比表格或图片，仅仅是用文字阐述了表现相当，这里似乎加入与 Gu2020 的结果做一个对比图会更好。
3. **跨资产类别的概率预测**:
   - 探索概率预测模型在不同资产类别（如公司债券、货币、基金）中的应用，以及这些模型如何帮助投资者在不同市场条件下做出更明智的投资决策。
   - 这与传统上主要通过回报预测视角来探索资产定价问题的方法形成对比。

Appendix1：计算股票超过市场的概率

对于超额回报 $R_{i,t+1} - R_{mkt,t+1}$，我们假设它服从均值为 $\mu_i - \mu_{mkt}$ 和标准差为 $\sigma_{t+1|t}$ 的正态分布。

标准化超额回报

标准化的目的是将这个分布转换为标准正态分布 $N(0,1)$，这样我们就可以使用标准正态分布的累积分布函数 Φ 来计算概率。标准化公式如下：$Z = \frac{X-\mu}{\sigma}$。其中：

X 是原始数据（在这里是超额回报 $R_{i,t+1} - R_{mkt,t+1}$）。

μ 是 X 的均值（在这里是 $\mu_i - \mu_{mkt}$）。

σ 是 X 的标准差（在这里是 $\sigma_{t+1|t}$）。

Z 是标准化后的数据，期望为 0，标准差为 1。

将超额回报应用此公式得到：$Z = \frac{(R_{i,t+1}-R_{mkt,t+1})-(\mu_i-\mu_{mkt})}{\sigma_{t+1|t}}$。$Z$ 值服从标准正态分布 $N(0,1)$。

Appendix2：应用标准正态分布的累积分布函数

$$R_{i,t+1} \mid \Omega_t \sim \mathcal{N}\left(\mu_i, \sigma^2_{i,t+1|t}\right)$$

$$R_{mkt,t+1} \mid \Omega_t \sim \mathcal{N}\left(\mu_{mkt}, \sigma^2_{mkt,t+1|t}\right)$$

$$R_{i,t+1} - R_{mkt,t+1} \sim N(\mu_i - \mu_{mkt}, \sigma^2_{i,t+1|t} + \sigma^2_{mkt,t+1|t} - 2\rho\sigma_{i,t+1|t}\sigma_{mkt,t+1|t})$$

累积分布函数 $\Phi(z)$ 给出了标准正态随机变量小于或等于 z 的概率。在我们的例子中，我们想要计算的是：$P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z)$。

其中 z 是我们标准化后的值。在例子中，Z 大于 0 的概率，即股票 i 超过市场 m 的概率。因此，我们有：$P(R_{i,t+1} - R_{mkt,t+1} > 0) = P(Z > 0) = 1 - \Phi\left(\frac{\mu_i - \mu_{mkt}}{\sigma_{t+1|t}}\right)$。

这就是如何从标准化超额回报到应用标准正态分布的累积分布函数来计算股票超过市场的概率。

Appendix3: Whether the stock outperforms the risk-free rate?

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
corr	0.24	0.21	0.18	0.19	0.17	0.15	0.18	0.20
Panel A: Probability Forecast								
Mean	0.41	3.10	3.09	3.02	3.02	3.07	3.12	2.96
SR	0.39	1.35	1.35	1.23	1.24	1.23	1.27	1.26
α	0.27	1.85	1.82	1.67	1.69	1.70	1.77	1.64
t_α	1.78	5.45	5.69	5.20	5.25	5.13	5.67	5.31
Panel B: Expected Return Forecast								
Mean	3.00	3.00	3.00	2.24	2.68	2.87	3.00	2.89
SR	1.43	1.43	1.43	1.18	1.23	1.30	1.43	1.34
α	2.72	2.72	2.72	1.76	2.29	2.52	2.72	2.54
t_α	4.45	4.45	4.45	3.52	3.41	3.70	4.45	4.06
Panel C: 1/N Combination of Probability and Expected Return Forecasts								
Mean	1.71	3.05	3.04	2.63	2.85	2.97	3.06	2.93
SR	1.33	1.79	1.81	1.55	1.61	1.66	1.75	1.68
α	1.50	2.29	2.27	1.72	1.99	2.11	2.25	2.09
t_α	4.52	6.27	6.41	5.64	5.17	5.64	6.37	5.61
Panel D: Mean-variance Combination of Probability and Expected Return Forecasts								
Mean	6.65	10.37	10.71	8.53	11.22	9.76	10.95	10.34
SR	1.37	1.75	1.76	1.55	1.70	1.66	1.73	1.70
α	6.55	8.68	8.99	6.46	9.28	8.02	9.17	8.52
t_α	4.46	5.39	5.32	4.35	4.63	4.83	5.16	4.78

Appendix4: Whether the stock outperforms its β*market return?

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
Corr	0.13	0.29	0.28	0.32	0.33	0.27	0.29	0.32
Panel A: Probability Forecast								
Mean	0.49	2.94	2.90	3.00	2.92	2.87	2.96	2.63
SR	0.41	1.32	1.23	1.25	1.13	1.21	1.20	1.06
α	0.19	1.73	1.66	1.77	1.61	1.64	1.64	1.31
t_α	1.18	5.83	5.35	4.94	4.11	5.23	4.46	3.51
Panel B: Expected Return Forecast								
Mean	3.00	3.00	3.00	2.24	2.68	2.87	3.00	2.89
SR	1.43	1.43	1.43	1.18	1.23	1.30	1.43	1.34
α	2.72	2.72	2.72	1.76	2.29	2.52	2.72	2.54
t_α	4.45	4.45	4.45	3.52	3.41	3.70	4.45	4.06
Panel C: 1/N Combination of Probability and Expected Return Forecasts								
Mean	1.74	2.97	2.95	2.62	2.80	2.87	2.98	2.76
SR	1.37	1.71	1.65	1.49	1.44	1.57	1.62	1.46
α	1.46	2.23	2.19	1.77	1.95	2.08	2.18	1.92
t_α	4.85	5.91	5.84	5.60	4.58	5.35	5.54	5.00
Panel D: Mean-variance Combination of Probability and Expected Return Forecasts								
Mean	6.66	9.25	8.97	7.97	9.04	8.17	8.92	8.02
SR	1.38	1.70	1.66	1.51	1.53	1.57	1.63	1.52
α	6.62	7.78	7.64	6.04	7.74	6.77	7.60	6.81
t_α	4.52	5.33	5.20	4.66	4.52	4.98	4.99	4.48