Empirical Asset Pricing with Probability Forecasts

Songrun He[1], Linying Lv[1], and Guofu Zhou[1]

[1]Washington University in St. Louis

01/04/2025
AFA Annual Meeting 2025

Motivation

- Traditional asset pricing focuses on expected returns predictions using characteristics like those described by Fama and MacBeth (1973), Lewellen (2014), Green, Hand, and Zhang (2017).
- Existing literature documents significant non-linearity in expected returns as a function of these characteristics:
  - Gu, Kelly, and Xiu (2020), Gu, Kelly, and Xiu (2021), Chen, Pelger and Zhu (2023).
- However, focusing on $\mathbb{E}[R]$ ignores the risk of a stock. The two are usually closely interconnected.

Information Ratio Perspective Probability Forecasts

- This paper proposes a new objective: predicting the probability that a stock will outperform a benchmark.
- For example, if the benchmark is the risk-free rate, under normality:

$$\mathbb{P}(r_{i,t+1} > r_{f,t}) = \Phi\left(\frac{\mu_{it} - r_{f,t}}{\sigma_{it}}\right) = \Phi\left(SR_{it}\right).$$

- So, this incorporates both return and risk in one number and can serve as a valuable objective.

Simple Model Works!

- In general, there is a growing consensus: researchers should embrace complexity (Kelly, Malamud, and Zhou (2024)).
  - With proper shrinkage & cross-validation, the machine-learning model can perform automatic model selection. $\Longrightarrow$ superb performance out-of-sample.
- We consider a wide range of complex models for probability forecast with a large number of characteristics.
- Surprisingly, simple models achieve similar performance as complex models.
  - They achieve comparable performance as state-of-the-art return prediction models.
  - Suggesting probability might be a simpler object to learn and model.

Literature

- Probability forecast in the time-series context: Christoffersen and Diebold (2006), Chevapatrakul (2013), Catania et al. (2019), Moskowitz et al. (2012), and Papailias et al. (2021).
- Machine learning in asset pricing: Gu, Kelly, and Xiu (2020), Kelly, Pruitt, and Su (2019), Chen, Pelger, and Zhu (2023), Han et al. (2023), Cong et al. (2022), Cong et al. (2023).
- Applications of machine learning to other asset classes and contexts: Bali et al. (2020), Bali et al. (2023), Filippou et al. (2020), Kaniel et al. (2023), and Wu et al. (2021).

Methodology

- Prediction models with different complexity:
  - Linear probability model: OLS, PLS. ◦ model details
  - Logit probability model: Logistic Regression, Neural Networks with 1 to 5 layers. ◦ model details; ◦ NN training
- Forecasting target: the probability of outperforming the market.

$$y_{i,t+1} = \mathbf{1}\left\{r_{i,t+1} > r_{t+1}^{mkt}\right\}.$$

Evaluation: Variance-adjusted Portfolio Sorting

- Sorting on probability will generate portfolios with different volatility:
  - Different loadings on common factors.
  - $\Longrightarrow$ Long-short strategy will take on additional factor risk.
- Adjust volatility of decile portfolios in realtime to have the same volatility as the decile 1 portfolio:

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

  - $\sigma_{j,t}$ is a real-time estimate of portfolio volatility based on a rolling window of the past 36 months of data.

Variance Adjusted Portfolio Performance

	OLS					Logit				
	$\widehat{Prob}$	$Prob$	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
2	0.14	0.44	0.31	5.48	0.19	0.41	0.43	0.14	7.73	0.06
3	0.29	0.44	0.10	5.20	0.07	0.43	0.44	0.48	7.72	0.21
4	0.38	0.44	0.34	5.25	0.23	0.45	0.45	0.62	7.64	0.28
5	0.42	0.45	0.53	5.10	0.36	0.46	0.46	0.79	7.63	0.36
6	0.44	0.46	0.51	5.15	0.35	0.47	0.47	0.93	7.71	0.42
7	0.47	0.47	0.63	5.04	0.43	0.48	0.48	1.23	7.49	0.57
8	0.49	0.48	0.73	4.98	0.51	0.49	0.48	1.46	7.40	0.69
9	0.52	0.48	0.89	4.87	0.63	0.50	0.49	1.57	7.36	0.74
High (H)	0.58	0.49	1.17	4.86	0.84	0.53	0.51	1.92	7.45	0.89
H-L	-	-	0.43	3.66	0.41	-	-	2.61	6.19	1.46

	PLS					NN4				
	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR
Low (L)	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.64	8.03	-0.27
⋮			⋮					⋮		
High (H)	0.53	0.51	1.91	7.48	0.89	0.50	0.50	1.95	7.70	0.88
H-L	-	-	2.54	6.44	1.36	-	-	2.59	6.74	1.33

Variance Adjusted Portfolio Performance

	OLS					Logit				
	$\widehat{Prob}$	*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
2	0.14	0.44	0.31	5.48	0.19	0.41	0.43	0.14	7.73	0.06
3	0.29	0.44	0.10	5.20	0.07	0.43	0.44	0.48	7.72	0.21
4	0.38	0.44	0.34	5.25	0.23	0.45	0.45	0.62	7.64	0.28
5	0.42	0.45	0.53	5.10	0.36	0.46	0.46	0.79	7.63	0.36
6	0.44	0.46	0.51	5.15	0.35	0.47	0.47	0.93	7.71	0.42
7	0.47	0.47	0.63	5.04	0.43	0.48	0.48	1.23	7.49	0.57
8	0.49	0.48	0.73	4.98	0.51	0.49	0.48	1.46	7.40	0.69
9	0.52	0.48	0.89	4.87	0.63	0.50	0.49	1.57	7.36	0.74
High (H)	0.58	0.49	1.17	4.86	0.84	0.53	0.51	1.92	7.45	0.89
H-L	-	-	0.43	3.66	0.41	-	-	2.61	6.19	1.46

	PLS					NN4				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.64	8.03	-0.27
⋮			⋮					⋮		
High (H)	0.53	0.51	1.91	7.48	0.89	0.50	0.50	1.95	7.70	0.88
H-L	-	-	2.54	6.44	1.36	-	-	2.59	6.74	1.33

Variance Adjusted Portfolio Performance

	OLS					Logit				
	$\widehat{Prob}$	$Prob$	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
2	0.14	0.44	0.31	5.48	0.19	0.41	0.43	0.14	7.73	0.06
3	0.29	0.44	0.10	5.20	0.07	0.43	0.44	0.48	7.72	0.21
4	0.38	0.44	0.34	5.25	0.23	0.45	0.45	0.62	7.64	0.28
5	0.42	0.45	0.53	5.10	0.36	0.46	0.46	0.79	7.63	0.36
6	0.44	0.46	0.51	5.15	0.35	0.47	0.47	0.93	7.71	0.42
7	0.47	0.47	0.63	5.04	0.43	0.48	0.48	1.23	7.49	0.57
8	0.49	0.48	0.73	4.98	0.51	0.49	0.48	1.46	7.40	0.69
9	0.52	0.48	0.89	4.87	0.63	0.50	0.49	1.57	7.36	0.74
High (H)	0.58	0.49	1.17	4.86	0.84	0.53	0.51	1.92	7.45	0.89
H-L	-	-	0.43	3.66	0.41	-	-	2.61	6.19	1.46

	PLS					NN4				
	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR
Low (L)	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.64	8.03	-0.27
⋮			⋮					⋮		
High (H)	0.53	0.51	1.91	7.48	0.89	0.50	0.50	1.95	7.70	0.88
H-L	-	-	2.54	6.44	1.36	-	-	2.59	6.74	1.33

Variance Adjusted Portfolio Performance

	OLS					Logit				
	$\widehat{Prob}$	$Prob$	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
2	0.14	0.44	0.31	5.48	0.19	0.41	0.43	0.14	7.73	0.06
3	0.29	0.44	0.10	5.20	0.07	0.43	0.44	0.48	7.72	0.21
4	0.38	0.44	0.34	5.25	0.23	0.45	0.45	0.62	7.64	0.28
5	0.42	0.45	0.53	5.10	0.36	0.46	0.46	0.79	7.63	0.36
6	0.44	0.46	0.51	5.15	0.35	0.47	0.47	0.93	7.71	0.42
7	0.47	0.47	0.63	5.04	0.43	0.48	0.48	1.23	7.49	0.57
8	0.49	0.48	0.73	4.98	0.51	0.49	0.48	1.46	7.40	0.69
9	0.52	0.48	0.89	4.87	0.63	0.50	0.49	1.57	7.36	0.74
High (H)	0.58	0.49	1.17	4.86	0.84	0.53	0.51	1.92	7.45	0.89
H-L	-	-	0.43	3.66	0.41	-	-	2.61	6.19	1.46

	PLS					NN4				
	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR
Low (L)	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.64	8.03	-0.27
⋮			⋮					⋮		
High (H)	0.53	0.51	1.91	7.48	0.89	0.50	0.50	1.95	7.70	0.88
H-L	-	-	2.54	6.44	1.36	-	-	2.59	6.74	1.33

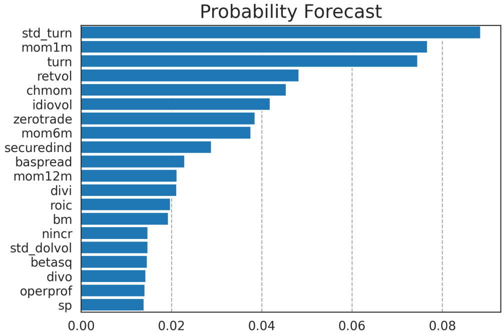
Incremental Information Relative to Expected Return Forecasts

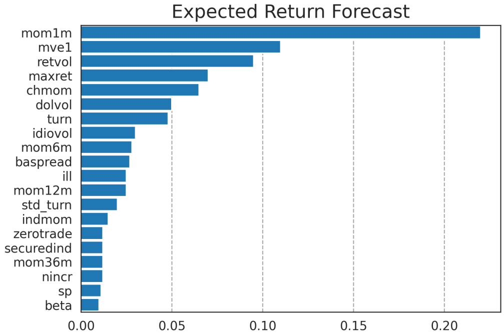
- Construct the following portfolios:
  - Probability forecast long-short portfolios.
  - Expected return forecast long-short portfolios from Gu, Kelly, and Xiu (2020).
  - Combinations of the two types of portfolios.
- Two combination approaches:
  - Equal-weighted combination: robust to estimation errors.
  - Realtime mean-variance efficient combination.

Incremental Information Relative to Expected Return Forecasts

	OLS	Logit	PLS	NN4
Corr	0.25	0.34	0.34	0.33
Panel A: Probability Forecast				
SR	0.41	1.46	1.36	1.33
t_{α}	2.15	5.72	5.60	5.52
Panel B: Expected Return Forecast (NN4)				
SR	1.43	1.43	1.43	1.43
t_{α}	4.45	4.45	4.45	4.45
Panel C: 1/N Combination of Probability and Expected Return Forecasts				
SR	1.33	1.76	1.71	1.69
t_{α}	4.36	5.65	5.58	5.73
Panel D: Mean-variance Combination of Probability and Expected Return Forecasts				
SR	1.38	1.73	1.68	1.65
t_{α}	4.56	5.37	5.19	5.15

Variable Importance: Probability vs. Expected Return Forecasts





Using Factor Models as Benchmark

- Assume that realized stock returns follow a factor model:

$$r_{i,t+1} - r_{f,t} = \alpha_{it} + \beta_{it}' f_{t+1} + \epsilon_{i,t+1}.$$

  - $\epsilon_{i,t+1} \sim \mathcal{N}(0, \sigma_{it}^2)$ is the idiosyncratic return.

- Probability of outperforming the factor model can be written as:

$$\mathbb{P}(r_{i,t+1} - r_{f,t} - \beta_{it} f_{t+1} > 0) = \Phi\left(\frac{\alpha_{it}}{\sigma_{it}}\right).$$

  - Sorting on probability generates a spread in magnitude of mispricing (information ratio) relative to the benchmark factor model.

Augmenting Factor Models with Probability Forecasts

	CAPM	FF3	FF6	HXZ	SY	DHS	IPCA	CA
Corr	0.06	-0.11	0.20	0.28	0.20	0.29	-0.28	-0.17
Panel A: Probability 'H-L' Portfolios								
Mean	2.74	2.66	2.32	2.28	2.20	2.48	2.17	2.58
SD	6.90	6.50	6.02	5.61	5.67	6.27	9.00	6.42
SR	1.37	1.42	1.34	1.41	1.34	1.37	0.84	1.39
Panel B: Factor Tangency Portfolios								
Mean	0.73	0.29	0.31	0.44	0.59	0.51	4.42	3.81
SD	4.51	2.05	1.18	1.03	1.48	1.31	9.00	6.42
SR	0.56	0.49	0.92	1.47	1.38	1.35	1.70	2.05
Panel C: Combination of Probability 'H-L' and Factor Tangency Portfolios								
Mean	1.73	0.88	0.60	0.74	0.99	0.91	3.30	3.20
SD	4.24	2.11	1.45	1.44	1.97	1.86	5.39	5.08
SR	1.42	1.45	1.43	1.79	1.74	1.69	2.12	2.18
IR	1.30	1.36	1.09	1.02	1.07	1.01	1.26	0.74

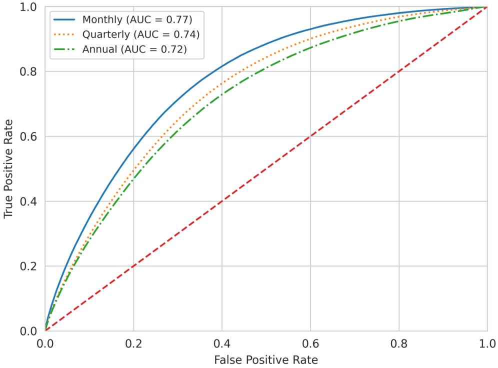
Forecasting and Managing Tail Risks

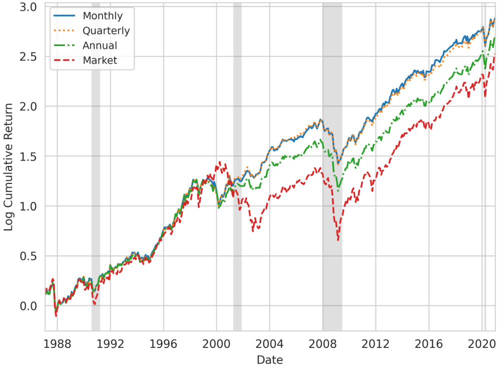
- Predicting the probability that a stock will experience a larger than -20% decline in the future 1 month to 1 year.

	$\widehat{\mathbb{P}}\{R^{1m}_{i,t+1} < -0.2\}$		$\widehat{\mathbb{P}}\{R^{3m}_{i,t+1} < -0.2\}$		$\widehat{\mathbb{P}}\{R^{12m}_{i,t+1} < -0.2\}$	
	L	H	L	H	L	H
$\widehat{\text{Prob}}$	0.006	0.198	0.028	0.376	0.072	0.554
Prob	0.005	0.204	0.019	0.356	0.062	0.523
Mean	0.763	-0.519	0.765	-0.524	0.723	-0.295
Std	3.257	11.635	3.274	11.297	3.418	10.988
SR	0.811	-0.155	0.809	-0.161	0.733	-0.093
MaxDD	-0.362	-0.999	-0.361	-0.999	-0.402	-0.997

- Contemporaneous market excess return: Mean 0.72%, Std: 4.51%, SR: 0.56, MaxDD: -0.54.

Forecasting and Managing Tail Risks





Conclusion

- We propose a simpler modeling objective: probability forecasts:
  - Unified objective of risk and return.
  - Simple model works.
  - Superior portfolio performance when combined with expected return forecasts and factor models.
- Probability learnings allow for forecasting and managing tail risks.
- Future research could explore wider applications of probability forecasts across different asset classes.

Thank you!

Information Ratio Perspective of Probability Forecast

- The probability of outperformance is linked to the Information Ratio (IR).
- Under normality: $r_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$ and $r_b \sim \mathcal{N}\left(\mu_b, \sigma_b^2\right)$.
- Probability of outperformance can be expressed as:

$$\begin{aligned}
\mathbb{P}(r_i - r_b > 0) &= 1 - \mathbb{P}(r_i - r_b < 0) \\
&= 1 - \mathbb{P}\left(\frac{(r_i - r_b) - (\mu_i - \mu_b)}{\sigma(r_i - r_b)} < \frac{-(\mu_i - \mu_b)}{\sigma(r_i - r_b)}\right) \\
&= \Phi\left(\frac{\mathbb{E}[r_i - r_b]}{\sigma(r_i - r_b)}\right).
\end{aligned}$$

Linear Probability Models

- The loss function for linear probability models is the mean-squared loss:

$$L(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t+1} - g\left(z_{i,t};\theta\right)\right)^2.$$

- OLS regression: $g\left(z_{i,t};\theta\right) = z_{i,t}'\theta$.
- PLS regression: $g(z_{i,t};\theta) = (z_{i,t}'\Omega)'\theta$.
  - Ω is a $K \times P$ matrix that transforms K predictors in $z_{i,t}$ into P lower dimensional components.
  - Each component is extracted to maximize covariance with prediction target y while being orthogonal to previous components.

$$\omega_j = \arg\max_{\omega} \text{Cov}(Y, Z\omega), \quad \text{s.t. } \omega'\omega = 1, \text{Cov}\left(Z\omega, Z\omega_i\right) = 0, i = 1, 2, \cdots j-1$$

Back

Logit Probability Models

- The loss function for logit probability is the cross-entropy loss:

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

  - Negative log-likelihood is equivalent to cross-entropy loss.
  - Minimizing cross-entropy is maximizing the log-likelihood of the model's predictions.

- Use the Sigmoid function to ensure the probability forecast is within 0 and 1. $S(x) = \frac{1}{1+e^{-x}}$ and $g\left(z_{i,t};\theta\right) = S\left(f\left(z_{i,t};\theta\right)\right)$.

- Logistic regression: $f\left(z_{i,t};\theta\right) = z_{i,t}'\theta$.

- Neural networks: $f\left(z_{i,t};\theta\right) = W_3 \times \sigma\left(W_2 \times \sigma\left(W_1 z_{i,t} + b_1\right) + b_2\right) + b_3$.
  - $\sigma(\cdot)$ is the ReLU activation function adding non-linearity.

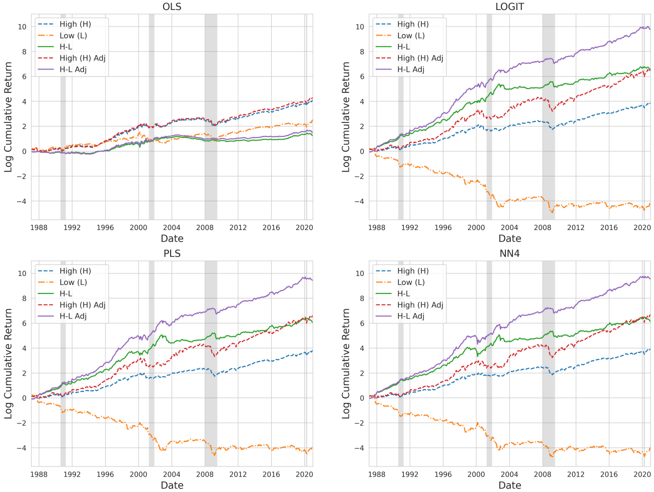
◦ Back

Details on Training Neural Networks

- Our neural network training procedure follows closely with Gu, Kelly, and Xiu (2020).
- Specifications with 1 to 5 layers: 32 neurons in the first layer, followed by 16, 8, 4, and 2 in subsequent layers.
- Regularization with L2 penalty on layer weights and early stop depending on validation loss.
- Hyperparameters include learning rate and regularization strength determined on the validation set.
- Optimized with the Adam optimizer.
- Model ensemble by retraining the NN 5 times with different random initialization.

◦ Back

Cumulative Portfolio Returns



○ Non-var adjusted portfolios

Sorted Portfolio Performance without Variance Adjustment

	OLS					Logit				
	$\widehat{Prob}$	*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
2	0.14	0.44	0.40	6.10	0.23	0.41	0.43	0.06	6.60	0.03
3	0.29	0.44	0.30	6.76	0.15	0.43	0.44	0.30	5.86	0.18
4	0.38	0.44	0.55	6.16	0.31	0.45	0.45	0.43	5.44	0.28
5	0.42	0.45	0.67	5.31	0.44	0.46	0.46	0.46	5.28	0.30
6	0.44	0.46	0.64	4.92	0.45	0.47	0.47	0.56	4.75	0.41
7	0.47	0.47	0.66	4.61	0.49	0.48	0.48	0.69	4.55	0.52
8	0.49	0.48	0.69	4.32	0.55	0.49	0.48	0.78	4.49	0.60
9	0.52	0.48	0.83	4.07	0.70	0.50	0.49	0.87	4.41	0.68
High (H)	0.58	0.49	1.10	4.45	0.86	0.53	0.51	1.06	4.65	0.79
H-L	-	-	0.36	3.38	0.37	-	-	1.75	5.73	1.06

○ Back

Sorted Portfolio Performance without Variance Adjustment

	PLS					NN4				
	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR
Low (L)	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.64	8.03	-0.27
2	0.41	0.43	-0.04	6.81	-0.02	0.40	0.43	0.09	7.45	0.04
3	0.43	0.44	0.38	6.19	0.22	0.42	0.45	0.42	6.27	0.23
4	0.45	0.45	0.43	5.48	0.27	0.43	0.46	0.38	5.65	0.23
5	0.46	0.46	0.43	5.24	0.29	0.44	0.46	0.54	5.20	0.36
6	0.47	0.47	0.59	4.82	0.43	0.46	0.47	0.58	4.95	0.41
7	0.48	0.48	0.66	4.69	0.48	0.47	0.48	0.61	4.40	0.48
8	0.49	0.48	0.70	4.45	0.54	0.48	0.48	0.75	4.53	0.58
9	0.50	0.49	0.87	4.46	0.67	0.49	0.49	0.86	4.40	0.68
High (H)	0.53	0.51	1.04	4.53	0.79	0.50	0.50	1.06	4.43	0.83
H-L	-	-	1.66	5.86	0.98	-	-	1.70	6.16	0.96

○ Back

Variable Importance for Tail Risk Forecasts