Estimating Stock Market Betas via Machine Learning

Wolfgang Drobetz, Fabian Hollstein, Tizian Otto, Marcel Prokopczuk

Journal of Finalcial and Quantitative Analysis, 2024

解读人：赵伟皓

武汉大学金融系

2025.11.09

Research Question

1. **Prediction level (Which works better)**
   - Can(When) ML **more accurately** predicts β than traditional methods? –**Yes**
     - Does this outperformance have economic significance? –**Yes**
2. **Mechanism level (Why it works)**
   - If yes, Where do the advantages of machine learning come from?
     - **Numerous charateristics**, **Non-linear** relationships and **Interaction effects**.

Motivation

β matters for firms and investors, estimation errors are significant.

β is time-varying and not directly observable. (Campbell et al., 2001)

1. **Rolling OLS** :Bias-Variance trade-off, susceptible to outliers.
2. **WLS** (Hollstein et al., 2019) or **Winsorized** (Welch, 2022) reduces impact of outliers.
3. Cosemans et al. (2016) use **shrinkage** methods to reduce estimation errors.
4. **Portfolio** β (Fama et al., 1992); **Long Memory** Model (Becker et al., 2021).

The research focus is shifting towards how to accurately estimate β.

Benchmark Estimators

Panel A: Benchmark estimators

Model	Description	Definition
ols_5y_m	Historical beta	Rolling regressions using a five-year window of monthly returns
ols_1y_d	Historical beta	Rolling regressions using a one-year window of daily returns
ewma_s	Exponentially-weighted beta	Rolling regressions using a one-year window of daily returns with exponentially decaying weights (short half-life)
ewma_l	Exponentially-weighted beta	Rolling regressions using a one-year window of daily returns with exponentially decaying weights (long half-life)
bsw	Slope-winsorized beta	Rolling regressions using a one-year window of *winsorized* daily returns
vasicek	Shrinkage beta	Shrinkage of *ols_1y_d* towards average beta within stock universe
karolyi	Shrinkage beta	Shrinkage of *ols_1y_d* towards average beta within industry portfolio
hybrid	Shrinkage beta	Shrinkage of *ols_1y_d* towards firm-specific beta prior
fama-french	Portfolio beta	Assignment of portfolio betas (rolling regressions using a one-year window of daily post-ranking portfolio returns) to individual stocks
long-memo	Long-memory beta	Application of fractionally integrated long-memory time-series process

Motivation

2. Traditional algorithms have several limitations

- Lack of integration of high-dimensional info on firm characteristics.
- Not capable of capturing nonlinear relationships or interactions.

3. Most studies lack testing of economic value of accurate β in portfolio construction.

4. Prior literature rarely addressed reasons why ML models are effective.

Contribution I

1. Literature on Traditional β Estimation algorithms

Prior: Mainly utilizing historical returns, lacking integration of numerous Z_t.

- Rolling OLS/WLS: Black et al. (1972); Fama et al. (1973); Hollstein et al. (2019)
- Shrinkage: Cosemans et al. (2016); Long memory: Becker et al. (2021)

Extend: Underscoring systematic relationship between β and firm characteristics.

- ML captures info content of numerous firm characteristics that affect betas.
- Outperformance of RF etc., comes from exploiting nonlinear and interactive patterns.

Contribution II

2. Literature on application of ML in Asset Pricing

Prior: ML has been widely applied to various types of prediction.

- Return prediction: Gu et al. (2020); Leippold et al. (2021);
- Volatility(total risk) of stocks: Christensen et al. (2023); Bucci. (2022);
- **Jourovski et al. (2020) predicts β using linear regression and Tree models.**
  - Small sample size (540 stocks), short time period (1999-2019)
  - Not comparing ML with the optimal benchmarks (shrinkage; long memory)
  - Limitations of economic assessment, only examining BAB.

Extend: Integrating ML into "systematic risk estimation"(market β).

Extending Jourovski et al. (2020) (sample size, model comparison, strategy).

Hypothesis

H1： Compared with traditional methods, ML methods are able to **more accurately predict** future market β.

- Statistically more accurate forecasts lead to economic gains.

H2： ML models perform better because they capture **nonlinear relationship and interaction effects** between β and firm characteristics.

- Instead of relying solely on more data or more complex algorithms.

Framework

Research Objective:

- Estimate and forecast time-varying market betas ($\beta_{i,t}$)
- Compare Machine Learning (ML) methods vs. traditional rolling OLS
- Understand *why* ML improves estimation: nonlinearities and interactions

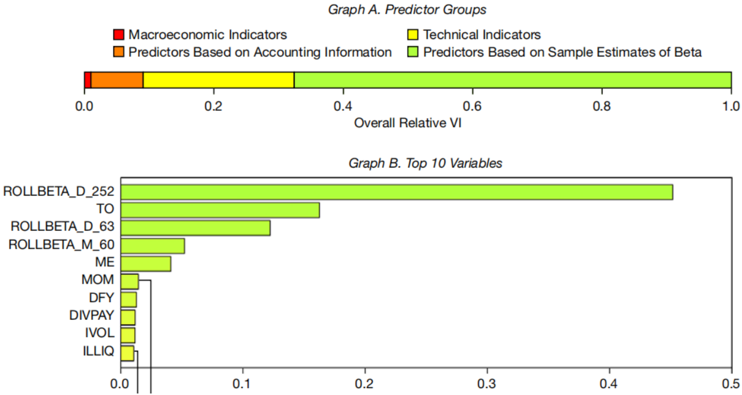
ML focus directly on the goal of predicting market β (Gu et al., 2020)

$$\beta^R_{i,t+k} = \mathbb{E}_t\left(\beta^R_{i,t+k}\right) + \epsilon_{i,t+k}$$

$$\mathbb{E}_t(\beta^R_{i,t}) = g^\star(Z_{i,t})$$

- Incorporate firm characteristics, market states, and macroeconomic variables.
- Use ML algorithms (RF, GBRT, NN) to estimate the unknown nonlinear mapping $g(\cdot)$

Variable Importance



ML capturing joint info content of a large set of $Z_{i,t}$ whichs affect betas.

Empirical Implementation

Data & Sample:

- CRSP + Compustat firms, NYSE/AMEX/NASDAQ (1970–2020)
- First estimated betas in Dec 1979; monthly frequency thereafter

Variables:

- *Target:* future beta $\beta_{i,t+12}$
- $Z_{i,t}$: 81 chars: past betas, Size, MOM, Turnover, IVOL, macro factors, etc.

Methodology:

- Rolling estimation: 9-year training, 1-year validation, 1-year test
- Compare: ML models (RF, GBRT, NN) vs. OLS, shrinkage, etc.
- Evaluation: MSE/MCS/DM, and economic gains via anomaly portfolios.

ML models(Linear v.s Nonlinear)

Panel B: Machine learning estimators

Model	Hyperparameter	Specification	Definition
lm			
	None		
elanet			
	λ	(0,1)	General strength of the penalization
	p	{0,0.5,1}	Weight on the lasso and ridge penalization
rf			
	L	(1,10)	Depth of the single regression trees
	M	{20,25,30,35,40}	Number of predictors randomly considered as potential split variables
	B	(10,500)	Number of trees added to the ensemble prediction
gbrt			
	L	(1,5)	Depth of the single regression trees
	v	{0.01,0.05,0.1}	Weight for the learning rate shrinkage
	B	(10,500)	Number of trees added to the ensemble prediction
nn_1–nn_5			
	batch size	1000	Batch size
	n_{epochs}	100	Number of epochs
	patience	25	Number of iterations during which the value-weighted mean squared error is allowed to increase in the validation sample
	dropout rate	0.1	Fractional rate of input variables that are randomly set to zero at each iteration
	n_{seeds}	10	Number of independent seeds used for each specification family

Model Comparison Framework: MCS & DM Tests

1. **Model Confidence Set (MCS)** Hansen et al. (2011)
   - Identifies models that are *statistically indistinguishable from the best model(s)*.
   - Reported value: fraction of months in which a model belongs to the MCS.
2. **Diebold–Mariano Test(DM)** Diebold & Mariano (1995)
   - Pairwise test of equal predictive ability between models j and l.
   - Based on the difference in squared forecast errors (SEs):

$$d_{i,t}^{(j,l)} = SE_{i,t}^{(j)} - SE_{i,t}^{(l)}, \quad SE_{i,t}^{(j)} = (\beta_{i,t+12}^{R} - \beta_{i,t+12|t}^{F,(j)})^2$$

MCS: overall robustness — how often a model performs best.

DM: pairwise dominance — how often a model significantly beats another.

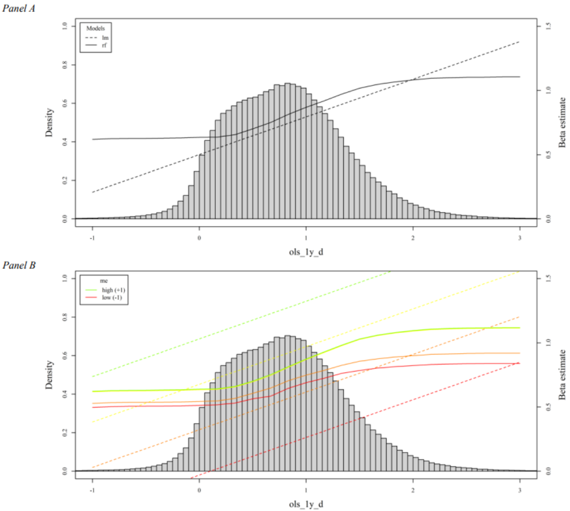
H1: Machine learning significantly improves prediction accuracy

	Benchmark Estimators										ML Estimators				
	OLS_5Y_M	OLS_1Y_D	EWMA_S	EWMA_L	BSW	VASICEK	KAROLYI	HYBRID	FAMA–FRENCH	LONG-MEMO	LM	ELANET	RF	GBRT	NN_1
Panel A. Average Forecast Errors															
MSE, value-weighted (%)	19.17	9.70	9.55	9.44	8.77	8.91	8.97	8.53	9.11	8.29	9.15	8.89	7.77	8.04	7.79
Panel B. Forecast Errors over Time															
In MCS	3.74	32.64	40.54	41.79	49.69	46.99	50.10	52.81	43.04	67.15	51.77	57.80	82.54	68.19	78.59
Benchmark Estimators															
vs. OLS_5Y_M		87.73	87.73	88.36	92.52	91.68	90.64	94.39	90.85	97.51	90.44	89.81	96.47	96.05	96.67
vs. OLS_1Y_D	1.87		32.22	48.44	74.22	81.50	83.58	72.35	50.31	60.29	46.78	50.73	68.61	61.75	62.99
vs. EWMA_S	1.25	28.27		39.29	52.60	50.52	48.02	54.05	40.75	55.72	43.04	45.74	62.99	59.88	59.88
vs. EWMA_L	1.25	16.84	20.58		55.51	52.60	50.52	59.46	39.50	54.26	43.24	46.15	62.79	58.21	59.46
vs. BSW	0.62	5.20	13.31	12.47		22.45	22.25	40.96	16.01	42.00	34.93	38.88	59.88	53.43	54.47
vs. VASICEK	1.04	6.86	14.55	13.72	29.73		19.75	50.52	12.68	43.04	35.97	39.50	60.50	56.55	56.13
vs. KAROLYI	1.04	5.82	15.59	12.89	29.94	30.15		49.69	18.50	43.04	34.93	40.54	57.80	52.39	54.05
vs. HYBRID	0.83	3.53	12.89	13.51	21.21	23.08	21.83		14.97	31.60	31.39	33.89	51.77	48.23	50.94
vs. FAMA–FRENCH	1.46	14.14	18.09	19.33	31.60	30.15	30.15	44.28		46.36	33.89	39.09	63.83	60.71	60.29
vs. LONG-MEMO	0.00	14.97	16.22	17.46	19.33	18.71	20.17	22.04	16.22		21.62	27.23	43.04	40.33	43.04
ML Estimators															
vs. LM	3.12	27.65	28.27	30.98	34.72	33.89	34.93	37.21	28.27	40.75		37.63	65.28	54.47	63.83
vs. ELANET	3.33	25.78	26.61	29.73	34.30	34.93	35.55	33.47	27.23	34.10	18.92		56.13	49.69	54.26
vs. RF	0.00	9.77	10.60	12.68	13.31	13.93	14.35	14.55	6.86	10.81	5.82	8.94		18.30	20.17
vs. GBRT	0.42	16.22	18.71	18.71	22.04	22.87	22.66	23.70	13.31	19.54	9.77	18.71	39.92		34.93
vs. NN_1	0.00	12.89	13.10	13.72	17.67	17.88	18.09	17.88	11.64	11.02	6.65	13.93	23.70	16.22	
t	481	481	481	481	481	481	481	481	481	481	481	481	481	481	481

H1: ML brings improvements in economic gains

Model	MOM α_{CAPM} [%]	MOM α_{FF5} [%]	MOM β	IVOL α_{CAPM} [%]	IVOL α_{FF5} [%]	IVOL β	BAB α_{CAPM} [%]	BAB α_{FF5} [%]	BAB β
Benchmark Estimators									
OLS_1Y_D	5.95	8.62	−0.09	10.94	8.93	0.05	8.13	6.95	0.41
	(1.67)	(2.46)	(−2.40)	(4.25)	(4.19)	(1.12)	(2.68)	(2.32)	(3.37)
BSW	6.20	8.89	−0.10	10.93	8.87	−0.05	6.91	5.64	0.23
	(1.74)	(2.57)	(−2.07)	(4.31)	(4.24)	(−1.20)	(2.31)	(1.93)	(2.33)
HYBRID	6.06	8.67	−0.11	10.84	8.81	0.01	9.54	7.89	0.23
	(1.70)	(2.48)	(−2.77)	(4.28)	(4.20)	(0.29)	(3.15)	(2.71)	(1.81)
FAMA–FRENCH	6.01	8.76	−0.12	10.91	8.86	−0.06	8.05	6.74	0.20
	(1.68)	(2.55)	(−2.48)	(4.18)	(4.11)	(−1.50)	(2.83)	(2.46)	(1.86)
LONG-MEMO	6.62	9.44	−0.11	11.46	9.45	−0.04	8.95	7.64	0.19
	(1.86)	(2.82)	(−2.17)	(4.52)	(4.45)	(−1.27)	(3.11)	(2.67)	(2.47)
ML Estimators									
LM	6.84	9.32	−0.09	11.68	9.54	−0.08	9.00	7.01	0.03
	(1.88)	(2.58)	(−1.34)	(4.59)	(4.47)	(−1.91)	(2.68)	(2.30)	(0.39)
ELANET	6.26	8.66	−0.09	11.61	9.35	−0.09	9.03	7.19	−0.03
	(1.70)	(2.34)	(−1.30)	(4.48)	(4.34)	(−2.10)	(2.83)	(2.51)	(−0.33)
RF	7.05	9.79	−0.07	11.54	9.46	−0.09	9.31	7.81	−0.05
	(1.97)	(2.91)	(−1.13)	(4.54)	(4.49)	(−2.23)	(2.91)	(2.57)	(−0.55)
GBRT	6.98	9.74	−0.04	11.74	9.65	−0.07	10.13	8.69	−0.02
	(1.95)	(2.90)	(−0.70)	(4.60)	(4.53)	(−1.57)	(3.13)	(2.82)	(−0.26)
NN_1	6.79	9.49	−0.04	11.80	9.69	−0.01	10.01	8.45	0.03
	(1.89)	(2.79)	(−0.66)	(4.63)	(4.58)	(−0.31)	(3.13)	(2.71)	(0.31)

H2: Nonlinearity and Interaction effects



Extension

1. **本文对时变性的刻画使用了历史 β，可以重写目标为：**

$$\beta^R_{i,t+k} = f(Z_t, \beta^R_{i,t-12}, \beta^R_{i,t-24}, \dots) + \epsilon_{i,t+k}$$

并让模型（例如 LSTM 或 Transformer）学习这种时间依赖关系。

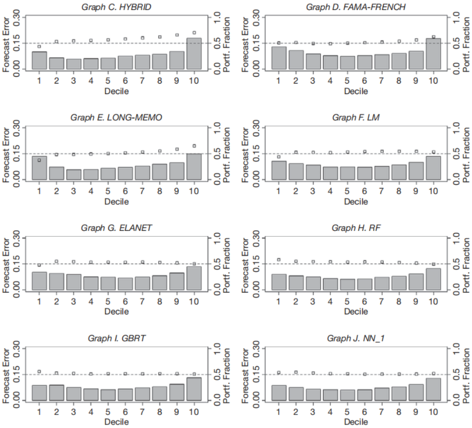
2. β 反映的是市场共动性，这往往受情绪与叙事驱动。可以构造情绪指标作为特征：

$$Z^\star_{i,t} = [Z_{i,t}, St]$$

其中 S_t 代表情绪或不确定性状态，可来自：

- **文本情绪指标**（如 RavenPack 新闻情绪、FinBERT 得分）；
- **市场不确定性指标**（如 VIX、Twitter 情绪指数）；**ESG 或政策风险事件指标**。

Appendix I - Average Forecast Errors of Portfolio Sorts Based on β



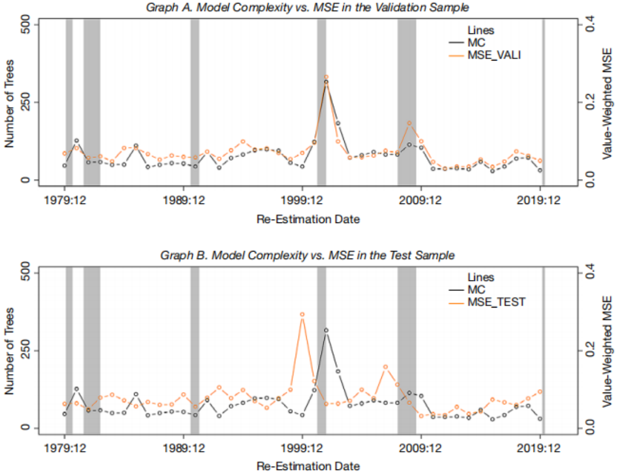
- Forecast error **distributions are more uniform** compared to classical ones.
- ML **does not systematically over/underestimatie** in the high/low-beta deciles.

Appendix II - MVP

Model	Std. Dev. (%)	DWND (%)	Min (%)	MaxDD (%)	TV (%)	Mean (%)	SR	β_{pv}	β_{mn}
	Panel A. Minimum Variance							*Panel B. Market Neutrality*	
Benchmark Estimators									
OLS_1Y_D	12.40	9.94	−24.24	43.55	27.98	9.12	0.74	0.39	−0.22
								(5.19)	(−6.41)
BSW	12.06	9.29	−21.91	43.07	27.29	9.01	0.75	0.37	−0.16
								(4.41)	(−4.01)
HYBRID	12.20	9.36	−21.88	41.12	26.24	8.93	0.73	0.36	−0.15
								(4.30)	(−2.96)
FAMA–FRENCH	12.02	9.52	−23.04	37.08	25.62	8.85	0.74	0.40	−0.07
								(7.95)	(−1.99)
LONG-MEMO	11.93	9.04	−20.35	34.73	28.96	9.15	0.77	0.36	−0.12
								(4.64)	(−3.23)
ML Estimators									
LM	11.71	8.75	−12.98	50.15	35.35	9.62	0.82	0.41	−0.13
								(4.36)	(−2.63)
ELANET	11.91	9.98	−22.34	49.34	30.94	9.32	0.78	0.41	−0.10
								(4.28)	(−1.96)
RF	11.42	8.31	−19.32	39.12	33.07	9.42	0.82	0.35	0.01
								(4.23)	(0.15)
GBRT	11.10	8.42	−18.82	38.85	42.41	10.01	0.90	0.36	0.00
								(4.24)	(−0.01)
NN_1	11.16	8.11	−16.28	35.08	37.51	9.70	0.87	0.35	−0.03
								(4.42)	(−0.61)

Appendix III - Model Complexity

More complex models are required when market β are more difficult to predict.



Appendix IV - Robustness

稳健性检验	结果
改变预测窗口（6M, 24M）	机器学习仍优于传统方法
替代 β 目标（SMB/HML 因子 β）	结果一致，机器学习预测更准
删除部分特征组	模型性能略降但排名稳定
子样本检验（危机期 vs 稳定期）	危机期机器学习优势更显著
不同误差指标（MSE, MSHE, MAE）	结论稳健一致