Attention Factors for Statistical Arbitrage

Elliot Epstein Rose Wang Jaewon Choi Markus Pelger

The 6th ACM International Conference on AI in Finance

This Paper: Statistical Arbitrage with Trading Costs

Statistical arbitrage

- Statistical arbitrage exploits temporal price difference between similar assets
- Prior work: Two-step approach
  1. Step: Identify similar assets (typically as residuals from factor model)
  2. Step: Residual portfolio weights based on time-series patterns
- **Problem**: High trading costs (turnover, short-selling), eroding net profitability

$$\epsilon_{n,t} = R_{n,t} - \beta_{n,t-1}^{\top} F_t, \qquad \epsilon_t^{\text{portfolio}} = \sum_{n=1}^{N} \omega_{n,t-1}^{\epsilon} \epsilon_{n,t}$$

Our approach:

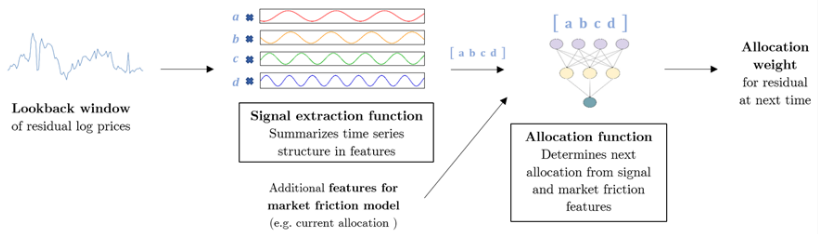
- One-step approach: Jointly estimate arbitrage factors and portfolio weights
- Objective function: portfolio performance after trading costs (net Sharpe ratio)
- Attention factors capture complex dependencies on firm characteristics
- Residual portfolio weights estimated with general sequence model

Results:

- Out-of-sample on 500 largest (most liquid) U.S. stocks from 1998-2021
- Gross Sharpe ratio **4.5**, net Sharpe ratio **2.3** (net of turnover and short-selling)

⇒ Best performing model in the literature under realistic trading frictions

Components: Arbitrage Portfolios, Signal and Allocation



Components of arbitrage trading:

1. Construct *arbitrage portfolio* as residual portfolios:

$$\epsilon_{n,t} = R_{n,t} - \beta_{n,t-1}^{\top} F_t$$

   - Factor models identify similar assets by similar exposures to risk factors
   - $\beta_{n,t-1}^{\top} F_t$ is "fair price" of $R_{n,t}$ and $\epsilon_{n,t}$ captures temporary mispricing

2. Extract arbitrage signal from time-series patterns of cumulative residuals

3. The arbitrage allocation assigns investment weights on residuals using the estimated signal.

Arbitrage Factors

Arbitrage portfolio:

$$\epsilon_{n,t} = R_{n,t} - \beta_{n,t-1}^{\top} F_t$$

- K factors F_t capture systematic risk.
- Loadings $\beta_{t-1} \in \mathbb{R}^{N_t \times K}$ are general function of information at time $t-1$.
- Factor models identify similar assets by similar exposures to risk factors

Factors are *tradable portfolios*:

$$F_t = \omega_{t-1}^F R_t, \qquad \omega_{t-1}^F \in \mathbb{R}^{K \times N}.$$

Candidate factor models:

1. Observed fundamental: Fama-French factors.
2. Statistical that explain correlations: PCA factors.
3. Conditional statistical where loadings are functions of firm characteristics: Instrumented PCA factors.

Our solution: Attention Factors

- Factor portfolio weights ω_{t-1}^F, are modeled with attention mechanism
- The attention mechanism learns embeddings of lagged firm characteristics
- Capture general dependencies and interactions with firm characteristics

$\Rightarrow$ Estimate attention factors with arbitrage trading objective!

Arbitrage Signal

Arbitrage Signal

$$\omega^{\text{port}}_{i,t-1} = \text{LongConv}_{\theta}(\epsilon_{i,(t-s,t-1)}),$$

- Key idea: Exploit predictable patterns in the time-series of residual portfolios
- Residual portfolio: Portfolio weight $\omega^{\text{port}}_{i,t-1}$ for trading residual i next period
- Sequence model: a 1-layer LongConv on the past s residuals $\epsilon_{i,(t-s,t-1)}$ estimates arbitrage signal

$\Rightarrow$ LongConv can capture complex time-series patterns

Arbitrage Trading:

- Residual mapping: Residuals are traded portfolios of stocks:

$$\epsilon_t = R_t - \beta^T_{t-1} F_t = R_t - \beta^T_{t-1} \omega^{\text{F}}_{t-1} R_t = \omega^{\epsilon}_{t-1} R_t,$$

- Arbitrage portfolio return: Residual portfolio $\omega^{\text{port}}_{t-1}$ maps into stock portfolio:

$$R^{\text{port}}_t = \epsilon^{\top}_t \omega^{\text{port}}_{t-1} = R^{\top}_t \left((\omega^{\epsilon}_{t-1})^{\top} \omega^{\text{port}}_{t-1} \right)$$

- Factor loadings: Under normalization $\beta_{t-1} = \omega^F_{t-1}$.

$\Rightarrow$ Functions to estimate: Factor weights ω^F_{t-1} and residual weights ω^{ϵ}_{t-1}

Arbitrage Trading Objective

Arbitrage Trading Objective

$$\max_{\omega^F,\omega^{\text{port}}} \underbrace{\frac{\bar{R}^{\text{port}}_{net} - R_f}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(R^{\text{port}}_{t,net} - \bar{R}^{\text{port}}_{net}\right)^2}}}_{\text{net Sharpe ratio}} + \lambda_{\text{Var}} \cdot \underbrace{\frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{\text{Var}(e^i)}{\text{Var}(R^i)}\right)}_{\text{explained variance}},$$

- Joint estimation of factors ω^F_{t-1} and arbitrage trading strategy ω^{ϵ}_{t-1}
- Economic objective of maximizing net Sharpe ratio and explaining variance
- Net return $R^{\text{port}}_{t,\text{net}} = R^{\text{port}}_t - \text{cost}(\omega_t, \omega_{t-1})$ for realistic transaction costs

$$\text{cost}(\omega_t, \omega_{t-1}) = 5 \times 10^{-4} \|\omega_t - \omega_{t-1}\|_1 + 10^{-4} \|\max(-\omega_t, 0)\|_1.$$

- Dual objective: λ_{Var} balances explained variance and net Sharpe ratio: Including explained variance is necessary for identifying latent factors
- Special case are conditional latent factors that maximize explained variance

Data and Empirical Implementation

Out-of-sample analysis on U.S. equity data

- 24 years of large cap U.S. daily stock returns from Jan 1998-2021
- Only largest 500 (most liquid) U.S. stocks
- Firm characteristics $M = 79$: past returns, value, investment,...

Implementation:

- All results are out-of-sample
- Train model on rolling windows of 8 years and retrain every year
- Lookback window of $L = 30$ days of returns as input for daily trading

Baseline Comparison Models:

1. **PCA+LongConv**: Two-Step approach
   a) Latent PCA factors ω^{PCA}.
   b) Residual weights ω^{port} estimated with LongConv
2. **PCA+OU**: Parametric benchmark (Avellaneda and Lee (2010))
   a) Latent PCA factors ω^{PCA}.
   b) Residual weights ω^{OU} estimated with parametric threshold rule

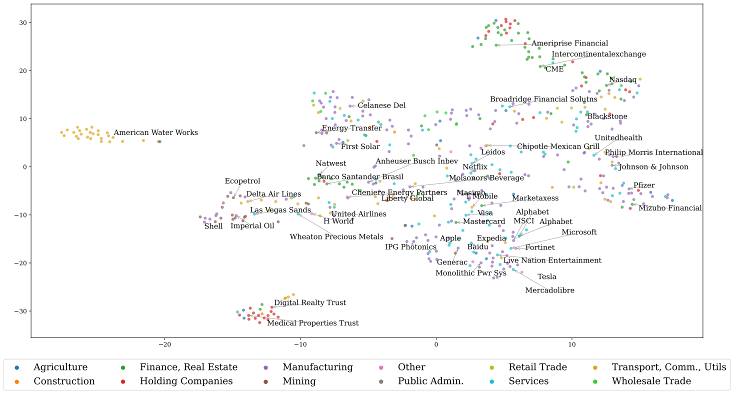
OOS Annualized Performance

Factors K	SR	μ	σ	$SR_{\mathbf{net}}$	$\mu_{\mathbf{net}}$	$\sigma_{\mathbf{net}}$
1	**3.05**	14.45	4.74	**1.68**	7.94	4.72
3	**3.05**	14.91	4.89	**1.69**	8.25	4.87
5	**2.92**	14.21	4.87	**1.58**	7.66	4.85
8	**3.35**	15.70	4.68	**1.94**	9.05	4.66
15	**3.81**	16.66	4.37	**2.25**	9.78	4.35
30	**3.97**	16.66	4.20	**2.28**	9.52	4.18
100	**4.52**	16.45	3.64	**2.19**	7.93	3.62
Market	0.42	8.61	20.37	0.42	8.61	20.37
PCA+OU (K=3)	1.26	4.18	3.33	-2.72	-9.04	3.33
PCA+LongConv (K=3)	2.76	14.61	5.30	1.57	8.29	5.28

Out-of-sample model performance for different number of attention factors

- New literature standard: best performance under realistic trading frictions
- Uncorrelated with market and other risk factors
- Strategy adjusts optimally to trading costs for high net performance
- Weak factors are important for arbitrage trading
- Simple PCA factors cannot adjust the factor construction to trading frictions

Interpretation: Factor Betas

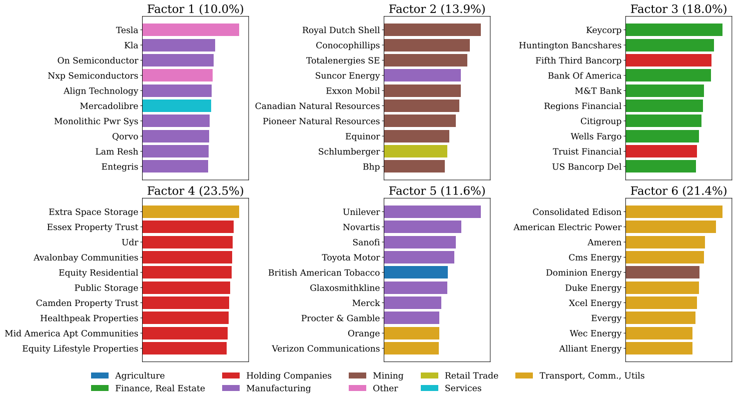


Interpretation of Attention Factor Betas (8 factors) on 2D (t-SNE) plot on Jan 1. 2021

- Use t-SNE to represent closeness of firms in the loading space
- Firms in similar industries are grouped together
- Upper right: banks and financial firms; lower right: energy,...

⇒ Learn industry patterns from data without explicit industry labels

Interpretation: Factor Portfolio Weights



Attention Weights for Factor Portfolio Weights on Jan. 1, 2021

- Factor weights have clear industry relationships
- Factor 1: technology; Factor 2: oil; Factor 3: finance,...

⇒ Attention factor structure has a clear economic interpretation

Drivers of Performance

Dropped Feature	SR	μ	σ	SR_{net}	μ_{net}	Beta
baseline (none excluded)	**3.97** (0.13)	16.66	4.20	**2.28**	9.52	0.05
past returns	1.50 (0.07)	7.82	5.23	0.59	3.09	0.08
investment	3.88 (0.17)	17.93	4.63	2.19	10.06	0.06
profitability	3.94 (0.15)	18.39	4.67	2.26	10.48	0.05
intangibles	3.91 (0.15)	18.18	4.65	2.24	10.34	0.06
value	4.08 (0.12)	18.45	4.53	2.32	10.44	0.04
trading frictions	2.90 (0.14)	13.36	4.61	1.34	6.14	0.06

Out-of-sample model performance when dropping characteristic groups

$\Rightarrow$ Past returns and trading frictions are driving performance

$\Rightarrow$ Robust performance across seeds of model estimation

Key Takeaways

Methodological Contribution

- Joint estimation of factors and arbitrage trading policy
- Attention Factors are conditional latent factors for arbitrage trading
- Joint estimation to maximize profitability after trading costs.

Empirical Results

- Comprehensive empirical study of 500 largest US equities over 24 years
- Unprecedented net Sharpe ratio of 2.3 and gross Sharpe ratio above 4
- Weak factors are important for arbitrage trading
- Loadings cluster by industry giving insight into the learned structure
- Performance is driven by return characteristics

$\Rightarrow$ Best performing model in the literature under realistic trading frictions