NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

STORM: A Spatio-Temporal Factor Model Based on Dual Vector Quantized Variational Autoencoders for Financial Trading

Yilei Zhao, Wentao Zhang, Tingran Yang, Yong Jiang, Fei Huang, Wei Yang Bryan Lim

Nanyang Technological University, Alibaba Group

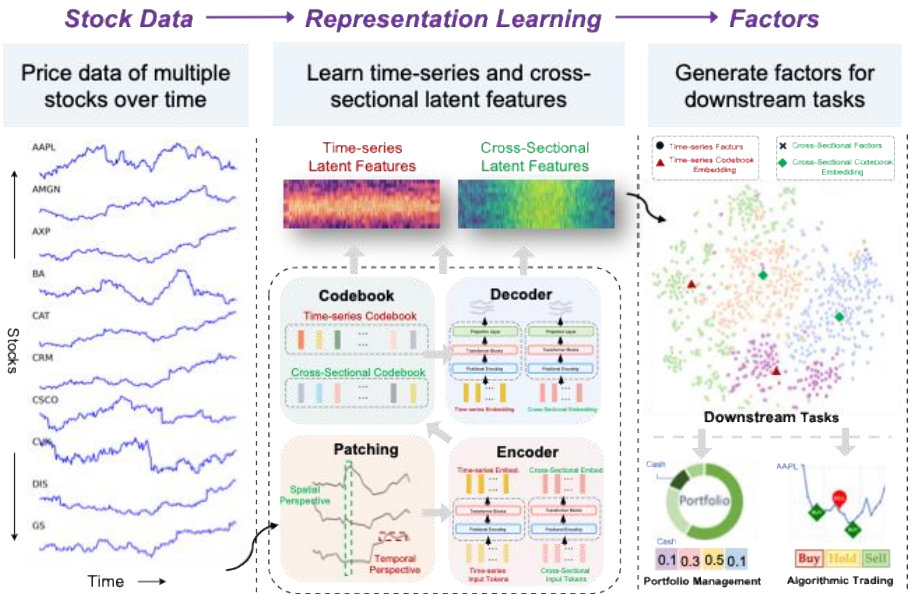
Content

- Introduction
- Problem Formulation
- Methodology
- Empirical Analysis
- Conclusion and Future Work

Introduction – Background of Factor models

Traditional Factor Models. In financial trading, factor models are fundamental tools for asset pricing and are widely used to predict asset returns. These models enhance pricing accuracy and risk management by identifying a set of key factors that explain excess returns.

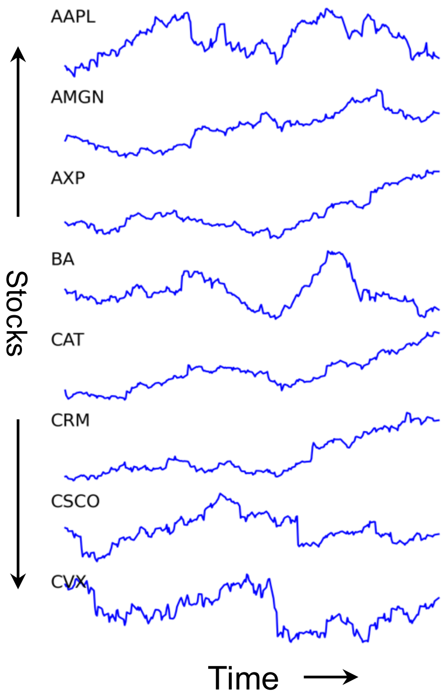
Latent Factor Models, connecting the factor model with the generative model, the variational autoencoder (VAE).



High-dimensional data (prices) → self-adaptively → Low-dimensional representations (factors)

Introduction – Challenges & Motivations

Stock Data



Factor Zoos

$$\text{ROE} = \frac{\text{Net Income}}{\text{Average Shareholder's Equity}}$$

$$\text{Momentum}_{i,t} = \frac{\text{P}_{i,t-1}}{\text{P}_{i,t-12}} - 1$$

$$\text{B/M Ratio} = \frac{\text{Book Value of Equity}}{\text{Market Capitalization}}$$

$$\text{E/P Ratio} = \frac{\text{Earnings Per Share (EPS)}}{\text{Price Per Share}}$$

…

$$R_{i,t} = \alpha_i + \sum_{j=1}^{K} \beta_{i,j} F_{j,t} + \epsilon_{i,t}$$

Challenge 1: Limited Reflection of Market Complexity. Single-value factor representations struggle to capture the non-linear nature of financial data, leading to reduced predictive stability.

Introduction – Challenges & Motivations

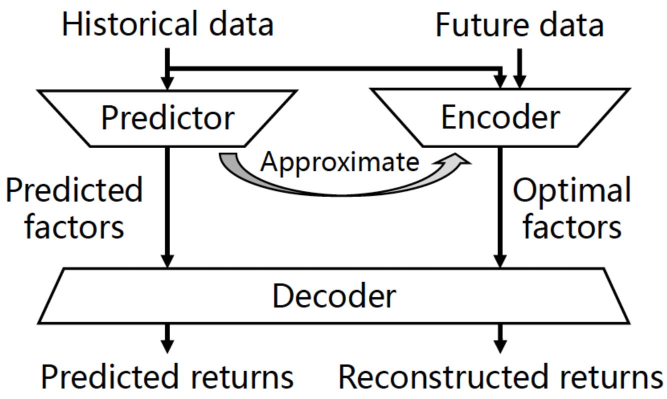
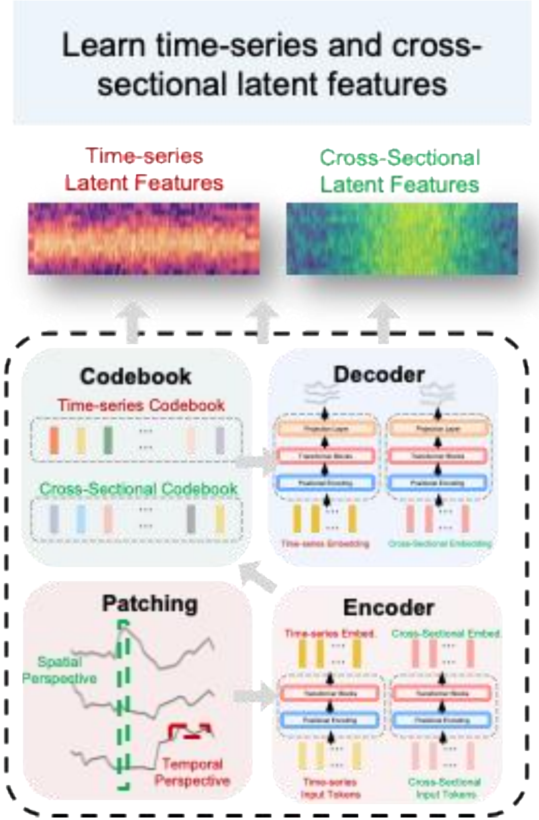


Fig from FactorVAE: A probabilistic dynamic factor model based on variational autoencoder for predicting cross-sectional stock returns. Duan Y, et al, 2022.

Challenge 2: Factor Inefficiency. VAE-based factors often ignore temporal dynamics, succumb to latent space noise, and suffer from multicollinearity due to a lack of factor independence.

Challenge 3: Lack of Factor Selection. Existing models prioritize factor generation over evaluation, failing to implement selection mechanisms necessary to identify and utilize the most impactful signals.

Introduction – Challenges & Motivations

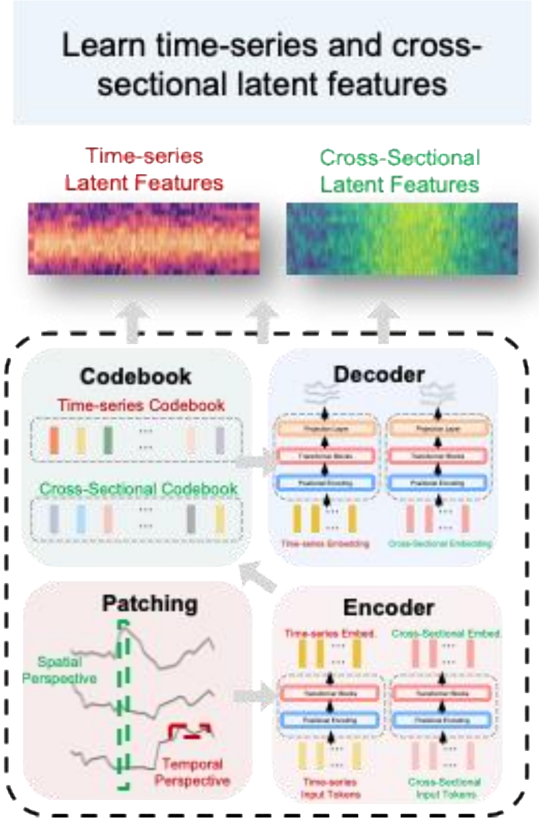


Challenge 1: Limited Reflection of Market Complexity. Single-value factor representations struggle to capture the non-linear nature of financial data, leading to reduced predictive stability.

➢ **High-Dimensional Vector Representation**: Captures intricate market complexity and non-linearity through high-dimensional latent vectors, overcoming the inherent constraints of traditional scalar-valued factors.

Introduction – Challenges & Motivations

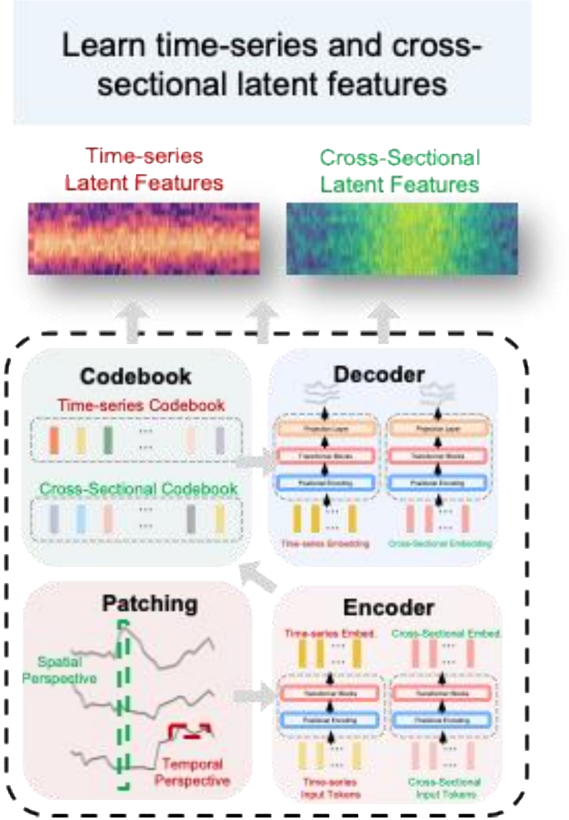
Alibaba-NTU Global e-Sustainability CorpLab



Challenge 2: Factor Inefficiency. VAE-based factors often ignore temporal dynamics, succumb to latent space noise, and suffer from multicollinearity due to a lack of factor independence.

➤ **Dual VQ-VAE Spatio-Temporal Architecture**: Integrates cross-sectional and time-series features using a dual VQ-VAE structure with diversity and orthogonality loss constraints to ensure factor efficacy and independence.

Introduction – Challenges & Motivations



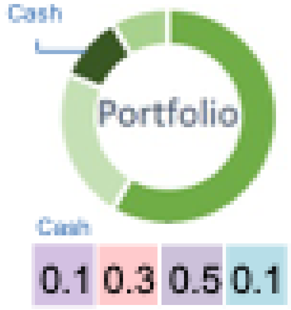
Challenge 3: Lack of Factor Selection. Existing models prioritize factor generation over evaluation, failing to implement selection mechanisms necessary to identify and utilize the most impactful signals.

- **Codebook-Based Categorization & Selection**: Employs discrete codebook embeddings as cluster centers to categorize and differentiate factors, providing a transparent mechanism for impactful factor selection.

Problem Formulation

Observed Data

Stock's historical price data $\boldsymbol{p}$ (open, high, low, and close) and technical indicators $\boldsymbol{d}$ as observed variables $\mathbf{x} \coloneqq [\boldsymbol{p}, \boldsymbol{d}] \in \mathbb{R}^{N\times W\times D}$. Then we use the D dimension data to predict the stock's future returns $y_{i,t+1} = \frac{p_{i,t+1}-p_{i,t}}{p_{i,t}}$.



Portfolio Management

Algorithmic Trading

Downstream tasks

- Portfolio Management: constructs an optimal portfolio based on predicted returns to test the profitability of factor models.
- Algorithmic Trading: trade on one single asset, aiming to find trading signals, and execute buy, hold and sell actions.

Method – Overall Structure

Alibaba-NTU Global e-Sustainability CorpLab

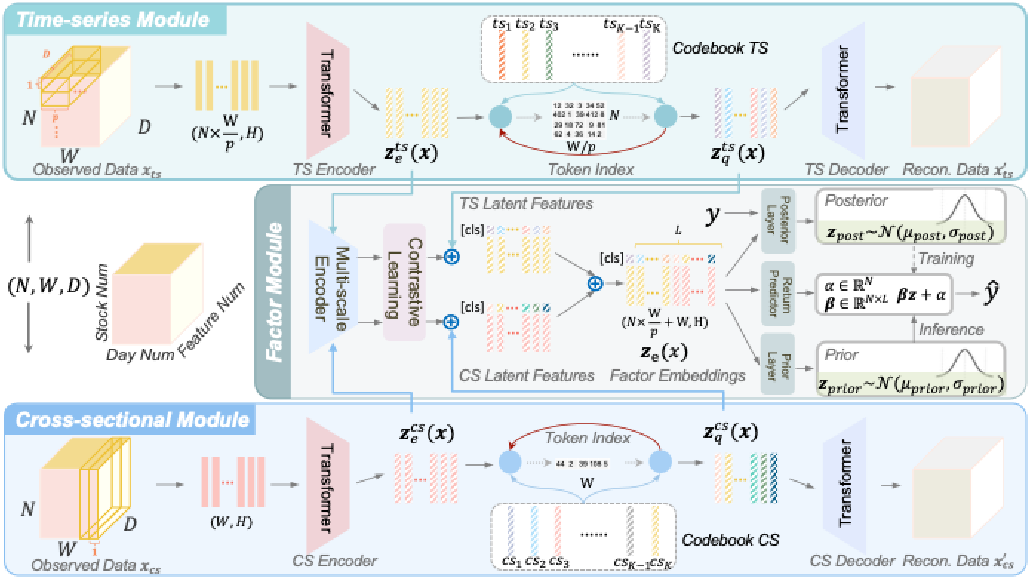
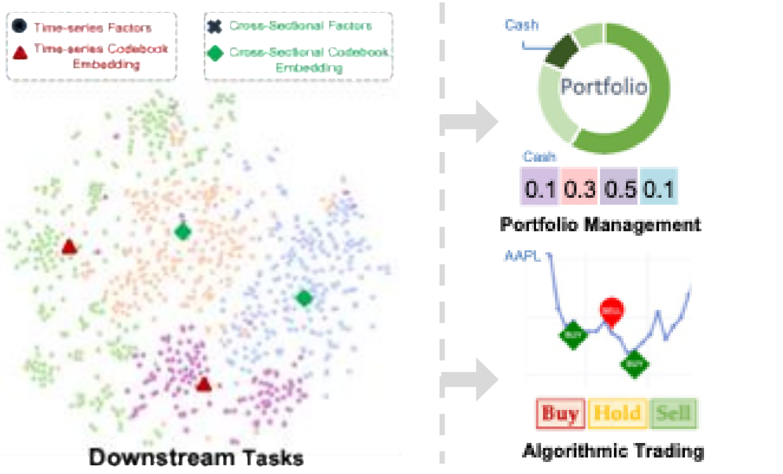


Figure 2: The architecture of our proposed STORM model

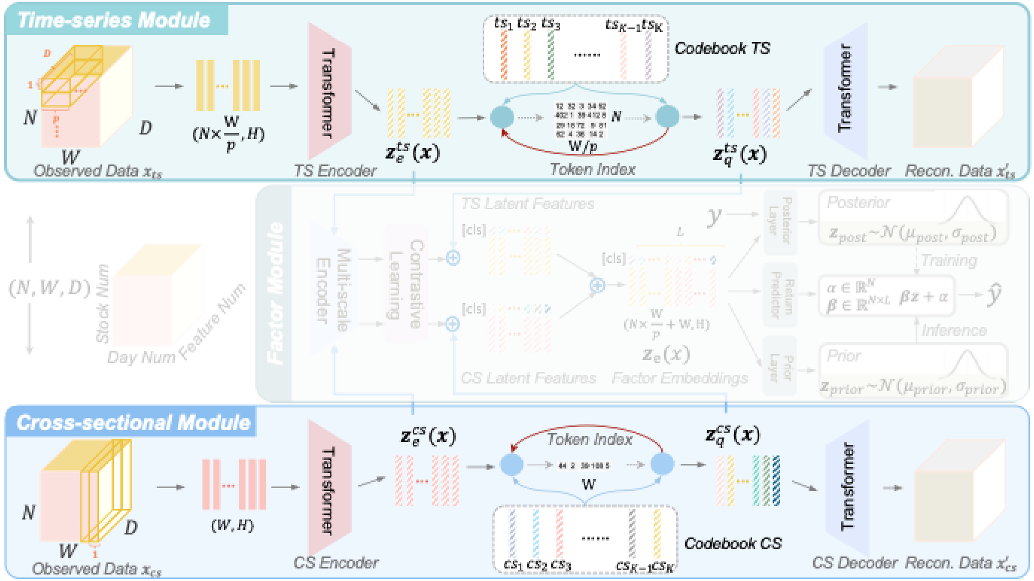
Then the learned factors will be used in downstream tasks, like portfolio management, and algorithmic trading.

Overall Structure

Our designed STORM, which has a dual VQ-VAE architecture that systematically constructs cross-sectional and time-series factors by capturing both spatial and temporal dependencies.



Method – TS and CS modules



Patching and Encoding

In TS module, observed data is divided along the stock number dimension. In CS module, it's divided along the time axis. Then we use Transformer as encoders and decoders to capture complex patterns.

Decoding and Reconstruction

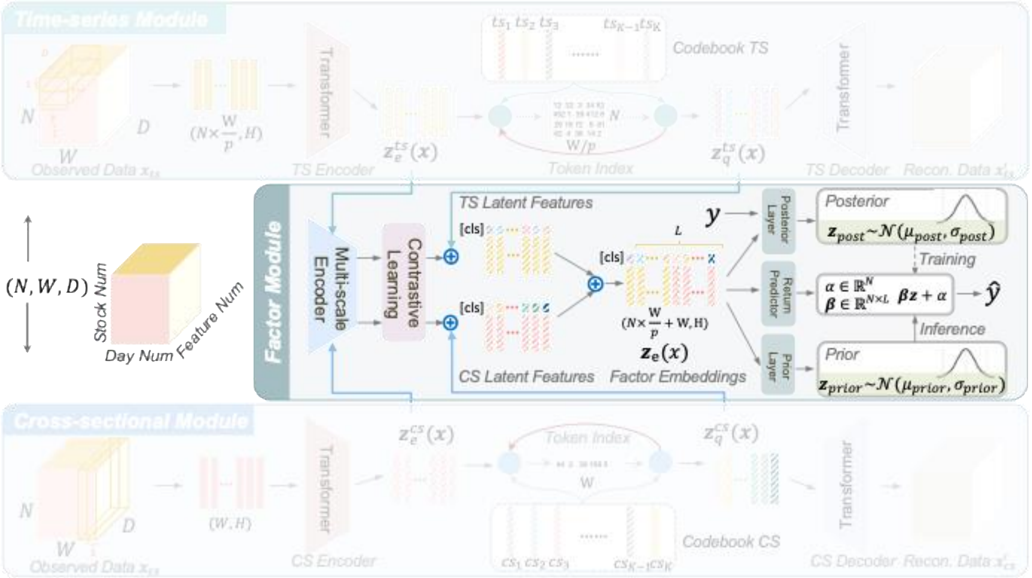
$$
\begin{aligned}
\mathcal{L}_1 =& \lambda_{\text{ortho}}\mathcal{L}_{\text{ortho}} + \lambda_{\text{div}}\mathcal{L}_{\text{div}} + \left\|\mathbf{x}-\mathbf{x}'_{ts}\right\|_2^2 + \left\|\mathbf{x}-\mathbf{x}'_{cs}\right\|_2^2 \\
&+ \left\|sg[\mathbf{z}_e^{ts}(\mathbf{x})] - \mathbf{z}_q^{ts}(\mathbf{x})\right\|_2^2 + \left\|sg[\mathbf{z}_q^{ts}(\mathbf{x})] - \mathbf{z}_e^{ts}(\mathbf{x})\right\|_2^2 \\
&+ \left\|sg[\mathbf{z}_e^{cs}(\mathbf{x})] - \mathbf{z}_q^{cs}(\mathbf{x})\right\|_2^2 + \left\|sg[\mathbf{z}_q^{cs}(\mathbf{x})] - \mathbf{z}_e^{cs}(\mathbf{x})\right\|_2^2
\end{aligned}
$$

Codebook Construction and Optimization

Diversity loss -> representational capacity. $\mathcal{L}_{\text{div}} = \frac{1}{GK}\sum_{g=1}^{G}\sum_{k=1}^{K}\bar{p}_{g,k}\log\bar{p}_{g,k},$

Orthogonality loss -> factor orthogonality. $\mathcal{L}_{\text{ortho}} = \frac{1}{K^2}\left\|\ell_2(\mathbf{e})^\top\ell_2(\mathbf{e}) - I_K\right\|_F^2$

Method – Factor Module



For the latent factor model, our proposed STORM with the dual VQ-VAE structure and the factor module can already learn the corresponding factors and then make future return predictions.

Feature Fusion and Alignment

We use multiscale encoder and contrastive learning layer to fuse and align TS and CS features at fine-grained and semantic levels.

Prior - Posterior Learning

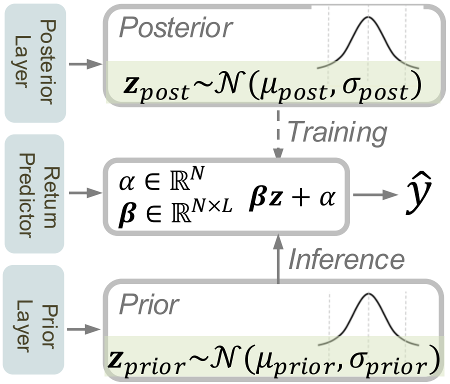
Concatenate two latent features, which are used to predict the future returns. The factors are then be used in portfolio and trading downstream tasks.

$$[\mu_{post}, \sigma_{post}] = \varphi_{FE}(y, \boldsymbol{z}_e(x))$$

$$[\mu_{prior}, \sigma_{prior}] = \varphi_{FP}(\boldsymbol{z}_e(x))$$

$$\hat{\boldsymbol{y}} = \boldsymbol{\alpha} + \sum_{k=1}^{K} \boldsymbol{\beta}^k \boldsymbol{z}^k + \boldsymbol{\epsilon}$$

Method – How to apply to downstream tasks



$$\hat{\boldsymbol{y}} = \boldsymbol{\alpha} + \sum_{k=1}^{K} \boldsymbol{\beta}^k \boldsymbol{z}^k + \boldsymbol{\epsilon}$$

MDP 5-tuple: $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$

Downstream Tasks

Portfolio Management Task: utilize the factor decoder network to generate stock future returns $\boldsymbol{y}$, and then apply the *TopK-Drop* strategy to construct a daily portfolio by selecting the top k stocks.

TopK-Drop strategy: At trading day t, $\mathbf{w_t} = [w_t^0, w_t^1, \dots, w_t^{N-1}]$ represents the portfolio weights for N stocks. Then, for day t and $t+1$, at most d stocks can be replaced. The replacement is defined as a stock whose weight changes from 0 (not held) to a positive value (newly bought) or from a positive value to 0 (fully sold). This constraint ensures that at least $k-d$ stocks remain in the portfolio, reducing excessive turnover.

Algorithmic Trading Task: executes buy, hold, and sell actions based on predicted asset states to balance returns and risks. We formulate it as a Markov Decision Process under reinforcement learning scenario. The latent factor embeddings $\mathbf{Z}$ are integrated into the observation set $\mathcal{O} = \{\mathbf{Z}, \mathcal{R}\}$, and $\mathcal{R}$ is the reward function that guides the agent's learning.

Empirical Analysis

Three Tasks:

- ✓ Future Return Prediction Task
  To evaluate the factor effectiveness and predictive power
- ✓ Portfolio Management & Algorithmic Trading
  Financial downstream tasks to evaluate profitability of learned factor

Metrics:

- ✓ Factor Correlated Criteria
  RankIC: correlation between the predicted rankings and the actual returns
  RankICIR: the information ratio of RankIC, measuring the stability of prediction
- ✓ Financial Standard Metrics
  Profit: Annualized percentage yield (APY), cumulative wealth(CW)
  Risk: Maximum drawdown (MDD) and annualized volatility (AVO)
  Risk-Adjusted Profit: Calmar ratio (CR) and annualized Sharpe ratio (ASR)

Datasets:

Two US stock markets, from 2008-04-01 to 2024-03-31, spanning 16 years.

- ✓ SP500
- ✓ DJ30

Training: from 2008-04-01 to 2021-03-31
Test: from 2021-04-01 to 2024-03-31

For algorithmic trading task, we choose 5 typical individual stocks inside the US stock markets.

Empirical Analysis

← *Results on Portfolio Management*

SP500 Dataset						
Strategies	Profit		Risk-Adj. Profit		Risk	
	APY↑	CW↑	CR↑	ASR↑	MDD↓	AVO↓
Market Index	0.058	1.184	0.228	0.142	0.254	0.410
LightGBM	0.059 ± 0.132	1.201 ± 0.456	0.304 ± 0.753	0.332 ± 0.755	0.238 ± 0.108	0.176 ± 0.008
LSTM	0.069 ± 0.020	1.221 ± 0.068	0.278 ± 0.065	0.371 ± 0.114	0.248 ± 0.042	0.186 ± 0.013
Transformer	0.076 ± 0.028	1.244 ± 0.098	0.389 ± 0.165	0.433 ± 0.161	0.198[1] ± 0.057	0.174 ± 0.004
CAFactor	0.075 ± 0.028	1.241 ± 0.100	0.342 ± 0.229	0.428 ± 0.165	0.223 ± 0.043	0.174 ± 0.006
FactorVAE	0.079 ± 0.025	1.256 ± 0.085	0.404 ± 0.177	0.460 ± 0.145	0.200 ± 0.040	0.173 ± 0.007
HireVAE	0.077 ± 0.029	1.249 ± 0.104	0.361 ± 0.180	0.448 ± 0.189	0.216 ± 0.048	0.172 ± 0.007
STORM	**0.188**[2] ± 0.055	**1.683** ± 0.226	**1.189** ± 0.661	**1.052** ± 0.329	0.166 ± 0.050	0.171 ± 0.020
STORM-w/o-TS	0.090 ± 0.074	1.300 ± 0.277	0.503 ± 0.390	0.570 ± 0.339	0.181 ± 0.043	0.175 ± 0.025
STORM-w/o-CS	0.089 ± 0.045	1.294 ± 0.163	0.623 ± 0.385	0.592 ± 0.252	**0.146** ± 0.073	**0.167** ± 0.020
Improvement(%)[3]	137.97	34.00	194.31	128.70	26.26	2.91

DJ30 Dataset						
Strategies	Profit		Risk-Adj. Profit		Risk	
	APY↑	CW↑	CR↑	ASR↑	MDD↓	AVO↓
Market Index	0.063	1.201	0.147	0.429	0.219	0.288
LightGBM	0.069 ± 0.040	1.221 ± 0.140	0.288 ± 0.262	0.430 ± 0.267	0.244 ± 0.047	0.160 ± 0.007
LSTM	0.060 ± 0.007	1.192 ± 0.023	0.243 ± 0.021	0.370 ± 0.043	0.248 ± 0.006	0.163 ± 0.005
Transformer	0.056 ± 0.013	1.179 ± 0.043	0.227 ± 0.067	0.367 ± 0.085	0.250 ± 0.033	0.154 ± 0.002
CAFactor	0.059 ± 0.015	1.186 ± 0.050	0.233 ± 0.058	0.382 ± 0.092	0.252 ± 0.021	0.153 ± 0.001
FactorVAE	0.076 ± 0.038	1.246 ± 0.135	0.352 ± 0.349	0.480 ± 0.235	0.225 ± 0.063	0.159 ± 0.005
HireVAE	0.072 ± 0.037	1.233 ± 0.132	0.298 ± 0.253	0.445 ± 0.256	0.247 ± 0.057	0.163 ± 0.012
STORM	**0.148** ± 0.046	**1.517** ± 0.188	**1.396** ± 0.679	**1.052** ± 0.297	**0.108** ± 0.026	**0.140** ± 0.014
STORM-w/o-TS	0.079 ± 0.057	1.259 ± 0.192	0.621 ± 0.425	0.603 ± 0.379	0.127 ± 0.038	0.142 ± 0.008
STORM-w/o-CS	0.073 ± 0.051	1.236 ± 0.169	0.533 ± 0.400	0.566 ± 0.343	0.138 ± 0.034	**0.140** ± 0.017
Improvement(%)	94.74	21.75	296.59	119.17	52.00	8.50

1 Underline indicates the best-performing baseline method result;
2 **Bold** indicates the best performance among our proposed models (including variants);
3 Improvement of STORM over the best-performing baselines.

↓ *Results on Algorithmic Trading*

Models	AAPL			JPM			IBM			INTC			MSFT		
	APY↑	CW↑	CR↑	APY↑	CW↑	CR↑	APY↑	CW↑	CR↑	APY↑	CW↑	CR↑	APY↑	CW↑	CR↑
Buy&Hold	0.120	1.404	0.383	0.096	1.316	0.236	0.145	1.499	0.727	-0.117	0.690	-0.184	0.214	1.784	0.569
LightGBM	0.135	1.390	0.487	0.116	1.335	0.333	0.227	1.654	1.091	-0.042	0.880	0.038	0.267	2.068	0.637
LSTM	0.053 ± 0.015	1.152 ± 0.042	0.283 ± 0.101	0.079 ± 0.019	1.290 ± 0.053	0.266 ± 0.023	0.134 ± 0.141	1.386 ± 0.406	0.754 ± 0.69	0.060 ± 0.119	1.262 ± 0.429	0.381 ± 0.62	0.178 ± 0.122	1.513 ± 0.352	0.893 ± 0.822
Transformer	0.083 ± 0.099	1.240 ± 0.284	0.512 ± 0.864	0.133 ± 0.07	1.384 ± 0.201	0.614 ± 0.509	0.131 ± 0.072	1.377 ± 0.207	0.782 ± 0.541	0.079 ± 0.174	1.290 ± 0.68	0.458 ± 0.824	0.138 ± 0.118	1.397 ± 0.338	0.726 ± 0.643
DQN	0.135 ± 0.075	1.374 ± 0.226	0.510 ± 0.239	0.105 ± 0.132	1.305 ± 0.383	0.607 ± 0.673	0.139 ± 0.055	1.400 ± 0.158	0.802 ± 0.270	0.061 ± 0.106	1.185 ± 0.305	0.442 ± 0.436	0.166 ± 0.035	1.475 ± 0.100	0.534 ± 0.107
SAC	0.147 ± 0.021	1.509 ± 0.060	0.528 ± 0.052	0.131 ± 0.038	1.383 ± 0.111	0.400 ± 0.198	0.207 ± 0.057	1.598 ± 0.163	1.170 ± 0.335	0.056 ± 0.210	1.165 ± 0.6	0.353 ± 0.994	0.229 ± 0.114	1.656 ± 0.331	0.929 ± 0.585
PPO	0.137 ± 0.073	1.379 ± 0.209	0.496 ± 0.213	0.128 ± 0.026	1.372 ± 0.071	0.356 ± 0.059	0.146 ± 0.047	1.422 ± 0.136	0.779 ± 0.205	-0.019 ± 0.177	0.954 ± 0.514	0.040 ± 0.418	0.216 ± 0.081	1.620 ± 0.232	0.569 ± 0.105
STORM	**0.229** ± 0.033	**1.857** ± 0.154	0.750 ± 0.066	**0.174** ± 0.032	**1.621** ± 0.133	0.559 ± 0.081	**0.236** ± 0.039	**1.893** ± 0.184	**1.470** ± 0.445	**0.173** ± 0.067	**1.625** ± 0.284	**0.773** ± 0.293	**0.290** ± 0.052	**2.154** ± 0.262	**1.216** ± 0.597
STORM-w/o-TS	0.199 ± 0.059	1.730 ± 0.260	**0.766** ± 0.274	0.127 ± 0.058	1.437 ± 0.214	**0.627** ± 0.187	0.154 ± 0.023	1.536 ± 0.089	1.294 ± 1.012	0.106 ± 0.120	1.212 ± 0.145	0.267 ± 0.135	0.254 ± 0.060	1.979 ± 0.093	0.964 ± 0.503
STORM-w/o-CS	0.148 ± 0.069	1.521 ± 0.283	0.641 ± 0.184	0.103 ± 0.043	1.347 ± 0.154	0.350 ± 0.106	0.152 ± 0.054	1.534 ± 0.219	0.814 ± 0.352	0.136 ± 0.082	1.474 ± 0.300	0.541 ± 0.328	0.181 ± 0.051	1.650 ± 0.221	0.648 ± 0.361
Improvement(%)	55.782	23.062	45.076	30.827	17.124	2.117	3.965	14.4501	20.408	118.987	26.969	66.594	8.614	4.159	30.893

Models	AAPL			JPM			IBM			INTC			MSFT		
	ASR↑	MDD↓	AVO↓	ASR↑	MDD↓	AVO↓	ASR↑	MDD↓	AVO↓	ASR↑	MDD↓	AVO↓	ASR↑	MDD↓	AVO↓
Buy&Hold	0.447	0.313	0.269	0.405	0.406	0.237	0.679	0.199	0.213	-0.324	0.635	0.359	0.777	0.376	0.275
LightGBM	0.950	0.309	0.017	0.921	0.386	0.015	1.822	0.175	0.011	0.098	0.553	0.023	1.456	0.370	0.017
LSTM	0.623 ± 0.163	0.244 ± 0.072	0.012 ± 0.003	0.801 ± 0.185	0.335 ± 0.061	0.012 ± 0.002	1.268 ± 0.924	0.163 ± 0.017	0.010 ± 0.001	0.287 ± 1.02	0.149 ± 0.101	0.010 ± 0.007	1.345 ± 0.842	0.246 ± 0.124	0.014 ± 0.003
Transformer	0.972 ± 0.748	0.239 ± 0.187	0.010 ± 0.006	1.332 ± 0.607	0.226 ± 0.100	0.010 ± 0.003	1.336 ± 0.688	0.182 ± 0.100	0.010 ± 0.003	0.604 ± 0.789	0.257 ± 0.194	0.014 ± 0.006	1.206 ± 0.754	0.182 ± 0.134	0.011 ± 0.006
DQN	0.974 ± 0.472	0.288 ± 0.159	0.016 ± 0.007	1.085 ± 1.270	0.212 ± 0.118	0.010 ± 0.003	1.415 ± 0.403	0.161 ± 0.049	0.009 ± 0.004	0.609 ± 0.584	0.243 ± 0.310	0.014 ± 0.009	1.194 ± 0.148	0.309 ± 0.079	0.015 ± 0.004
SAC	1.054 ± 0.223	0.292 ± 0.022	0.016 ± 0.002	1.044 ± 0.276	0.350 ± 0.085	0.014 ± 0.001	1.794 ± 0.458	0.157 ± 0.068	0.011 ± 0.005	0.527 ± 1.061	0.359 ± 0.191	0.017 ± 0.004	1.638 ± 0.736	0.229 ± 0.215	0.013 ± 0.005
PPO	0.974 ± 0.412	0.300 ± 0.030	0.016 ± 0.001	1.000 ± 0.189	0.381 ± 0.018	0.015 ± 0.001	1.386 ± 0.231	0.177 ± 0.027	0.011 ± 0.003	-0.044 ± 1.689	0.511 ± 0.115	0.019 ± 0.007	1.276 ± 0.299	0.353 ± 0.036	0.017 ± 0.001
STORM	**1.346** ± 0.162	0.282 ± 0.000	0.016 ± 0.001	1.245 ± 0.106	0.299 ± 0.016	0.014 ± 0.001	**2.107** ± 0.302	0.137 ± 0.033	0.009 ± 0.001	**1.189** ± 0.386	**0.227** ± 0.055	**0.015** ± 0.002	**1.672** ± 0.201	**0.210** ± 0.067	**0.015** ± 0.002
STORM-w/o-TS	1.199 ± 0.300	0.262 ± 0.052	0.017 ± 0.001	**1.587** ± 0.296	**0.194** ± 0.038	**0.011** ± 0.001	1.868 ± 0.956	**0.131** ± 0.077	**0.008** ± 0.003	0.580 ± 0.229	0.344 ± 0.061	0.017 ± 0.000	1.553 ± 0.187	0.244 ± 0.093	**0.015** ± 0.004
STORM-w/o-CS	1.027 ± 0.326	**0.229** ± 0.030	**0.015** ± 0.001	0.858 ± 0.236	0.327 ± 0.064	0.013 ± 0.001	1.397 ± 0.405	0.177 ± 0.028	0.010 ± 0.001	0.914 ± 0.397	0.312 ± 0.136	0.018 ± 0.004	1.401 ± 0.224	0.295 ± 0.090	0.016 ± 0.002
Improvement(%)	27.704	4.184	-	19.144	8.491	-	15.642	16.561	11.111	95.238	-	-	2.076	-	-

Empirical Analysis

Table 4: Factor quality evaluation task results on RankIC and RankICIR

Models	SP500 RankIC↑	SP500 RankICIR↑	DJ30 RankIC↑	DJ30 RankICIR↑
Future Return Prediction in Portfolio Management Task				
LightGBM	0.027 ± 0.006	0.274 ± 0.084	0.031 ± 0.005	0.272 ± 0.049
LSTM	0.034 ± 0.006	0.333 ± 0.042	0.031 ± 0.004	0.329 ± 0.056
Transformer	0.035 ± 0.007	0.340 ± 0.078	0.033 ± 0.005	0.343 ± 0.045
CAFactor	0.037 ± 0.005	0.356 ± 0.084	0.040 ± 0.003	0.380 ± 0.043
FactorVAE	0.052 ± 0.010	0.543 ± 0.122	0.056 ± 0.012	0.520 ± 0.081
HireVAE	0.057 ± 0.006	0.558 ± 0.058	0.058 ± 0.006	0.563 ± 0.053
STORM	**0.062** ± 0.018	**0.673** ± 0.155	**0.065** ± 0.038	**0.668** ± 0.287
STORM-w/o-TS	0.053 ± 0.017	0.513 ± 0.145	0.055 ± 0.015	0.563 ± 0.127
STORM-w/o-CS	0.054 ± 0.014	0.522 ± 0.118	0.053 ± 0.016	0.559 ± 0.156
Improvement(%)[1]	8.772	20.609	12.069	18.650

We compare STORM with other latent factor models, and use RankIC and RankICIR metrics to show the factor correlations.

To further show the quality of learned factors, we represent in Figure 3(a) the distribution of factor usage in downstream tasks.

In Figure 3(b), we show the sensitivity results of hyperparameter codebook size. This variable balance the factor diversity and representational quality.

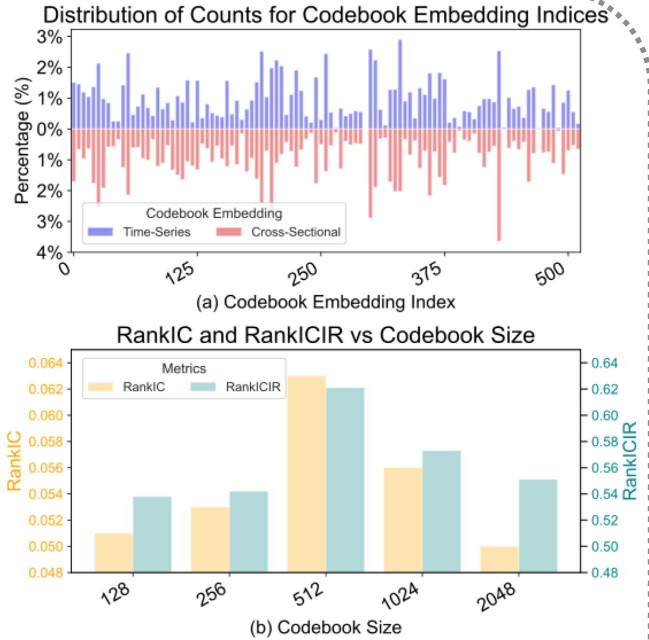


Figure 3: (a) Distribution of counts for codebook embedding indices. (b) Hyperparameter experiment results for different codebook sizes.

Conclusion and Future Work

Summary

- Propose a spatio-temporal latent factor model to learn both time-series and cross-sectional factors
- Ensure the factor orthogonality and diversity
- Factor selection in financial trading

Future Work

- More side information.
- Exogenous factors