MULTI-HORIZON UNCERTAINTY-AWARE LATENT SCENARIO LEARNING FOR REAL-TIME SEQUENTIAL FINANCIAL ANALYTICS AND DECISION-MAKING WITH ASYNCHRONOUS DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Machine learning in the finance domain always involves multi-task sequential quantitative analytics and decision-making, in which research analysts extract predictive signals from diverse firm characteristics, form prospects of future financial disclosure and valuations of stock prices, then construct investment portfolios based on their return expectations. In reality, they have to deal with asynchronous financial ìdata and thereby obsolete information, e.g., accounting-based fundamentals are released quarterly or annually with lags while market-based characteristics such as trading activities can be collected in a more timely manner. This imposes huge challenges in the sequential investment process, where input uncertainty can propagate through multi-task learning, leading to (i) distorted beliefs about firm fundamentals in upstream data analytics, (ii) misinformed return expectations in midstream asset pricing, and (iii) suboptimal downstream portfolio decisions. To tackle these challenges, we propose a novel unified multi-horizon uncertainty-aware latent scenario learning network that simultaneously (i) infers real-time multi-horizon probabilistic firm fundamentals jointly from the stale prints and the latest market data, (ii) estimates probability-weighted systematic risk exposures to common risk factors that produce realistic multi-horizon cross-sectional expected returns, and (iii) constructs various multi-horizon robust investment portfolios, which take the generative latent scenarios into account in optimization. Extensive experiments, ablation studies, and sensitivity analyses on the most comprehensive open-source stock market dataset show that our proposed model consistently outperforms seminal baselines in relevant literature across all evaluation metrics for sequential financial tasks. The results also emphasize the importance of real-time uncertainty-aware design of the latent scenario learning mechanism.

1 INTRODUCTION

Applications of machine learning have shown economic gains in various financial tasks in the investment management industry. However, these financial tasks are not independent but interconnected, suggesting that the input uncertainty can be propagated through multi-task learning. For example, when extracting economic insights from a large set of financial data, research analysts have to deal with asynchronous firm fundamentals, as these financial disclosures are scheduled quarterly or annual with lags while financial market-based characteristics such as prices and volumes can be collected in a more timely manner. Huge challenges arise from inferring prospective firm fundamentals from high-dimensional mixed-frequency data for downstream stock valuation (asset pricing) and investment (portfolio selection) tasks.

Financial markets are forward-looking, investment decisions are made based on subjective beliefs about the financial prospects of firms when evaluating stock prices, and the time-varying market overall assessment determines dynamics of asset prices. In the sequential investment process, if stale accounting-based fundamentals as a part of firm characteristics are fed into the downstream asset pricing or portfolio selection models, we essentially incorporate distorted beliefs about firm

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

fundamentals into the investment valuations and decisions, which, in turns, lead to misinformed return expectations and suboptimal asset allocation. However, in reality, it is almost impossible to gain access to real-time metrics or a full spectrum of investors' subjective beliefs about firm fundamentals for stock valuations or portfolio construction.

Formally, standard cross-sectional factor pricing models characterize the relation between firm states and expected returns as $r_{i,t+1} = \beta_{i,t}^\top f_{t+1} + \epsilon_{i,t+1}$ and $\mathbb{E}_t[r_{i,t+1}] = \beta_{i,t}^\top \lambda_t$, where f_{t+1} denotes common factor realizations, $\beta_{i,t}$ measures asset i's exposure to those factors, and $\lambda_t = \mathbb{E}_t[f_{t+1}]$ is the conditional factor risk premium. Neural asset pricing models typically parameterize exposures as a nonlinear function of firm characteristics, $\beta_{i,t} = g(X_{i,t-1})$, which means asynchronous characteristics directly distort pricing through the exposure function. At any decision time t, the characteristic panel contains genuinely missing entries, not-yet-released measurements, and stale values carried forward from earlier disclosures. Simple heuristics such as last-observation-carried-forward or cross-sectional mean imputation all collapse the state distribution to a single point estimate, as do deterministic learned completion methods. For nonlinear exposure functions, this introduces a systematic bias: in general $\mathbb{E}[g(X_{i,t}) \mid \mathcal{I}_t] \neq g(\mathbb{E}[X_{i,t} \mid \mathcal{I}_t])$, so collapsing the state distribution eliminates exposure uncertainty before it can propagate through factor premia to expected returns and portfolio weights. Independently fitted horizon-specific models further discard shared structure across one- to twelve-month returns, eroding term structure consistency.

We propose a unified multi-horizon uncertainty-aware latent scenario learning framework for sequential financial analytics and decision-making. The framework operates through three tightly coupled stages that systematically address each layer of cascading bias. First, a disclosure-aware information layer organizes characteristics by economic category and disclosure frequency, conditions on disclosure status and staleness, and generates plausible contemporaneous firm-state scenarios via selectively anchored latent diffusion. Second, a scenario-conditioned multi-horizon pricing layer decomposes returns into common factor premia and firm-specific exposures: FactorNet learns latent risk factors and a shared low-rank horizon decoder exclusively from historical returns, enforcing strict causal validity; BetaNet then maps each state scenario to scenario-specific factor loadings, and the frozen decoder converts their interaction with rolling factor premia into multi-horizon return forecasts. Third, a portfolio optimization stage translates the resulting return scenario distributions into robust investment decisions. Our contributions are threefold:

- We formalize predictive asset pricing under asynchronous disclosures as a problem of learning return distributions under state uncertainty, and identify conditional factor exposures as the key channel for propagating disclosure uncertainty through the pricing pipeline.
- We design a disclosure-aware hierarchical latent diffusion module that unifies backward imputation, real-time nowcasting, and forward forecasting of firm fundamentals, with observational anchoring that preserves known values while sampling plausible unknown states.
- We develop a scenario-conditioned multi-horizon factor pricing mechanism with strictly causal factor extraction, scenario-dependent exposures, and a shared frozen horizon decoder, achieving end-to-end uncertainty propagation while maintaining cross-horizon structural consistency and economic interpretability.

2 Related Work

Financial State Inference under Asynchronous Disclosures. Firm-characteristic panels combine continuously observed market variables with accounting variables released quarterly or annually and after reporting lags. Missingness in such panels is therefore heterogeneous: an entry may be structurally unavailable, not yet disclosed, or merely stale. Asset-pricing studies recover incomplete characteristics through low-rank, Bayesian, and attention-based structure Bryzgalova et al. (2025a); Freyberger et al. (2025); Beckmeyer & Wiedemann (2022); Chen & McCoy (2024). Related time-series methods use bidirectional recurrence, self-attention, or conditional diffusion to model missing-value distributions Cao et al. (2018); Tashiro et al. (2021); Du et al. (2023); Liu et al. (2023). STDM further exploits conditional diffusion and cross-stock dependence for missing and real-time financial inference Fang et al. (2024). Together, they improve fidelity under irregular sampling but remain detached from economic supervision supplied by future returns. However, these

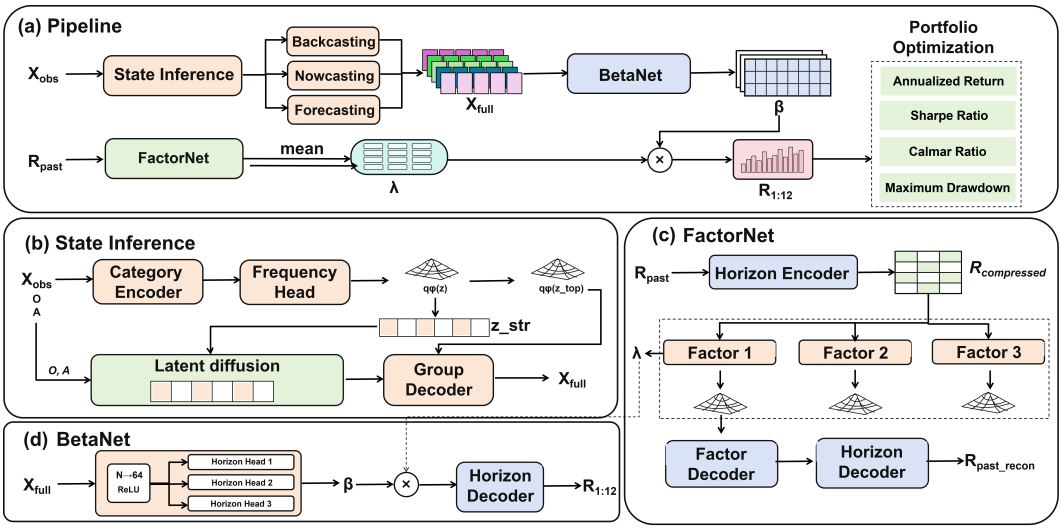

Figure 1: Architecture of the Proposed Network.

approaches largely evaluate masked-value reconstruction and encode availability through generic masks; they do not jointly distinguish disclosure status from information age while anchoring values already observed at the decision date.

Neural Conditional Asset Pricing and Multi-Horizon Prediction. Conditional asset-pricing models seek economically structured representations of returns. IPCA models characteristic-dependent loadings, whereas autoencoder and adversarial formulations learn nonlinear factors and exposures Kelly et al. (2019); Gu et al. (2021); Chen et al. (2024). FactorVAE, CQVAE, CAVB, HireVAE, RSAP, and GraphVAE introduce variational, quantile, hierarchical, relational, or graph-based latent-factor representations Duan et al. (2022); Yang et al. (2024); Liu et al. (2025); Wei et al. (2023); Xiang et al. (2024); Jia et al. (2024). FactorGCL enriches factor discovery through hypergraph and contrastive learning Duan et al. (2025). In parallel, direct return forecasters such as MASTER and StockMixer model temporal and cross-stock dependence without explicitly separating factor premia from firm-specific exposures Li et al. (2024); Fan & Shen (2024). Despite rapid progress, most methods condition on a single completed characteristic panel and either target one horizon or model horizons independently. The remaining gap is a point-in-time framework that propagates disclosure-induced state uncertainty into scenario-specific betas, combines those betas with factor premia inferred only from historical returns, and jointly predicts multiple investment horizons for portfolio decisions.

3 Methodology

3.1 Asset Pricing with Asynchronously Observed Characteristics

Consider N assets observed at monthly frequency. Let $\mathbf{X}^{\text{obs}}_{\leq t} \in \mathbb{R}^{N\times t\times D}$ denote D firm characteristics available by decision time t, where high-frequency monthly market variables coexist with quarterly/annually disclosed accounting variables carried forward between disclosures. We augment the characteristic panel with two disclosure metrics: a binary *disclosure indicator* $\mathbf{O}_{\leq t} \in \{0,1\}^{N\times t\times D}$ marking actual disclosures vs. stale values, and a *staleness tensor* $\mathbf{A}_{\leq t} \in [0,1]^{N\times t\times D}$ recording normalized time elapsed since the latest disclosure. Realized monthly returns $\mathbf{r}_{\leq t}$ are also fully observable by time t. The complete causal information set at decision time t is:

$$\mathcal{I}_t = \left\{\mathbf{r}_{\leq t}\,;\, \mathbf{X}^{\text{obs}}_{\leq t},\, \mathbf{O}_{\leq t},\, \mathbf{A}_{\leq t}\right\}. \tag{1}$$

All subsequent modeling strictly uses only variables within $\mathcal{I}_t$ to eliminate look-ahead bias. To enforce strict causality, we distinguish trailing and forward return panels for horizon $h \in \{1, \dots, H\}$,

$$R_{i,t}^{\text{past},(h)} = \frac{1}{h} \sum_{j=0}^{h-1} r_{i,t-j}, \qquad R_{i,t}^{\text{future},(h)} = \frac{1}{h} \sum_{j=1}^{h} r_{i,t+j}. \tag{2}$$

FactorNet receives only $\mathbf{R}_t^{\text{past}}$ to identify latent risk factors, while $\mathbf{R}_t^{\text{future}}$ serves exclusively as the prediction target, eliminating future information leakage from factor construction. We model average returns over holding periods $h = 1, \dots, 12$: this avoids compounding errors from recursive single-step forecasting and discontinuous factor loadings from direct monthly sequence forecasting, enabling robust out-of-sample performance via shared latent factor structure. Our primary objective is to learn the conditional distribution $p(\mathbf{R}_t^{\text{future}} \mid \mathcal{I}_t)$ rather than a single point predictor per horizon. As illustrated in Figure 1, the framework decomposes this problem into three sequentially coupled stages:

$$\mathcal{I}_t \xrightarrow{\text{state inference}} \{\widetilde{\mathbf{X}}_t^{(s)}\}_{s=1}^{S} \xrightarrow{\text{factor pricing}} \{\widehat{\mathbf{R}}_t^{\text{future},(s)}\}_{s=1}^{S} \xrightarrow{\text{allocation}} \mathbf{w}_t. \tag{3}$$

where S is the number of sampled firm-state scenarios, superscript (s) indexes the s-th scenario, $\widetilde{\mathbf{X}}_t$ are inferred complete characteristic panels, $\widehat{\mathbf{R}}_t^{\text{future}}$ are predicted multi-horizon returns, and $\mathbf{w}_t$ denotes the optimal portfolio weight vector. The first stage unifies imputation, nowcasting and forecasting in a scenario-based diffusion framework to quantify disclosure uncertainty. The second stage eliminates look-ahead bias via causally historical factor premia and enforces cross-horizon consistency via a shared horizon decoder. The final stage translates pricing uncertainty into robust investment decisions. We detail each stage below.

3.2 State Inference Stage: Unified Three-Casting Diffusive Firm-State Modeling

Exisitng approaches for asynchronous financial characteristics suffer from two core flaws: deterministic point estimates suppress inherent uncertainty, and task-siloed pipelines break temporal coherence across historical filling, current inference and future prediction. Rooted in the natural mismatch between monthly trading windows and quarterly/annual accounting disclosure cycles, our single diffusion framework with one training pipeline unifies three inference tasks spanning past, present, and future: (1) **Backward imputation**: Infers undisclosed quarterly/annual values for all prior months in the historical sequence, replacing stale carry-forwards in monthly panels to eliminate measurement bias and look-ahead leakage. (2) **Real-time nowcasting**: At the current monthly decision point, infers the contemporaneous undisclosed quarterly/annual fundamentals for the present month, resolving the real-time information gap at portfolio rebalancing. (3) **Forward forecasting**: Predicts next-period firm states conditional on observed information, providing coherent inputs for forward-looking risk management. This unified paradigm eliminates task silos and ensures temporal consistency across all horizons, enabled by two core technical designs: structured hierarchical encoding resolving dual feature heterogeneity, and observational anchoring diffusion reconciling observational and distributional consistency.

Structured Hierarchical Encoding. The first design resolves the encoding dilemma from financial features' dual heterogeneity along two orthogonal dimensions: economic function and disclosure frequency. Flat joint encoders suffer from *frequency misalignment*: high-frequency market features dominate and drown out low-frequency accounting signals. Fully separate group encoders cause *semantic isolation*, breaking cross-feature economic relatedness and wasting statistical strength. Our insight is to disentangle the two dimensions via a shared-before-specialized hierarchical architecture. Let $\mathcal{D}_g$ denote characteristics in economic category g, and $\nu(d)$ the disclosure frequency of characteristic d. A category-specific base encoder maps raw values, disclosure indicators and staleness into shared intermediate representations capturing common economic semantics:

$$\mathbf{h}_g = E_{\phi,g}\Big(\Big[\mathbf{X}_{\mathcal{D}_g}^{\text{obs}}, \mathbf{O}_{\mathcal{D}_g}, \mathbf{A}_{\mathcal{D}_g}\Big]\Big). \tag{4}$$

Frequency-specific branches then output a Gaussian latent variable for each category-frequency pair, preserving distinct updating dynamics per disclosure cycle:

$$q_\phi(\mathbf{z}_{g,\nu} \mid \mathbf{h}_g) = \mathcal{N}\big(\boldsymbol{\mu}_{g,\nu}, \operatorname{diag}(\boldsymbol{\sigma}_{g,\nu}^2)\big), \qquad \mathbf{z}_{\text{str}} = \{\mathbf{z}_{g,\nu}\}_{g,\nu}. \tag{5}$$

This design boosts statistical strength for data-sparse low-frequency features via shared semantics, while avoiding frequency misalignment across cycles. Group-wise encoding alone risks cross-group incoherence in generated outputs. We thus introduce a global latent state $q_\phi(\mathbf{z}_{\text{top}} \mid \mathbf{z}_{\text{str}})$ aggregating information across all groups and parameterizing a conditional prior $p_\theta(\mathbf{z}_{\text{str}} \mid \mathbf{z}_{\text{top}})$. The decoder injects this global firm condition into each group and reconstructs characteristics with dedicated heads:

$$\widetilde{\mathbf{z}}_{\text{str}} = \mathbf{z}_{\text{str}} + P_\theta(\mathbf{z}_{\text{top}}), \qquad (\widehat{\boldsymbol{\mu}}_g, \widehat{\log \boldsymbol{\sigma}}_g^{\,2}) = D_{\theta,g}(\widetilde{\mathbf{z}}_g). \tag{6}$$

The hierarchical variational regularizer imposes both global regularization and group-level coherence:

$$\mathcal{L}_{\text{hier}} = D_{\text{KL}}(q_\phi(\mathbf{z}_{\text{top}}) \,\|\, \mathcal{N}(\mathbf{0}, \mathbf{I})) + D_{\text{KL}}(q_\phi(\mathbf{z}_{\text{str}}) \,\|\, p_\theta(\mathbf{z}_{\text{str}} \mid \mathbf{z}_{\text{top}})) \,. \tag{7}$$

This structured latent space forms the semantic foundation for the unified three-casting diffusion.

Conditional Diffusion with Observational Anchoring. The second design resolves the fundamental trade-off between observational consistency and distributional consistency in partial-observability generation. Standard diffusion applies uniform noise across all dimensions, perturbing disclosed information and producing economically invalid samples. Deterministic imputation fixes observed values entirely, eliminates uncertainty, breaking the joint feature distribution, and deviating from the true data manifold. Our insight is an observational anchoring mechanism operating directly in the structured latent space: at every reverse diffusion step, disclosed latent dimensions remain anchored to their encoded observational values, while undisclosed dimensions evolve freely under the learned joint distribution. Critically, adjusting the anchoring mask's temporal scope enables all three inference tasks within one trained model. During training, let $\mathbf{z}_{\text{str}}^{\text{tgt}}$ be the latent encoding of fully disclosed panels, and $\mathbf{z}_{\text{str}}^{\text{obs}}$ the encoding of decision-time available information. At forward step k, we add Gaussian noise to the target latent:

$$\mathbf{z}_k = \sqrt{\bar{\alpha}_k}\, \mathbf{z}_{\text{str}}^{\text{tgt}} + \sqrt{1 - \bar{\alpha}_k}\, \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{8}$$

and train the denoiser to predict the added noise conditioned on observed partial information:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{k,\boldsymbol{\epsilon}} \left[\left\| \boldsymbol{\epsilon} - \epsilon_\theta(\mathbf{z}_k, k, \mathbf{z}_{\text{str}}^{\text{obs}}) \right\|_2^2 \right]. \tag{9}$$

At reverse sampling, we apply the anchoring rule at every denoising step:

$$\mathbf{z}_k \leftarrow \mathbf{a} \odot \mathbf{z}_{\text{str}}^{\text{obs}} + (\mathbf{1} - \mathbf{a}) \odot \mathbf{z}_k^{\text{gen}}, \tag{10}$$

where $\mathbf{a}$ is a binary disclosure mask, and $\mathbf{z}_k^{\text{gen}}$ is the denoiser output at step k. This ensures disclosed dimensions remain strictly consistent with public information, while undisclosed dimensions evolve naturally under the joint distribution. Repeated reverse trajectories yield a distribution of economically coherent firm-state scenarios:

$$\left\{ \widetilde{\mathbf{X}}_t^{(s)}, \mathbf{z}_{\text{str},t}^{(s)} \right\}_{s=1}^{S} \sim p_{\phi,\theta} \left(\widetilde{\mathbf{X}}_t, \mathbf{z}_{\text{str},t} \mid \mathcal{I}_t \right). \tag{11}$$

This anchored diffusion mechanism is the technical foundation of the unified three-casting paradigm.

3.3 Asset Pricing Stage: Scenario-Aware Multi-Horizon Factor Pricing

Our pricing framework builds on the standard cross-sectional factor pricing paradigm, where asset returns follows $r_{i,t+1} = \boldsymbol{\beta}_{i,t}^\top \mathbf{f}_{t+1} + \epsilon_{i,t+1}$, with $\boldsymbol{\beta}_{i,t}$ measuring exposure to K common latent risk factors $\mathbf{f}_{t+1} \in \mathbb{R}^K$, and $\epsilon_{i,t+1}$ denotes idiosyncratic noise. Existing deep factor models effectively extract latent risk factors from cross-sectional return snapshots, yet for multi-horizon pricing scenarios, they leave considerable room for improvement in term structure continuity. We propose a two-stage cascaded pricing framework coupled with the upstream state inference module: FactorNet identifies latent risk factors and learns a unified horizon mapping exclusively from historical returns, while BetaNet maps scenario-wise complete firm characteristics to factor loadings.

FactorNet. We propose FactorNet as a causally constrained horizon conditional variational autoencoder that learns a shared low-dimensional horizon manifold of latent risk factors using only historical return trajectories, following a three-tier compression-inference-reconstruction architecture

over single-time-step trailing return windows. Instead of treating 12 holding periods as independent dimensions, a horizon compression encoder projects the 12-dimensional horizon space into a 3-dimensional compressed embedding via two successive tanh-activated linear layers:

$$\mathbf{C}_t = \mathbf{W}_{h2} \cdot \tanh\left(\mathbf{W}_{h1} \cdot \mathbf{R}_t^{\text{past}} + \mathbf{b}_{h1}\right) + \mathbf{b}_{h2}, \tag{12}$$

where $\mathbf{R}_t^{\text{past}} \in \mathbb{R}^{N\times 12}$ is the trailing multi-horizon return cross-section at time t, and $\mathbf{C}_t \in \mathbb{R}^{N\times 3}$ is the compressed horizon representation. This design captures nonlinear dependence structures across holding periods and embeds the full term structure into a smooth low-dimensional manifold, eliminating redundant parameterization of horizon-specific factors. For each compressed horizon dimension, we perform independent conditional variational inference to extract K latent risk factors, with a shared standard normal prior to ensure factor space consistency. For horizon $c \in \{0, 1, 2\}$, the encoder maps the returns to Gaussian posterior parameters:

$$\boldsymbol{\mu}_{t,c}, \log \boldsymbol{\sigma}_{t,c}^2 = \text{FactorEnc}_c\left(\mathbf{C}_{t,:,c}\right). \tag{13}$$

We take the posterior mean as the deterministic factor representation and stack across horizons to form $\mathbf{f}_t = [\boldsymbol{\mu}_{t,0}, \boldsymbol{\mu}_{t,1}, \boldsymbol{\mu}_{t,2}] \in \mathbb{R}^{K\times 3}$. A symmetric horizon decoder maps compressed factor representations back to the full 12-dimensional horizon space, reconstructing the trailing return panel. A core design choice is that this decoder is learned once in FactorNet and reused in a permanently frozen state in the downstream BetaNet. This implements cross-module representation alignment: the mapping from latent factor space to return space is identical across factor extraction and pricing stages, eliminating cross-module misalignment error present in standard two-stage factor models. At any decision time t, the future factor realization f_{t+1} is unobservable. To uphold strict causal validity in forward-looking estimation, we compute λ_t as the rolling mean of historical latent factors using only information available by time t: $\lambda_t = \frac{1}{W}\sum_{\tau=t-W}^{t} f_\tau$ where W denotes the rolling estimation window size. Economically, λ_t serves as the empirical estimate of the conditional expected factor risk premium $\mathbb{E}_t[f_{t+1}]$, making it a causally consistent substitute for the unobserved future factor f_{t+1} in the standard factor pricing formula.

BetaNet. The beta network maps firm characteristics to factor loadings. Unlike standard models that operate directly on raw panels with stale carry-forward values, we feed scenario-wise complete characteristics from the upstream inference stage, correcting for staleness bias and naturally propagating fundamental uncertainty into return forecasts. We adopt a three-head architecture with shared feature transformation, corresponding to the three compressed horizon dimensions from FactorNet. For the s-th scenario at decision time t, all inputs are strictly limited to information available by $t-1$: lagged complete characteristics $\mathbf{Z}_{t-1}^{(s)} \in \mathbb{R}^{N\times 96}$ inferred from the upstream stage. First, a shared ReLU linear layer map 96-dimensional characteristics to a 64-dimensional shared representation, capturing common economic signals across horizons. Three independent linear heads then produce factor loadings for each compressed horizon dimension, stacked into a full loading tensor:

$$\boldsymbol{\beta}_{t-1}^{(s)} = \left[\boldsymbol{\beta}_{t-1,0}^{(s)}, \boldsymbol{\beta}_{t-1,1}^{(s)}, \boldsymbol{\beta}_{t-1,2}^{(s)}\right] \in \mathbb{R}^{N\times K\times 3}. \tag{14}$$

Expected returns in compressed horizon space are computed via element-wise product over the factor dimension, and we then use the frozen horizon decoder from FactorNet to map to the full 12-horizon future return panel:

$$\widehat{\mathbf{R}}_t^{\text{future},(s)} = \text{HorizonDec}_{\text{frozen}}\left(\sum_{k=1}^{K} \boldsymbol{\beta}_{t-1,:,k}^{(s)} \odot \boldsymbol{\lambda}_{t-1,k,:}\right). \tag{15}$$

This design establishes an end-to-end uncertainty propagation chain: upstream state inference generates plausible firm-state scenarios, each scenario produces distinct factor loadings, and the ensemble of forecasts forms the conditional distribution of future returns $p(R_t^{future} \mid \mathcal{I}_t)$, providing robust inputs for downstream portfolio optimization. We then feed multi-scenario predicted returns into practical allocation schemes such as Top-K stock selection, translating uncertainty-aware pricing signals derived from asynchronous characteristic recovery into real-world actionable investment decisions. Complete optimization details, full algorithm pseudocode, and loss function derivations are deferred to the Appendix.

4 EXPERIMENTS

4.1 SETUP

Data and Baselines. We evaluate on the Open Source Asset Pricing (OSAP) data Chen & Zimmermann (2021), comprising $N = 405$ U.S. stocks, $D = 96$ firm characteristics, and monthly returns. We use chronological splits: 1980/01–2007/12 for training, 2008/01–2009/12 for validation, and 2010/01–2022/12 for testing. We compare with five direct return forecasters: **SFM** Zhang et al. (2017), **ALSTM** Qin et al. (2017), **TRA** Lin et al. (2021), **AdaRNN** Du et al. (2021), and **GAT** Vrahatis et al. (2024), and seven structured asset-pricing models: **AGDL** Chatigny et al. (2021), **CAE** Gu et al. (2021), **FactorVAE** Duan et al. (2022), **CQVAE** Yang et al. (2024), **GAN** Chen et al. (2024), **AP-Tree** Bryzgalova et al. (2025b), and **FactorGCL** Duan et al. (2025). Implementation details are in Appendix.

Metrics. We evaluate our framework from two perspectives: cross-sectional return ranking and portfolio decision-making. **(1) Ranking Metrics:** Since asset pricing relies on cross-sectional discrimination, we adopt Rank Information Coefficient $\mathrm{RankIC}_t = \mathrm{corr}\big(\mathrm{rank}(\hat{r}_t), \mathrm{rank}(r_t)\big)$ to evaluate the consistency between predicted and realized return rankings. The average RankIC over the testing period and $\mathrm{RankICIR} = \mathrm{mean}(\mathrm{RankIC}_t)/\mathrm{std}(\mathrm{RankIC}_t)$ are reported to measure prediction quality and temporal stability. **(2) Portfolio Decision Metrics:** To evaluate the economic effectiveness of uncertainty-aware investment decisions, we construct portfolios based on predicted return scenarios and report Annualized Return (AR), Sharpe Ratio (SR), Maximum Drawdown (MDD), and Calmar Ratio (CR). Higher AR, SR, and CR indicate better risk-adjusted performance, whereas lower MDD indicates improved downside risk control.

4.2 PERFORMANCE EVALUATION

Overall asset ranking and portfolio selection. Table 1 reports one-month cross-sectional ranking accuracy and out-of-sample SOTA results across different economically core metrics, achieving the highest RankIC, RankICIR, annualized return, Calmar ratio, and the lowest maximum drawdown. For cross-sectional return discrimination, our model improves RankIC by approximately 15% relative to the strongest baseline AGDL, and lifts RankICIR by 8% over the previous best FactorGCL. This improvement directly traces to our disclosure-aware latent scenario inference: by correcting staleness bias in asynchronously disclosed firm fundamentals and propagating state uncertainty through conditional factor exposures, we generate more reliable and temporally stable ranking signals for cross-sectional stock selection. In portfolio terms, the model delivers a 3.6% higher annualized return than the top baseline GAT, while reducing maximum drawdown by 3.1% compared to FactorGCL, translating to an 8.4% improvement in Calmar ratio. This confirms that our uncertainty-aware pricing pipeline, paired with scenario-based portfolio construction, effectively translates more accurate ranking signals into stronger downside risk protection and superior risk-adjusted returns. Notably, we do not attain the highest Sharpe ratio, which narrowly belongs to ALSTM, a pure time-series forecaster optimized for volatility fitting. This is consistent with our design priority: our framework targets accurate cross-sectional selection and robust drawdown control rather than pure volatility normalization, which aligns more closely with real-world quantitative investment objectives where ranking accuracy and downside protection drive long-term portfolio value.

Multi-horizon pricing and portfolio performance. Table 2 reports out-of-sample performance across holding horizons from one to twelve months. Two key patterns emerge that align with our model design. First, cross-sectional ranking strength grows almost consistently with longer horizons. RankIC rises from 1.93% at the one-month horizon to 6.24% at twelve months, and RankICIR more than doubles from 15.0 to 40.0 over the same range. This pattern reflects the nature of fundamental-driven return predictability: short-horizon returns are dominated by idiosyncratic noise and transient liquidity shocks, while longer-horizon returns align more closely with the firm fundamentals captured by our latent state inference. Temporary dips at intermediate horizons do not alter the overall upward trend, confirming that our shared low-rank horizon decoder preserves coherent term structure rather than overfitting to individual horizons. Second, portfolio performance remains remarkably stable across horizons. Annualized returns hold steady around 14.3–16.5%, Sharpe ratios fluctuate within a narrow band around 0.80, and Calmar ratios stay near 0.47, with no systematic degradation as the holding period lengthens. Maximum drawdown also shows no monotonic worsening, remaining between 29.5% and 36.0% across all horizons. This cross-horizon

Table 1: Performance comparison on the OSAP dataset.

Model	RankIC (%) ↑	RankICIR (%) ↑	AR (%) ↑	SR ↑	CR ↑	MDD (%) ↓
SFM	0.398 ± 0.154	4.135 ± 4.130	12.493 ± 1.718	0.819 ± 0.089	0.405 ± 0.071	31.031 ± 1.860
GAT	0.144 ± 0.108	4.685 ± 1.067	$\underline{13.836 \pm 2.459}$	0.822 ± 0.108	0.447 ± 0.109	31.374 ± 2.918
ALSTM	0.571 ± 0.041	5.718 ± 0.852	13.550 ± 1.158	$\mathbf{0.836 \pm 0.085}$	0.390 ± 0.074	35.456 ± 5.094
TRA	0.324 ± 0.048	2.133 ± 0.530	12.356 ± 4.493	0.614 ± 0.217	0.293 ± 0.118	44.447 ± 8.267
AdaRNN	0.373 ± 0.087	1.164 ± 0.748	12.433 ± 1.134	0.720 ± 0.050	0.392 ± 0.079	32.741 ± 6.980
AGDL	$\underline{1.675 \pm 0.310}$	3.566 ± 1.416	9.134 ± 1.366	0.386 ± 0.093	0.165 ± 0.049	56.311 ± 4.505
AP-Tree	0.733 ± 0.161	3.320 ± 0.557	10.312 ± 1.933	0.591 ± 0.130	0.376 ± 0.072	38.850 ± 8.561
GAN	0.837 ± 0.191	4.851 ± 0.651	11.313 ± 1.024	0.609 ± 0.127	0.439 ± 0.056	42.029 ± 6.357
CAE	0.628 ± 0.121	5.078 ± 0.896	13.239 ± 3.082	0.675 ± 0.139	0.277 ± 0.096	49.041 ± 5.315
CQVAE	0.844 ± 0.129	1.442 ± 0.538	10.765 ± 2.501	0.467 ± 0.107	0.210 ± 0.076	53.462 ± 9.119
FactorVAE	0.761 ± 0.046	9.973 ± 0.624	13.022 ± 1.648	0.754 ± 0.057	0.412 ± 0.056	31.889 ± 4.288
FactorGCL	1.213 ± 0.171	$\underline{13.931 \pm 3.704}$	13.812 ± 0.630	0.809 ± 0.008	$\underline{0.454 \pm 0.021}$	$\underline{30.420 \pm 0.629}$
Ours	$\mathbf{1.927 \pm 0.473}$	$\mathbf{15.036 \pm 5.361}$	$\mathbf{14.329 \pm 1.314}$	$\underline{0.825 \pm 0.035}$	$\mathbf{0.492 \pm 0.042}$	$\mathbf{29.462 \pm 5.044}$

Note. ↑ indicates that larger values are better, whereas ↓ indicates that smaller values are better. Bold and underlined values denote the best and second-best mean results, respectively. Results are reported as mean ± standard deviation over five runs. Portfolio metrics are computed using the one-month mean–variance (MV) strategy. MDD is reported as its absolute magnitude.

Table 2: Out-of-sample Multi-horizon Pricing and Portfolio Performance.

Metric	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$	$h=11$	$h=12$
RankIC (%) ↑	1.927 ± 0.473	2.388 ± 0.697	3.170 ± 1.172	3.249 ± 0.908	3.261 ± 1.193	4.021 ± 0.743	4.510 ± 0.895	4.765 ± 0.910	5.241 ± 1.365	5.660 ± 0.846	5.978 ± 1.410	6.237 ± 0.896
RankICIR (%) ↑	15.036 ± 5.361	18.686 ± 6.361	24.614 ± 9.188	22.454 ± 6.004	20.973 ± 5.688	27.390 ± 7.460	29.126 ± 8.355	31.881 ± 10.140	36.822 ± 13.308	37.095 ± 12.106	39.852 ± 13.717	39.972 ± 12.058
AR (%) ↑	14.329 ± 1.314	14.969 ± 0.854	15.261 ± 0.769	16.475 ± 1.777	15.105 ± 1.096	15.327 ± 0.635	15.231 ± 0.556	15.010 ± 0.691	14.545 ± 1.557	14.978 ± 0.622	14.972 ± 0.642	14.947 ± 0.474
SR ↑	0.825 ± 0.035	0.820 ± 0.028	0.829 ± 0.037	0.839 ± 0.074	0.796 ± 0.029	0.766 ± 0.089	0.806 ± 0.026	0.759 ± 0.098	0.743 ± 0.135	0.814 ± 0.027	0.813 ± 0.032	0.815 ± 0.025
CR ↑	0.492 ± 0.042	0.481 ± 0.033	0.487 ± 0.030	0.506 ± 0.056	0.467 ± 0.029	0.451 ± 0.050	0.470 ± 0.029	0.431 ± 0.083	0.428 ± 0.090	0.470 ± 0.027	0.469 ± 0.028	0.470 ± 0.027
MDD (%) ↓	29.462 ± 5.044	31.357 ± 3.801	31.465 ± 3.046	32.688 ± 3.246	32.470 ± 3.276	34.349 ± 4.429	32.561 ± 3.087	35.997 ± 7.634	34.932 ± 5.681	32.037 ± 3.022	32.098 ± 3.170	31.922 ± 2.773

Note. All entries are out-of-sample results reported as mean ± standard deviation over five seeds.

stability distinguishes our framework from horizon-specific models, which often suffer from inconsistent performance across holding periods. The shared horizon representation ensures that pricing signals are consistent from short to long horizons, supporting practical multi-period investment decision-making.

4.3 Ablation Study and Sensitivity Analysis

Upstream Asynchronous State Inference Ablations. We first evaluate the contributions of hierarchical latent encoding and staleness-aware scenario sampling within the upstream inference module. All variants in this group correspond to the *Disclosure-aware State Inference* panel in Table 3.

(1) Raw Lagged Features: As our most aggressive ablation (Table 3, row "Raw Lag Feat."), feeding raw stale lagged characteristics directly into the downstream pipeline causes substantial performance degradation. Without probabilistic state correction, RankIC falls to 0.712% and RankICIR drops to 8.934. Portfolio metrics also deteriorate sharply: annualized return declines to 11.156%, Sharpe ratio drops to 0.593, and maximum drawdown rises to 37.841. This supports our core motivation: naively carrying forward asynchronous stale characteristics injects heavy noise and erodes cross-sectional predictive signals. (2) Deterministic Mean Imputation: Replacing stochastic multi-scenario sampling with deterministic mean imputation (Table 3, row "Det. Impute") increases both RMSE and MAE, indicating worse feature reconstruction quality. Although deterministic imputation partially mitigates stale-data artifacts, discarding distributional uncertainty hurts ranking stability: RankIC drops to 1.225% and RankICIR falls to 10.817. Portfolio performance is correspondingly weakened, demonstrating the value of probabilistic scenario representations. (3) Without Hierarchical Encoder: Removing the hierarchical encoder (Table 3, row "w/o Hier. Enc.") slightly degrades reconstruction RMSE and MAE. Downstream pricing and portfolio metrics consistently worsen: RankIC decreases to 1.503%, RankICIR to 13.162, and annualized return shrinks to 12.987%. (4) Without Staleness Channel: Dropping the dedicated staleness-awareness channel (Table 3, row "w/o Stale Chan.") yields only minor changes in reconstruction error, yet still impairs ranking and portfolio outcomes: RankICIR falls to 13.785 and Calmar ratio declines to 0.469.

Table 3: Ablation Study and Sensitivity Analysis.

Variant	RMSE ↓	MAE ↓	RankIC (%) ↑	RankICIR (%) ↑	AR (%) ↑	Sharpe ↑	Calmar ↑	MDD (%) ↓
Full Model	0.273	0.209	1.799	15.529	14.080	0.842	0.527	26.742
Disclosure-aware State Inference								
Raw Lag Feat.	–	–	0.712	8.934	11.156	0.593	0.312	37.841
Det. Impute	0.297	0.233	1.225	10.817	12.203	0.701	0.392	33.675
w/o Hier. Enc.	0.289	0.225	1.503	13.162	12.987	0.758	0.448	30.246
w/o Stale Chan.	0.284	0.219	1.579	13.785	13.342	0.781	0.469	29.583
Return Prediction								
Recur. Roll	–	–	1.446	12.391	12.618	0.735	0.415	32.279
Contemp. f	–	–	1.402	12.047	12.374	0.729	0.403	33.015

Table 4: Upstream State-inference Errors across Three Tasks. Results are reported as mean ± standard deviation over five seeds.

	Real-time		Quarterly		Annual		Overall	
Casting	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓
Nowcast	–	–	0.2920 ± 0.0437	0.2378 ± 0.0338	0.2938 ± 0.0068	0.2280 ± 0.0069	0.2938 ± 0.0094	0.2287 ± 0.0086
Backcast	–	–	0.3075 ± 0.0327	0.2597 ± 0.0224	0.3057 ± 0.0156	0.2574 ± 0.0102	0.3059 ± 0.0166	0.2575 ± 0.0108
Forecast	0.2081 ± 0.0023	0.1490 ± 0.0026	0.2954 ± 0.0316	0.2443 ± 0.0234	0.2862 ± 0.0117	0.2297 ± 0.0097	0.2129 ± 0.0033	0.1533 ± 0.0031

These two variants highlight an important practical insight: *feature-level reconstruction metrics (RMSE, MAE) are poor proxies for investment-relevant downstream performance*. Optimizing purely for imputation quality can be misleading for quantitative investment tasks.

Midstream Factor Pricing Ablations. We next validate our causal multi-horizon factor-pricing design. Variants below correspond to the *Return Prediction* panel in Table 3. (1) Recursive Monthly Rollout: Substituting our unified low-rank term-structure decoder with recursive monthly rollout (Table 3, row "Recur. Roll") degrades cross-sectional discrimination. RankIC drops to 1.446% and RankICIR falls to 12.391. Recursive forecasting accumulates sorting errors across horizons, damaging portfolio-selection signals. (2) Contemporaneous Latent Factors: Replacing forward-looking causal rolling risk premia with in-sample contemporaneous latent factors (Table 3, row "Contemp. f") also hurts ranking performance: RankIC falls to 1.402% and RankICIR drops to 12.047. Even though contemporaneous factors can fit in-sample return magnitudes, they fail to deliver stable cross-sectional ranking signals required for profitable out-of-sample portfolio construction.

4.4 Upstream state-inference quality

While our upstream state-inference module is ultimately designed to feed probabilistic firm-state scenarios into downstream asset-pricing and portfolio optimization, we report reconstruction RMSE and MAE in Table 4 as auxiliary diagnostics across its three tasks: nowcasting, backcasting, and forecasting. Consistent with intuition, backcasting masked historical quarterly/annual fundamentals yields larger reconstruction errors than nowcasting and forecasting. Real-time market characteristics attain the lowest reconstruction error under forecasting.

5 Conclusion

We address the cascading bias caused by asynchronous firm-level disclosures across sequential financial analytics, from state inference and asset pricing to portfolio decision-making. Our unified multi-horizon uncertainty-aware latent scenario learning framework resolves this problem by modeling firm fundamentals as probabilistic scenarios and propagating state uncertainty through conditional factor exposures, while enforcing strict causal validity throughout the pipeline. Empirical results on the OSAP dataset show that our approach consistently outperforms state-of-the-art baselines across return prediction, cross-sectional ranking, and out-of-sample portfolio performance, and ablation studies validate the material contribution of each core component. For future work, we aim to extend the framework to endogenous disclosure timing and regime-adaptive scenario weighting to further improve real-world applicability.

REFERENCES

Heiner Beckmeyer and Timo Wiedemann. Recovering missing firm characteristics with attention-based machine learning. 2022.

Svetlana Bryzgalova, Sven Lerner, Martin Lettau, and Markus Pelger. Missing financial data. *The Review of Financial Studies*, 38(3):803–882, 2025a.

Svetlana Bryzgalova, Markus Pelger, and Jason Zhu. Forest through the trees: Building cross-sections of stock returns. *The Journal of Finance*, 80(5):2447–2506, 2025b.

Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems*, 31, 2018.

Philippe Chatigny, Ruslan Goyenko, and Chengyu Zhang. Asset pricing with attention guided deep learning. *Available at SSRN 3971876*, 2021.

Andrew Y Chen and Jack McCoy. Missing values handling for machine learning portfolios. *Journal of Financial Economics*, 155:103815, 2024.

Andrew Y Chen and Tom Zimmermann. Open source cross-sectional asset pricing. *Critical Finance Review, Forthcoming*, 2021.

Luyang Chen, Markus Pelger, and Jason Zhu. Deep learning in asset pricing. *Management Science*, 70(2):714–750, 2024.

Wenjie Du, David Côté, and Yan Liu. Saits: Self-attention-based imputation for time series. *Expert Systems with Applications*, 219:119619, 2023.

Yuntao Du, Jindong Wang, Wenjie Feng, Sinno Pan, Tao Qin, Renjun Xu, and Chongjun Wang. AdaRNN: Adaptive learning and forecasting of time series. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 402–411, 2021.

Yitong Duan, Lei Wang, Qizhong Zhang, and Jian Li. FactorVAE: A probabilistic dynamic factor model based on variational autoencoder for predicting cross-sectional stock returns. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, volume 36, pp. 4468–4476, 2022.

Yitong Duan, Weiran Wang, and Jian Li. FactorGCL: A hypergraph-based factor model with temporal residual contrastive learning for stock returns prediction. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, volume 39, pp. 173–181, 2025.

Jinyong Fan and Yanyan Shen. Stockmixer: a simple yet strong mlp-based architecture for stock price forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 8389–8397, 2024.

Yupeng Fang, Ruirui Liu, Huichou Huang, Peilin Zhao, and Qingyao Wu. A spatio-temporal diffusion model for missing and real-time financial data inference. In *Proceedings of the 33rd ACM international conference on information and knowledge management*, pp. 602–611, 2024.

Joachim Freyberger, Bjoern Hoeppner, Andreas Neuhierl, and Michael Weber. Missing data in asset pricing panels. *The Review of Financial Studies*, 38(3):760–802, 2025.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Autoencoder asset pricing models. *Journal of Econometrics*, 222(1):429–450, 2021.

Yulong Jia, Guanxing Li, Ganlong Zhao, Xiangru Lin, and Guanbin Li. Graphvae: Unveiling dynamic stock relationships with variational autoencoder-based factor modeling. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 3807–3811, 2024.

Bryan T Kelly, Seth Pruitt, and Yinan Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524, 2019.

Tong Li, Zhaoyang Liu, Yanyan Shen, Xue Wang, Haokun Chen, and Sen Huang. Master: Market-guided stock transformer for stock price forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 162–170, 2024.

Hengxu Lin, Dong Zhou, Weiqing Liu, and Jiang Bian. Learning multiple stock trading patterns with temporal routing adaptor and optimal transport. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1017–1026, 2021.

Mingzhe Liu, Han Huang, Hao Feng, Leilei Sun, Bowen Du, and Yanjie Fu. Pristi: A conditional diffusion framework for spatiotemporal imputation. In *2023 IEEE 39th international conference on data engineering (ICDE)*, pp. 1927–1939. IEEE, 2023.

Ruirui Liu, Huichou Huang, and Johannes Ruf. Asset pricing with contrastive adversarial variational Bayes. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence*, pp. 7572–7579, 2025.

Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*, 2017.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems*, 34:24804–24816, 2021.

Aristidis G Vrahatis, Konstantinos Lazaros, and Sotiris Kotsiantis. Graph attention networks: a comprehensive review of methods and applications. *Future Internet*, 16(9):318, 2024.

Zikai Wei, Anyi Rao, Bo Dai, and Dahua Lin. Hirevae: an online and adaptive factor model based on hierarchical and regime-switch vae. *arXiv preprint arXiv:2306.02848*, 2023.

Quanzhou Xiang, Zhan Chen, Qi Sun, and Rujun Jiang. Rsap-dfm: Regime-shifting adaptive posterior dynamic factor model for stock returns prediction. In *IJCAI*, pp. 6116–6124, 2024.

Xuanling Yang, Zhoufan Zhu, Dong Li, and Ke Zhu. Asset pricing via the conditional quantile variational autoencoder. *Journal of Business and Economic Statistics*, 42(2):681–694, 2024.

Liheng Zhang, Charu Aggarwal, and Guo-Jun Qi. Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2141–2149, 2017.

A Point-in-Time Data Construction

Characteristic panel. The monthly panel contains 96 cross-sectionally rank-normalized characteristics. By economic content, 51 variables are fundamental, 34 are price related, and 11 describe trading activity. By update frequency, 64 variables are available monthly, three are disclosed quarterly, and 29 are disclosed annually. For every asset, month, and characteristic, we retain both a disclosure indicator and the normalized age since its latest disclosure. Thus, an observed release, a stale carry-forward value, and an unavailable value remain distinct inputs.

Trailing inputs and forward labels. FactorNet receives only $\mathbf{R}_t^{\text{past}}$. For a prediction indexed by t, BetaNet uses the inferred state $\mathbf{Z}_{t-1}$ and the historical premium

$$\boldsymbol{\lambda}_t = \frac{1}{48} \sum_{\tau=t-48}^{t} \mathbf{f}_\tau, \qquad \mathbf{f}_\tau = F_\phi(\mathbf{R}_\tau^{\text{past}}), \tag{16}$$

where every factor is computed from returns observed no later than t. The forward panel is used only as the supervised target. This separation is preserved under the chronological train–validation–test protocol and during test-time scenario generation.

Algorithm 1 Sequential training of the proposed framework

Require: Point-in-time panel $(\mathbf{X}^{\mathrm{obs}}, \mathbf{O}, \mathbf{A})$, monthly returns $\mathbf{r}$, horizons H, premium window W
1: Construct $\mathbf{R}^{\mathrm{past}}$ and $\mathbf{R}^{\mathrm{future}}$
2: Train the state model on the training split; select its checkpoint on validation data
3: Train FactorNet F_ϕ to reconstruct $\mathbf{R}^{\mathrm{past}}$; freeze F_ϕ and its horizon decoder D_H
4: Compute $\mathbf{f}_\tau = F_\phi(\mathbf{R}^{\mathrm{past}}_\tau)$ for every admissible training and validation origin
5: **for** each BetaNet training origin $t \geq W$ **do**
6: Form $\boldsymbol{\lambda}_t = W^{-1} \sum_{\tau=t-W}^{t} \mathbf{f}_\tau$
7: Infer a complete point-in-time state $\widetilde{\mathbf{X}}_{t-1}$ and compute $\boldsymbol{\beta}_t = B_\psi(\widetilde{\mathbf{X}}_{t-1})$
8: Decode $\widehat{\mathbf{R}}^{\mathrm{future}}_t = D_H(\sum_k \boldsymbol{\beta}_{t,:,k} \odot \boldsymbol{\lambda}_{t,k})$
9: Update ψ using $\mathcal{L}_{\mathrm{beta}}$; keep the state model, F_ϕ, and D_H fixed
10: **end for**
11: Select the BetaNet checkpoint on the validation split and freeze all parameters
Ensure: Frozen state model, FactorNet, horizon decoder, and BetaNet

B Training and Implementation Details

Table 5 records the settings used for the reported experiments. We distinguish the state-model latent dimension K_z from the pricing factor dimension K_f to avoid overloading the factor notation in the main text.

Table 5: Implementation settings.

Setting	Value
Assets / characteristics	405 / 96
Return horizons / compressed horizons	12 / 3
State latent dimension K_z	8
Pricing factor dimension K_f	16
Factor-premium window	48 months
Diffusion / DDIM sampling steps	100 / 20
State scenarios per origin	100
Optimizer / learning rate	Adam / 10^{-3}
Training epochs	60 per stage
Factor KL / BetaNet ℓ_1 weight	10^{-3} / 10^{-4}
Portfolio rule / transaction cost	top 10% long-only / 10 bps
Independent runs	5

The state model, FactorNet, and BetaNet are trained sequentially. Their objectives are

$$\mathcal{L}_{\mathrm{state}} = \mathcal{L}_{\mathrm{rec}} + \eta_{\mathrm{KL}} \mathcal{L}_{\mathrm{KL}} + \eta_{\mathrm{diff}} \mathcal{L}_{\mathrm{diff}}, \tag{17}$$

$$\mathcal{L}_{\mathrm{factor}} = \|\mathbf{R}^{\mathrm{past}} - \widehat{\mathbf{R}}^{\mathrm{past}}\|_2^2 + \alpha_{\mathrm{KL}} D_{\mathrm{KL}}(q_\phi \| p), \tag{18}$$

$$\mathcal{L}_{\mathrm{beta}} = \|\mathbf{R}^{\mathrm{future}} - \widehat{\mathbf{R}}^{\mathrm{future}}\|_2^2 + \alpha_\beta \|\psi\|_1. \tag{19}$$

After FactorNet is fitted on the training period, both its parameters and its horizon decoder are frozen. BetaNet is then optimized using the rolling historical premia in Equation 16; validation data select checkpoints and hyperparameters. At test time, all network parameters remain frozen and only the information set available at each origin is updated.

C Portfolio Evaluation Protocol

For the main comparison, we average the one-month return forecasts over scenarios, rank stocks cross-sectionally, and equally weight the top decile subject to long-only and fully invested constraints. Portfolio returns are evaluated one month ahead and charged 10 bps per unit turnover. AR, SR, CR, and MDD are computed from the resulting net monthly return series; the identical rule is applied to every model in Table 1.

Algorithm 2 Causal scenario prediction and portfolio construction

Require: Frozen models, point-in-time information $\mathcal{I}_{t-1}$, window W, scenarios S, previous weights $\mathbf{w}_t$

1: Compute $\{\mathbf{f}_\tau\}_{\tau=t-W}^{t}$ from trailing returns and form $\boldsymbol{\lambda}_t$
2: **for** $s = 1, \ldots, S$ **do**
3: Sample an anchored state $\widetilde{\mathbf{X}}_{t-1}^{(s)} \sim p_\theta(\cdot \mid \mathcal{I}_{t-1})$
4: Compute $\boldsymbol{\beta}_t^{(s)} = B_\psi(\widetilde{\mathbf{X}}_{t-1}^{(s)})$
5: Decode $\widehat{\mathbf{R}}_t^{\mathrm{future},(s)} = D_H\Big(\sum_k \boldsymbol{\beta}_{t,:,k}^{(s)} \odot \boldsymbol{\lambda}_{t,k}\Big)$
6: **end for**
7: Average the one-month forecasts and equally weight the top decile for the main comparison
8: For robust allocation, solve Equation 20 using the same scenario panel
9: Realize the next-month return, deduct turnover cost, and update $\mathbf{w}_{t-1}$

Ensure: Return scenarios $\{\widehat{\mathbf{R}}_t^{\mathrm{future},(s)}\}_{s=1}^{S}$ and portfolio weights $\mathbf{w}_t$

The scenario forecasts also support a common robust-allocation template,

$$\max_{\mathbf{w}\in\mathcal{W}} \min_{\mathbf{q}\in\mathcal{Q}} \sum_{s=1}^{S} q_s\, \mathbf{w}^\top \widehat{\mathbf{r}}_t^{(s)} - \frac{\gamma}{2}\mathbf{w}^\top \widehat{\boldsymbol{\Sigma}}_t \mathbf{w} - c_{\mathrm{tc}}\|\mathbf{w} - \mathbf{w}_{t-1}\|_1, \qquad \mathcal{W} = \{\mathbf{w} \geq 0 : \mathbf{1}^\top \mathbf{w} = 1\}. \tag{20}$$

Different ambiguity sets $\mathcal{Q}$ recover scenario max–min, nominal mean–CVaR, box, ellipsoidal, and KL-robust allocations; mean–variance and top-k serve as deterministic references. All solvers receive the same scenario panel and use the same rebalancing dates, constraints, and cost convention, so portfolio differences isolate the decision rule rather than the predictive input.