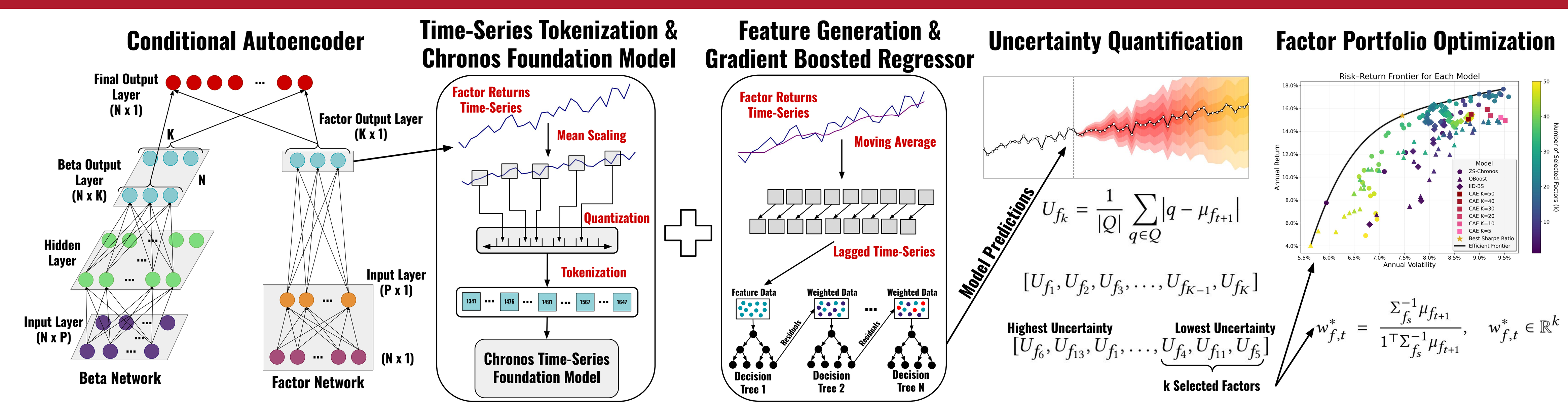
Scaling Conditional Autoencoders for Portfolio Optimization via Uncertainty-Aware Factor Selection

Ryan Engel[1], Yu Chen[1], Pawel Polak[1], Ioana Boier[2]

[1]Stony Brook University, [2]NVIDIA Corporation

Stony Brook University

Introduction



Architecture for Latent Factor Forecasting and Selection

Traditional factor models in asset pricing rely on low-dimensional representations due to concerns that higher dimensionality degrades out-of-sample performance. We address this fundamental trade-off by proposing a scalable framework that combines high-dimensional Conditional Autoencoders (K up to 50 latent factors) with a novel uncertainty-aware factor selection mechanism. For three forecasting models—including a zero-shot time-series foundation model (ZS-Chronos), quantile regression gradient-boosted trees (Q-Boost), and an I.I.D bootstrap resampling method (IID-BS)—we rank the latent factors by their forecast uncertainty and include only the top-κ most predictable factors for portfolio construction, where κ denotes the selected subset of latent factors. We analyze both individual model performance and explore a combined ensemble over a 25-year backtest.

Methodology

We Propose a Two-Stage Framework:

1. Latent Factor Construction via CAE
2. Select Factors based on Uncertainty

Stage 1: Conditional Autoencoder (CAE)

Model returns for stock i and time s as:

$$r_{i,s} = \beta_i(z_{i,s-1})^\top f_s + u_{i,s}$$

where $\beta_i(\cdot)$ is a neural network mapping characteristics $z_{i,s-1}$ to factor loadings, and latent factors are generated via:

$$f_s = W_f\left((Z_{s-1}^\top Z_{s-1})^{-1} Z_{s-1}^\top r_s\right)$$

We train the CAE model parameters by minimizing cross-sectional pricing loss:

$$\min_{\Theta} \sum_{s=1}^{t} \sum_{i=1}^{N} \left(r_{i,s} - \beta_i(z_{i,s-1})^\top f_s\right)^2$$

Stage 2: Uncertainty-Based Factor Selection

Generate forecasts $\hat{f}_{t+1}^{(\kappa)}$ and quantiles $\hat{f}_{t+1}^{(\kappa,\alpha)}$ for each model:

- **IID-BS:** I.I.D Bootstrap Resampling
- **Q-Boost:** Quantile Boosted Regression Trees
- **ZS-Chronos:** Zero-Shot Chronos Model

Compute uncertainty as mean absolute deviation:

$$U_{t+1}^{(\kappa)} = \frac{1}{|\mathcal{Q}|} \sum_{\alpha \in \mathcal{Q}} \left|\hat{f}_{t+1}^{(\kappa,\alpha)} - \hat{f}_{t+1}^{(\kappa)}\right|$$

Select κ most predictable factors with lowest $U^{(\kappa)}$ and construct factor tangency portfolio:

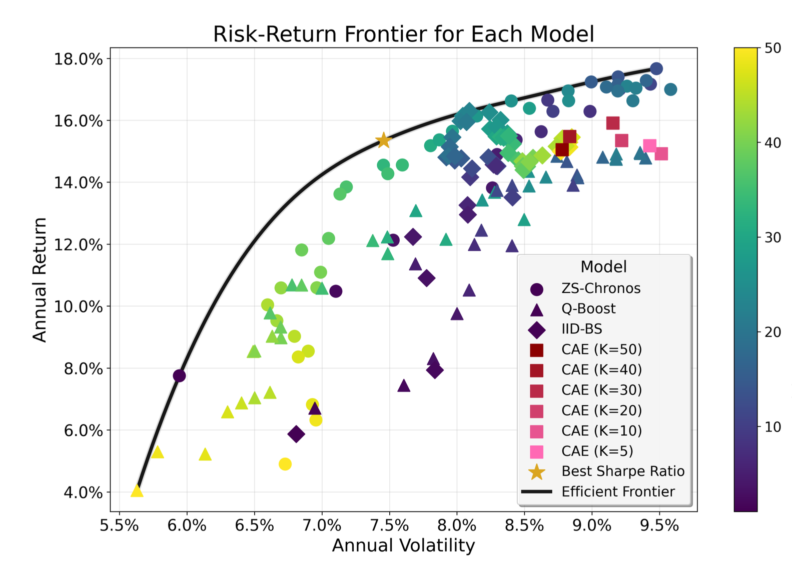
$$w_{f,t} = \frac{\Sigma_f^{-1}\mu_{t+1}}{\mathbf{1}^\top \Sigma_f^{-1}\mu_{t+1}}, \quad w_{r,t} = Z_t(Z_t^\top Z_t)^{-1} W_f^{(\kappa)\top} w_{f,t}$$

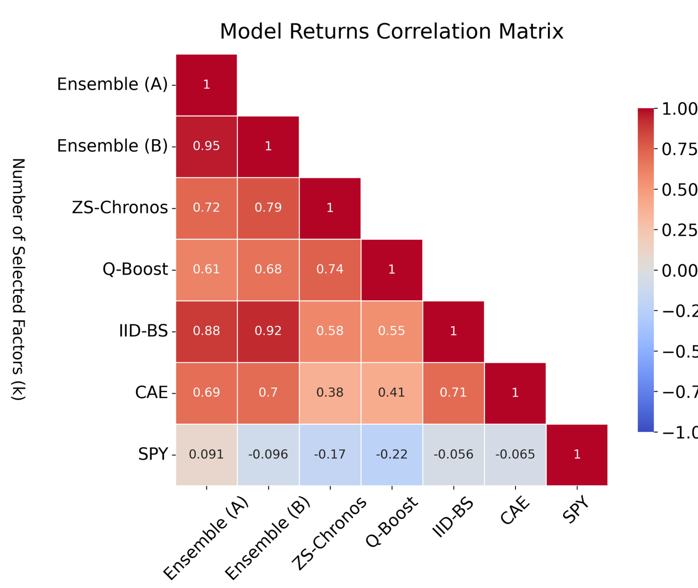
We denote the optimal number of factors as κ^*, and implement an adaptive κ^* selection algorithm. This allows us to select the best κ^* on a rolling basis rather than selecting the optimal value post-hoc. The method employs a smooth log-sum-exp approximation to optimize the Sortino ratio while incorporating temporal regularization that encourages gradual evolution of factor complexity over time.

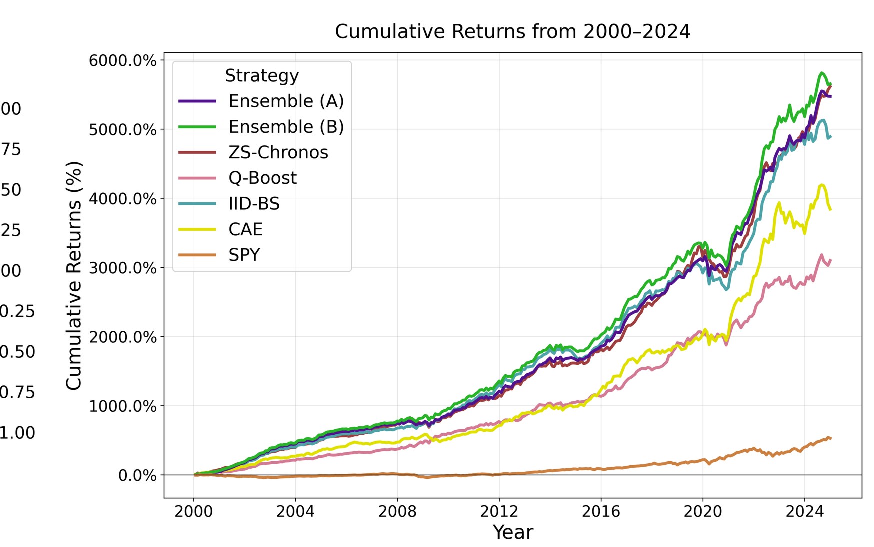
Results

	Ensemble (A)	Ensemble (B)	ZS-Chronos	Q-Boost	IID-BS	CAE ($K = 50$)	SPY
Total Return (%)	5473.927	**5655.296**	5615.464	3098.857	4890.608	3842.661	528.379
CAGR (%)	17.448	**17.599**	17.566	14.868	16.930	15.833	7.629
Annual Return (%)	16.475	16.632	**16.675**	14.315	16.085	15.226	8.551
Annual Volatility (%)	**7.633**	7.980	8.886	8.719	8.343	9.440	15.265
Sharpe Ratio	**2.158**	2.084	1.876	1.642	1.928	1.613	0.560
Sortino Ratio	**4.518**	4.464	3.052	2.786	3.577	3.064	0.791
Omega Ratio	**6.431**	5.552	5.169	3.806	4.640	3.437	1.513
Max Drawdown (%)	**8.442**	9.688	15.067	8.847	12.057	15.789	50.785
Market (SPY) Beta	0.046	-0.050	-0.100	-0.126	**-0.031**	-0.040	1
Alpha (%)	16.084	17.060	**17.532**	15.390	16.348	15.571	0

The table above shows out-of-sample performance metrics from January 2000 to December 2024 for portfolios constructed using adaptive κ^* selection. We report results for each forecasting model individually—ZS-Chronos, Q-Boost, IID-BS—as well as the base CAE model with $K = 50$, and two ensemble variants: Ensemble (A), a performance-weighted ensemble including SPY, and Ensemble (B), a performance-weighted ensemble excluding SPY.







Our methodology reveals that fewer factors often yield superior out-of-sample performance. The risk-return frontier plot demonstrates that configurations with $\kappa < K$ consistently outperform the full 50-factor CAE benchmark. Additionally, the low correlation between model configurations motivates the construction of ensemble models that leverage their complementary performance.

Conclusion

- **Summary:** This paper presents a scalable framework for high-dimensional Conditional Autoencoder (CAE) models in asset pricing that uses uncertainty-aware factor selection to strengthen performance as latent dimensionality increases.
- **Uncertainty-Driven Factor Selection:** We show that portfolios based on only a subset of available latent factors ($\kappa < K$) consistently outperform those using all factors from the CAE model.
- **Ensemble Strategy Success:** Three forecasting methods (IID-BS, Q-Boost, and ZS-Chronos) generate largely uncorrelated signals, enabling ensemble strategies that outperform individual models.