End-to-End Variational Inference for Robust Portfolio Construction

Xinyu Rain Wei, Giorgio Costa, Garud N. Iyengar

1. Introduction and Motivation

Portfolio Construction: Traditionally relies on two-stage '**predict-then-optimize**' models.

- Prediction stage: Parametrize **distribution of asset returns** y_t **conditioned on relevant features** x_t at time t. In practice, predictions are often unreliable.
- Limitation: Prediction models, e.g. linear regression, feed-forward neural networks, **produce point-predictions not conditional distribution**.

End-to-End (E2E) Learning: E2E systems integrate **prediction and decision** into a single unified model.

- Advantage: **Information can be back-propagated** through the prediction and decision layers during training.
- Calibrates prediction model with **respect to downstream decision (the 'task')**.
- Limitation: Basic E2E system still makes **conditional point prediction**.

2. Contributions

Variational Inference: Learn the conditional distribution using a Bayesian approach. We use a Variational Inference Neural Network (VINN):

- **Input**: features x_t **Output**: $\hat{y}_t^i \sim P(Y \mid X = x_t)$
- Apply an **information bottleneck** to identify meaningful, regularized latent features z_t.
- VINN is a **generative model** for a heteroscedastic conditional distribution.

Embedding generative models in E2E learning: Our E2E system consists of (i) a **VINN as the prediction layer**, and (ii) a Monte Carlo-based **constrained quadratic program** as the decision layer.

- System trained E2E using the task-based loss function composed of: (i) prediction loss (MSE), (ii) regularization (KL divergence), and (iii) task loss (realized portfolio return).
- VINN learns a "task-based" conditional distribution. Potentially different from solely prediction-based learning.

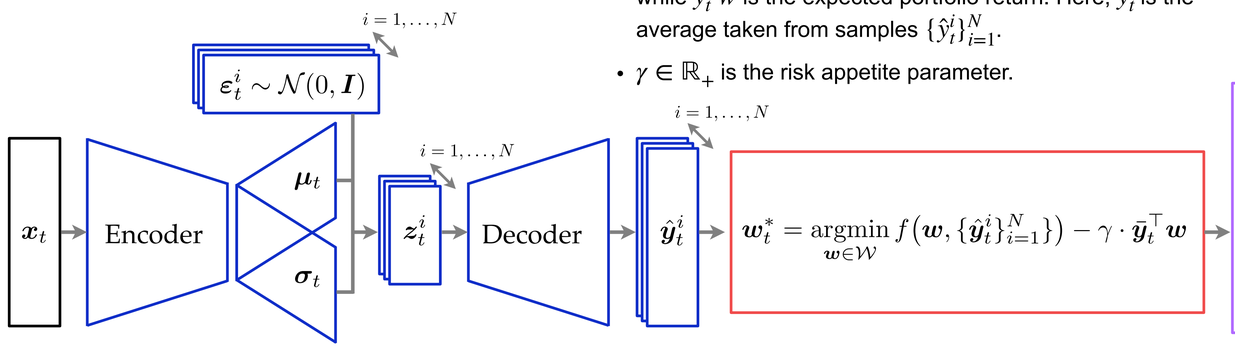
References

- Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize". *Management Science*, 68(1), 9-26.
- Donti, P., Amos, B., & Kolter, J. Z. (2017). Task-based end-to-end model learning in stochastic optimization. *Advances in neural information processing systems*, 30.
- Elmachtoub, A. N., Lam, H., Zhang, H., & Zhao, Y. (2023). Estimate-Then-Optimize Versus Integrated-Estimation-Optimization: A Stochastic Dominance Perspective. *arXiv preprint arXiv:2304.06833*.
- Costa, G., & Iyengar, G. N. (2023). Distributionally robust end-to-end portfolio construction. *Quantitative Finance*, 23(10), 1465-1482.

3. Methodology

Prediction Layer: Variational Inference NN

- Features $x_t \in \mathbb{R}^n$ are passed through the encoder.
- Encoder parametrizes the conditional distribution $P(Z \mid X = x_t) = \mathcal{N}(\mu_t, \sigma_t)$.
- Decoder: Transforms samples of $\{z_t^i \in \mathbb{R}^m\}_{i=1}^N$ into samples of $\{\hat{y}_t^i \in \mathbb{R}^p\}_{i=1}^N$.



Decision Layer: Monte Carlo-based QP

- Samples $\{\hat{y}_t^i\}_{i=1}^N$ = empirical approximation of conditional distributional.
- The optimal portfolio weights $w_t^* \in W$ determine our investment allocation at time t. The set $W \subseteq \mathbb{R}^q$ contains all pertinent portfolio constraints.
- The convex function $f : W \times \mathbb{R}^{N \times p} \to \mathbb{R}_+$ is a risk measure, while $\bar{y}_t^\top w$ is the expected portfolio return. Here, $\bar{y}_t$ is the average taken from samples $\{\hat{y}_t^i\}_{i=1}^N$.
- $\gamma \in \mathbb{R}_+$ is the risk appetite parameter.

Task Loss

- MSE loss compares the average prediction $\bar{y}_t$ against realizations y_t to ensure some predictive power is retained.
- The KL divergence regularizes the latent space in the VINN.
- Task loss evaluates the out-of-sample performance of the decision w_t^* with respect to the realizations y_t.

4. Numerical Experiment

Using synthetic data to better understand model behavior. We generate a time series of *true* latent features $\{\zeta_t \in \mathbb{R}^m\}_{t=0}^T$ $(m = 4)$, which are then linearly combined to create observed features $\{x_t \in \mathbb{R}^n\}_{t=0}^T$ $(n = 6)$. The features are then linearly combined with some white noise to generate realizations $\{y_t \in \mathbb{R}^p\}_{t=0}^T$ $(p = 2)$. Data composed of 1,600 observations; 960 are used for training and 640 are reserved for testing.

Competing models: Both models were initialized and pre-trained identically for 125 epochs. Each model was trained for a further 50 epochs using their respective loss functions.

- **Predict-then-optimize model**: VINN is trained with respect to MSE, KL divergence ~~**and task loss**~~.
- **E2E model**: VINN is trained with respect to MSE, KL divergence **and task loss**.

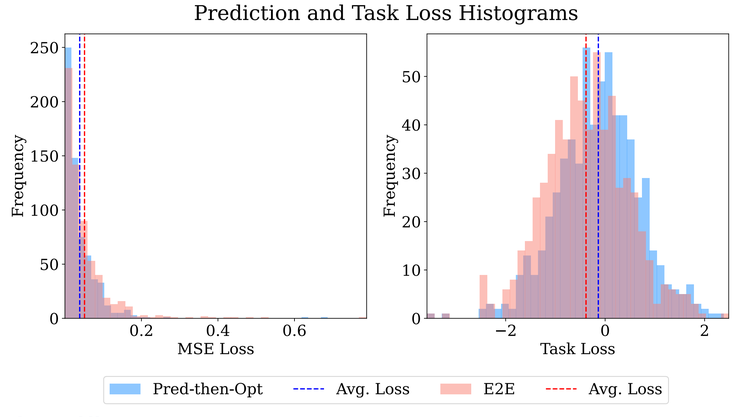


Figure 1: Loss Histograms

- MSE and Task loss over the 640 test samples.
- A lower loss indicates better performance.
- Difference in MSE significant (p-value = 0.03).
- Difference in Task loss significant (p-value < 0.001).

5. Future Work

- **Learning** γ: The risk appetite parameter γ can be learned during training in addition to the VINN weights. We expect this will significantly impact the VINN distribution since γ dictates the relative importance of the estimated conditional mean with respect to the variance.
- **Distributional Robustness**: The Monte Carlo-based approach of this model lends itself very well to formulate a distributionally robust optimization problem.

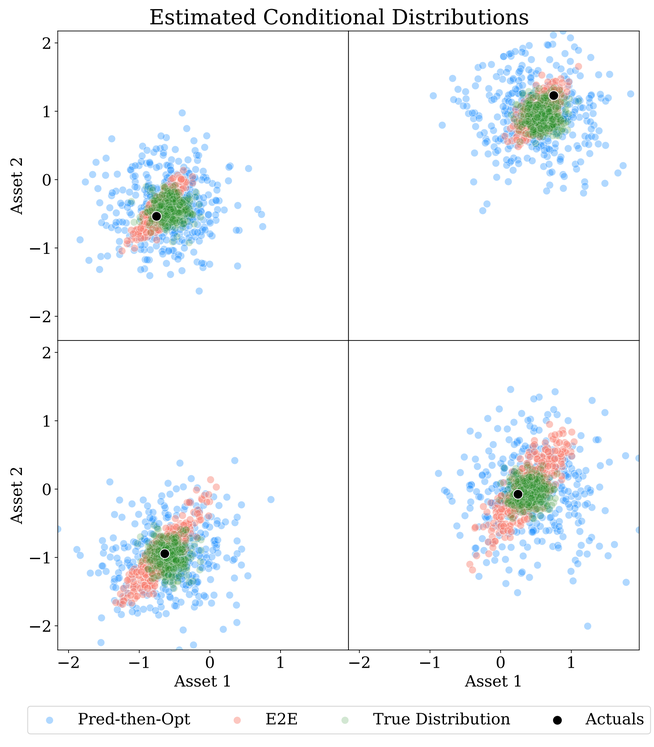


Figure 2: Conditional Distributions

- Conditional distributions learned by the two VINNs.
- Each sub-plot corresponds to a different x_t.
- E2E with task loss changes the distribution:
  - conditional variance diminishes
  - correlation between the assets increases.