# Factor Learning Portfolio Optimization Informed by Continuous-Time Finance Models

Sinong Geng[1], Houssam Nassif*[2], Zhaobin Kuang[3], Anders Max Reppen[4], Ronnie Sircar[1]

Princeton University[1], Meta[2], Google[3], Boston University[4] *Work done at Amazon

## Abstract

We study financial portfolio optimization in the presence of unknown and uncontrolled system variables referred to as stochastic factors. We propose FaLPO (factor learning portfolio optimization). a framework that interpolates between deep policy learning and continuous-time finance models.

## Problem Definition

Notations are provided below:

- Assets prices $S_t := [S_t^1, S_t^2, \cdots S_t^{d_S}]^\top$ and factors $Y_t$.
- Risk-free return as zero.
- Terminal wealth under a policy: $Z_T^\pi$.

Portfolio optimization aims to maximize the expected terminal utility: $\mathbb{E}[U(Z_T^\pi)]$, with two examples: the power utility $U(z;\gamma) := \frac{1}{1-\gamma} z^{1-\gamma}$ with $\mathscr{Z} = \mathbb{R}^+$, $\gamma > 0$, and $\gamma \neq 1$; and the exponential utility $U(z;\gamma) := -\frac{\exp(-\gamma z)}{\gamma}$ with $\mathscr{Z} = \mathbb{R}$ and $\gamma > 0$.

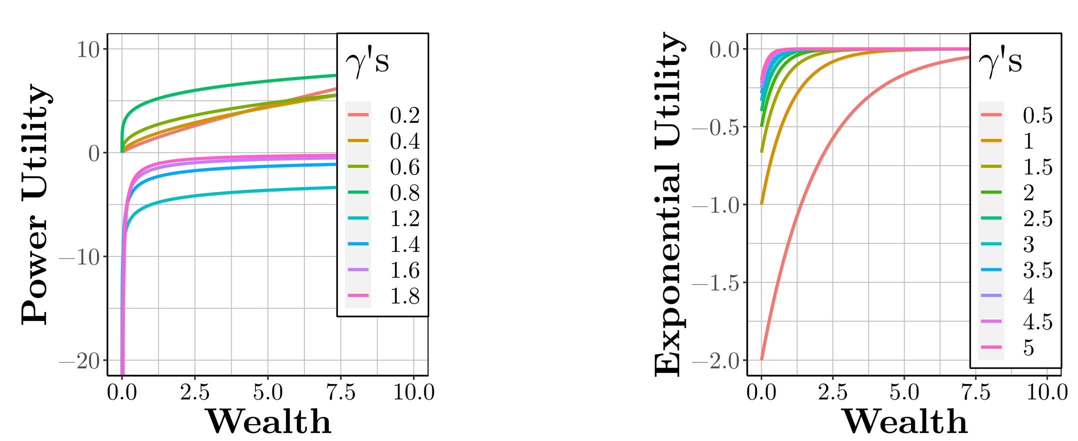


Figure: Power & exponential utilities.

## Background

**DDPG** directly maximizes the following performance objective:

$$\max_{\theta_D} V(\theta_D) \text{ with } V(\theta_D) := \mathbb{E}[U(Z_T^{\pi(\cdot;\theta_D)})]. \tag{1}$$

**Stochastic Factor Models** explicitly formulate the dynamics

$$\frac{dS_t^i}{S_t^i} = f_S^i(Y_t;\theta_S^*)dt + \sum_{j=1}^{d_W} g_S^{ij}(Y_t;\theta_S^*)dW_t^j, \quad dY_t = f_Y(Y_t;\theta_S^*)dt + g_Y(Y_t;\theta_S^*)^\top dW_t. \tag{2}$$

## FaLPO

We propose FaLPO with a neural stochastic factor model and a model-regularized policy learning method.

**Neural Stochastic Factor Model**

$$\frac{dS_t^i}{S_t^i} = f_S^i(X_t;\theta_S^*)dt + \sum_{j=1}^{d_W} g_S^{ij}(X_t;\theta_S^*)dW_t^j,$$
$$dX_t = f_X(X_t;\theta_S^*)dt + g_X(X_t;\theta_S^*)^\top dW_t,$$
$$X_t = \phi(Y_t;\theta_\phi^*).$$

**Model-Regularized Policy Learning**

- From the model we can derive the functional form of an optimal continuous-time policy:
$$\tilde{\pi}_t^* = \Pi(t, S_t, Z_t, X_t; \theta_{\tilde{\pi}}^*), \tag{3}$$
where the functional form of $\Pi$ can be obtained in many existing stochastic factor models.
- Given the specific functional forms in (2), FaLPO conducts model calibration :
$$\max_{\theta_S} L(\theta_\phi, \theta_S). \tag{4}$$

The policy learning procedure can be summarized as:

$$\max_{(\theta_\phi, \theta_\pi \theta_S) \in \mathscr{A}} H(\theta_\phi, \theta_\pi, \theta_S), \text{ with } H(\theta_\phi, \theta_\pi, \theta_S) := (1-\lambda)V(\theta_\phi, \theta_\pi) + \lambda L(\theta_\phi, \theta_S). \tag{5}$$

## Theory

**Theorem 1** Define $V_{\Delta t}^* := V(\pi^*)$ where $\pi^*$ is an optimal discrete-time admissible policy with time interval $\Delta t$, and $\theta_{\Delta t}^* := (\theta_{\phi,\Delta t}^*, \theta_{\pi,\Delta t}^*, \theta_{S,\Delta t}^*) \in \arg\max_{(\theta_\phi, \theta_\pi, \theta_S) \in \mathscr{A}} H(\theta_\phi, \theta_\pi, \theta_S)$ with the policy functional form (3) With assumptions above,

$$\lim_{\Delta t \to 0} \left(V_{\Delta t}^* - V(\theta_{\Delta t}^*)\right) = 0.$$

**Theorem 2** Finite-sample performance bounds are provided.

## Experiments

Table: Competing methods and their characteristics.

| Methods | Factor Representation | Parametric Modeling | Joint Optimization |
|---|---|---|---|
| MMMC | ✖ | ✔ | ✖ |
| DDPG | ✔ | ✖ | ✖ |
| SLAC | ✔ | ✖ | ✔ |
| RichID | ✔ | ✔ | ✖ |
| CT-MB-RL | ✖ | ✔ | ✖ |
| **FaLPO** | ✔ | ✔ | ✔ |

Table: Average terminal utility after tuning with standard deviation for synthetic data

| Annual Volatility | 0.1 | 0.2 | 0.3 |
|---|---|---|---|
| **FaLPO** | $\mathbf{-0.465} \pm 0.446$ | $\mathbf{-1.35} \pm 0.155$ | $\mathbf{-2.737} \pm 0.219$ |
| DDPG | $-1.650 \pm 0.456$ | $-3.30 \pm 1.294$ | $-5.495 \pm 1.269$ |
| SLAC | $-0.750 \pm 0.210$ | $-5.50 \pm 0.011$ | $-6.160 \pm 0.012$ |
| RichID | $-3.350 \pm 0.111$ | $-5.65 \pm 0.102$ | $-6.325 \pm 0.048$ |
| CT-MB-RL | $-2.850 \pm 0.014$ | $-5.35 \pm 0.020$ | $-6.160 \pm 0.026$ |
| MMMC | $-4.723 \pm 7.619$ | $-5.602 \pm 4.299$ | $-6.124 \pm 3.217$ |

Table: Average terminal utility for real-world data. Mix denotes a mix of stocks in the previous three sectors.

| Methods | Energy | Material | Industrials | Mix |
|---|---|---|---|---|
| **FaLPO** | $\mathbf{-2.4} \pm 1.9$ | $\mathbf{-3.2} \pm 1.0$ | $\mathbf{-6.3} \pm 2.3$ | $\mathbf{-3.5} \pm 1.5$ |
| DDPG | $-6.6 \pm 1.2$ | $-7.3 \pm 1.5$ | $-7.3 \pm 2.1$ | $-2.5 \times 10^4 \pm 3.3 \times 10^8$ |
| SLAC | $-6.8 \pm 0.2$ | $-7.0 \pm 1.5$ | $-342.4 \pm 886.8$ | $-3.0 \times 10^8 \pm 4.3 \times 10^{12}$ |
| RichID | $-6.5 \pm 0.1$ | $-6.9 \pm 1.4$ | $-6.9 \pm 0.4$ | $-8.1 \pm 3.9$ |
| CT-MB-RL | $-4.2 \pm 6.2$ | $-5.4 \pm 4.3$ | $-11655 \pm 32947.5$ | $-5.7 \pm 3.1$ |
| MMMC | $-8.5 \pm 7.6$ | $-6.5 \pm 1.7$ | $-11.0 \pm 5.4$ | $-7.5 \pm 4.4$ |

## References

[1] Tong Suk Kim and Edward Omberg. Dynamic nonmyopic portfolio behavior. *The Review of Financial Studies*, 9(1):141–161, 1996.

[2] David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In *International conference on machine learning*, pages 387–395. PMLR, 2014.