A PRELIMINARIES OF ECONOMIC AND FINANCIAL THEORY

A.1 UTILITY THEORY

Utility theory is the foundation of the choice theory under uncertainty including economics and financial theory. A utility function $u(\cdot)$ measures agent's relative preference for different levels of total wealth W. According to Morgenstern & Von Neumann (1953), a rational agent makes an investment decision to maximize the expected utility of wealth among a set of competing feasible investment alternatives. For simplicity, the following assumptions are often made for the utility function used in economics and finance. First, the utility function is assumed to be at least twice continuous differentiable. The first derivative of the utility function (the marginal utility of wealth) is always positive, i.e., $U^{'}(W) > 0$, because of the assumption of non-satiation. The second assumption concerns risk attitude of the agents called "risk averse". When we assume that an agent is risk averse, the utility function is described as a curve that increases monotonically and is concave. The most often used utility function of a risk-averse agent is the quadratic utility function as follows:

$$u(W; \alpha, \beta) = \alpha W - \beta \frac{1}{2} W^2 \tag{5}$$

where $\alpha \geq 0, \beta > 0$. Taking the expected value of the quadratic utility function in (5) yields:

$$E[u(W; \alpha, \beta)] = \alpha \mathbb{E}[W] - \frac{1}{2} \beta \mathbb{E}[W^2] \tag{6}$$

Substituting $\mathbb{E}[W^2] = \mathbb{V}[W] + \mathbb{E}[W]^2$ into (5) gives

$$\mathbb{E}[u(W; \alpha, \beta)] = \alpha \mathbb{E}[W] - \frac{1}{2} \beta (\mathbb{V}[W] + \mathbb{E}[W]^2) \tag{7}$$

Equation (7) shows that expected quadratic utility can be described in terms of mean $E[W]$ and variance $\mathbb{V}[W]$ of wealth. Therefore, the assumption of a quadratic utility function is crucial to the mean-variance analysis.

Remark 5 (Approximation by quadratic utility function)**.** Readers may be interested in how the quadratic utility function approximates other risk averse utility functions. Kroll et al. (1984) empirically answered this question by comparing MV portfolio (maximizer of expected quadratic utility function) and maximizers of other utility functions. In their study, maximizers of other utility functions also almost located in MV Pareto efficient frontier; that is, expected quadratic utility function approximates other risk averse utility function well.

Remark 6 (Non-vNM utility functions)**.** Unlike MV trade-off, utility functions maximized by some recently proposed new risk criteria, such as VaR and Prospect theory, do not belong to traditional vNM utility function.

A.2 MARKOWITZ'S PORTFOLIO

Considering the mean-variance trade-off in a portfolio and economic activity is an essential task in economics as Tamar et al. (2012) and Xie et al. (2018) pointed out. The mean-variance trade-off is justified by assuming either quadratic utility function to the economic agent or multivariate normal distribution to the financial asset returns (Borch, 1969; Baron, 1977; Luenberger et al., 1997). This means that if either the agent follows the quadratic utility function or asset return follows the normal distribution, the agent's expected utility function is maximized by maximizing the expected reward and minimizing the variance. Therefore, the goal of Markowitz's portfolio is not only to construct the portfolio itself but also to maximize the expected utility function of the agent. See Figure 3.

Markowitz (1952) proposed the following steps for constructing a MV controlled portfolio (Also see Markowitz (1959), page 288, and Luenberger et al. (1997)):

- Constructing portfolios minimizing the variance under several reward constraint;
- Among the portfolios constructed in the first step, the economic agent chooses a portfolio maximizing the utility function.

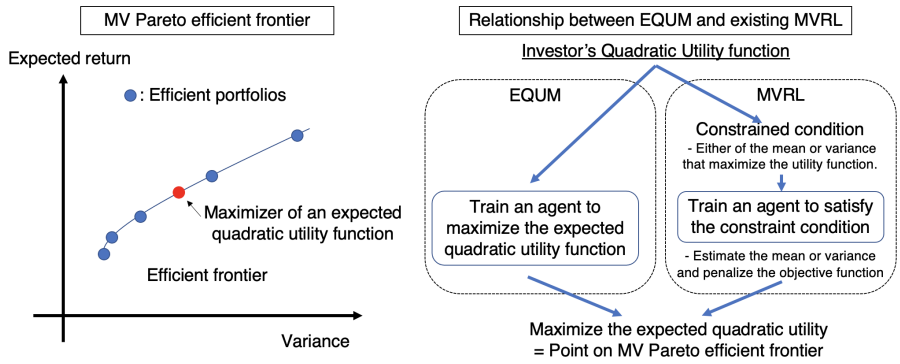

Figure 3: The concept of MV Preto efficient frontier and difference between EQUMRL and existing MVRL approaches.

Conventional financial methods often adopt this two-step approach because directly predicting the reward and variance to maximize the expected utility function is difficult; therefore, first gathering information based on analyses of an economist, then we construct the portfolio using the information and provide the set of the portfolios to an economic agent. However, owing to the recent development of machine learning, we can directly represent the complicated economic dynamics using flexible models, such as deep neural networks. In addition, as Tamar et al. (2012) and Xie et al. (2018) reported, when constructing the mean-variance portfolio in RL, we suffer from the double sampling issue. Therefore, this paper aims to achieve the original goal of the mean-variance approach; that is, the expected utility maximization. Note that this idea is not restricted to financial applications but can be applied to applications where the agent utility can be represented only by the mean and variance.

A.3 Markowitz's portfolio and capital asset pricing model

Markowitz's portfolio is known as the mean-variance portfolio (Markowitz, 1952; Markowitz & Todd, 2000). Constructing the mean-variance portfolio is motivated by the agent's expected utility maximization. When the utility function is given as the quadratic utility function, or the financial asset returns follow the multivariate normal distribution, a portfolio maximizing the agent's expected utility function is given as a portfolio with minimum variance under a certain standard expected reward.

The Capital Asset Pricing Model (CAPM) theory is a concept which is closely related to Markowitz's portfolio (Sharpe, 1964; Lintner, 1965; Mossin, 1966). This theory theoretically explains the expected return of investors when the investor invests in a financial asset; that is, it derives the optimal price of the financial asset. To derive this theory, as well as Markowitz's portfolio, we assume the quadratic utility function to the investors or the multivariate normal distribution to the financial assets.

Merton (1969) extended the static portfolio selection problem to a dynamic case. Fishburn & Burr Porter (1976) studied the sensitivity of the portfolio proportion when the safe and risky asset distributions change under the quadratic utility function. Thus, there are various studies investigating relationship between the utility function and risk-averse optimization (Tobin, 1958; Kroll et al., 1984; Bulmuş & Özekici, 2014; Bodnar et al., 2015a;b).

A.4 MV portfolio and MVRL

Traditional portfolio theory have attempted to maximize the expected quadratic utility function by providing MV portfolios. This is because MV portfolio is easier to interpret than EQUMRL, and we can obtain a solution by quadratic programming. One of the main goals of MVRL methods is also to construct MV portfolio under a dynamic environment (Tamar et al., 2012). However, we conjecture that there are three significant differences between them. First, unlike static problem, MVRL suffers computational difficulties. Second, while static problem solves quadratic programming given the expected reward and variance, MVRL methods simultaneously estimate these values and solve the constrained problem. Third, while static MV portfolios gives us an exact solution of the constrained problem, MVRL often relaxes the constrained problem by penalized method, which

causes approximation errors. In particular, for the second and third points, the difference in how to handle the estimators of expected reward and variance is essential.

A.5 Empirical studies on the utility functions

The standard financial theory is built on the assumption that the economic agent has the quadratic utility function. For supporting this theory, there are several empirical studies to estimate the parameters of the quadratic utility function. Markowitz & Todd (2000) discussed how the quadartic utility function approximates the other risk-averse utility functions. Ziemba et al. (1974) investigated the change of the portfolio proportion when the parameter of the quadratic utility function changes using the Canadian financial dataset. Recently, Bodnar et al. (2018) investigate the risk parameter (α and β in our formulation of the quadratic utility function) using the markets indexes in the world. They found that the utility function parameter depends on the market data model.

A.6 Criticism

For the simple form of the quadratic utility function, the financial models based on the utility are widely accepted in practice. However, there is also criticism that the simple form cannot capture the real-world complicated utility function. For instance, Kallberg et al. (1983) criticized the use of the quadratic utility function and proposed using a utility function, including higher moments. This study also provided empirical studies using U.S. financial dataset for investigating the properties of the alternative utility functions. However, to the best of our knowledge, financial practitioners still prefer financial models based on the quadratic utility function. We consider this is because the simple form gains the interpretability of the financial models.

A.7 Economics and finance

To mathematically describe an attitude toward risk, economics and finance developed expected utility theory, which assumes the Bernoulli utility function $u(R)$ on an agent. In the expected utility theory, an agent acts to maximize the von Neumann-Morgenstern (vNM) utility function $U(F(r)) = \int u(r)dF(r)$, where $F(r)$ is a distribution of R (Mas-Colell et al., 1995). We can relate the utility function form to agents with three different risk preferences: the utility function $u(R)$ is concave for risk averse agents; linear for risk neutral agents; and convex risk seeking agents. For instance, an agent with $u(R) = R$ corresponds to a risk neutral agent attempting to maximize their expected cumulative reward in a standard RL problem. For more detailed explanation, see Appendix A or standard textbooks of economics and finance, such as Mas-Colell et al. (1995) and Luenberger et al. (1997).

To make the Bernoulli utility function more meaningful, we assume that it is increasing function with regard to R; that is, $R \leq \frac{\alpha}{\beta}$ for all possible R. Even without the assumption, for a given pair of (α, β), a optimal policy maximizing the expected utility does not change; that is, the assumption is only related to interpretation of the quadratic utility function and does not affect the optimization.

On the other hand, the constraint condition is determined by an investor to maximize its expected utility (Luenberger et al., 1997). Finally, in theories of economics and finance, investors can maximize their utility by choosing a portfolio from MV portfolios. In addition, when $R \leq \frac{\alpha}{\beta}$ for all possible value of R, a policy maximizing the expected utility function is also located on MV Pareto efficient frontier.

B Per-step variance perspective

Bisi et al. (2020) defines the per-step reward random variable R, a discrete random variable taking the values in the image of r, by defining its probability mass function as $p(R = x) = \sum_{s,a} d_{\pi_\theta}(s,a)\mathbb{1}[r(s,a) = x]$, where $\mathbb{1}$ is the indicator function and d_{π_θ} is the normalized discounted state-action distribution, $d_{\pi_\theta} = (1-\gamma)\sum_{t=0}^{\infty}\gamma^t \mathrm{Pr}_{\mu_0,\pi_\theta,p}(S_t = s, A_t = a)$. Here, it is known that for $\gamma < 1$, $\mathbb{E}_{\pi_\theta}[G] = \frac{1}{1-\gamma}\sum_{s,a} d_{\pi_\theta}(s,a)r(s,a)$. It follows that $\mathbb{E}_{\pi_\theta}[R] = (1-\gamma)J(\pi_\theta)$. Bisi et al. (2020) showed that the per-step variance $\mathbb{V}_{\pi_\theta}(R) \leq \frac{\mathbb{V}_{\pi_\theta}(r(S_t,A_t))}{(1-\gamma)^2}$, which implies that the minimiza-

tion of the per-step variance $\mathbb{V}_{\pi_\theta}(r(S_t, A_t))$ also minimizes trajectory-variance $\mathbb{V}_{\pi_\theta}(R)$. Therefore, they train a policy π_θ by maximizing $J_\kappa(\pi) = \mathbb{E}_{\pi_\theta}[r(S_t, A_t)] - \kappa\mathbb{V}_{\pi_\theta}(r(S_t, A_t))$, where $\kappa > 0$ is a parameter of the penalty function. Bisi et al. (2020) reveals that $J_\kappa(\pi) = \mathbb{E}_{\pi_\theta}[R - \kappa(R - \mathbb{E}_{\pi_\theta}[R])^2]$, which implies that optimizing $J_\kappa(\theta)$ is equivalent to optimize the canonical risk-neutral objective of a new MDP with the same as the original MDP except that the new reward function $r'(s, a) = r(s, a) - \lambda\big(r(s, a) - (1 - \gamma)J(\pi)\big)^2$. This reward function depends on the policy π, making the reward function nonstationary and conventional RL method unusable. This problem is called policy-dependent-reward issue. To solve this policy-dependent-reward issue, the methods of Bisi et al. (2020) and Zhang et al. (2021b) are based on the trust region policy optimization (Schulman et al., 2015) and coordinate descent with Legendre-Fenchel duality (Xie et al., 2018), respectively.

For constants $\alpha, \beta > 0$, we define the quadratic utility function for per-step reward setting as

$$u^{\text{per-step}}(r(S_t, A_t); \alpha, \beta) = \alpha r(S_t, A_t) - \beta r^2(S_t, A_t).$$

For an infinite horizon case with $\gamma < 1$, we consider maximizing the following expected cumulative quadratic utility function:

$$\mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{\infty}\gamma^{t-1}u^{\text{per-step}}(r(S_t, A_t); \alpha, \beta)\right] = \alpha\mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{\infty}\gamma^{t-1}r(S_t, A_t)\right] - \beta\mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{\infty}\gamma^{t-1}r^2(S_t, A_t)\right]$$

For an finite horizon case with $\gamma = 1$, we consider maximizing the following expected cumulative quadratic utility function:

$$\mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T}u^{\text{per-step}}(r(S_t, A_t); \alpha, \beta)\right] = \alpha\mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T}r(S_t, A_t)\right] - \beta\mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T}r^2(S_t, A_t)\right]$$

When applying REINFORCE-based algorithm to train an agent by maximizing the objective function, the sample approximation of the gradient, for instance, is given as

$$\begin{aligned}
&\widehat{\nabla}\mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T}u^{\text{per-step}}(r(S_t^k, A_t^k); \alpha, \beta)\right] \\
&= \alpha\sum_{t=1}^{T}r(S_t^k, A_t^k)\nabla_\theta \log \pi_\theta(S_t^k, A_t^k) - \beta\sum_{t=1}^{T}\big(r(S_t^k, A_t^k)\big)^2\nabla_\theta \log \pi_\theta(S_t^k, A_t^k)
\end{aligned}$$

We investigate the empirical performance using the same setting as the portfolio management experiment of Section 6.1. Let ζ be α/β in the trajectory EQUMRL, and ρ be α/β in the per-step EQUMRL. Under both the trajectory and per-step EQUMRLs, we train policies with the REINFORCE-based algorithm. The other settings are identical to that in Section 6.1. As Figure 4 shows, under both the trajectory and per-step EQUMRLs, the REINFORCE-based algorithms train MV-efficient policies well for the parameters ζ and ρ.

C AMERICAN-STYLE OPTION WITH A SYNTHETIC DATASET

An American-style option refers to a contract that we can execute an option right at any time before the maturity time τ; that is, a buyer who bought a call option has a right to buy the asset with the strike price K_{call} at any time; a buyer who bought a put option has a right to sell the with the strike price K_{put} at any time.

In the setting of Tamar et al. (2014) and Xie et al. (2018), the buyer simultaneously buy call and put options, which have the strike price $K_{\text{call}} = 1.5$ and $K_{\text{put}} = 1.$, respectively. The maturity time is set as $\tau = 20$. If the buyer executes the option at time t, the buyer obtains a reward $r_t = \max(0, K_{\text{put}} - x_t) + \max(0, x_t - W_{\text{call}})$, where x_t is an asset price. We set $x_0 = 1$ and define the stochastic process as follows: $x_t = x_{t-1}f_u$ with probability 0.45 and $x_t = x_{t-1}f_d$ with probability 0.55, where f_u and f_d. These parameters follows Xie et al. (2018).

As well as Section 6.1, we compare the EQUMRL with policy gradient (EQUM) with the REINFORCE, Tamar, and Xie. The other settings are also identical to Section 6.1. We show performances under several hyperparameter in Figure 5 and CRs and Vars of some of their results in the upper table

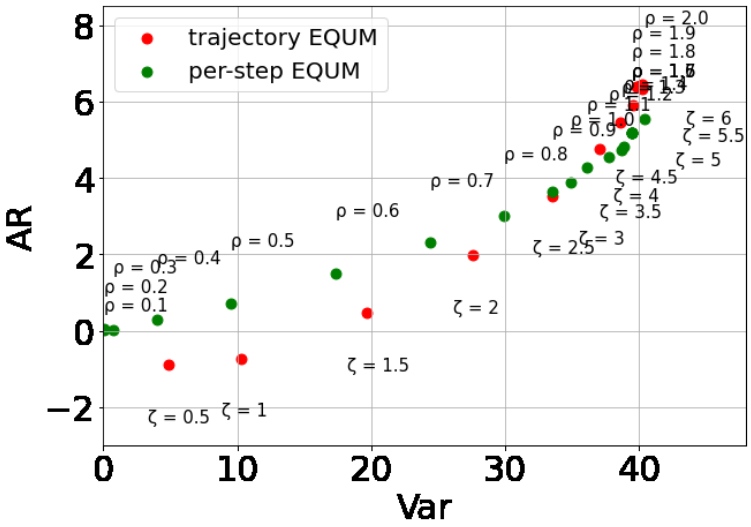

Figure 4: MV efficiency of the portfolio management experiment with the trajectory and per-step EQUMRLs. Higher CRs and lower Vars methods are MV Pareto efficient.

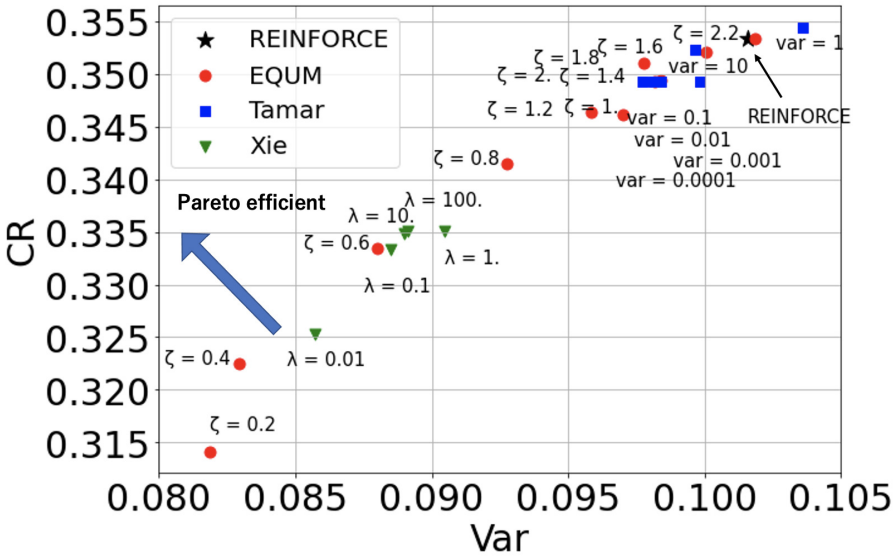

Figure 5: MV efficiency of the American-style option experiment. Higher CRs and lower Vars methods are MV Pareto efficient.

of Table 3. We also denote MSE between ζ and achieved CR of in the lower table of Table 3. From the table, we can find that while EQUMRL minimize the MSE for lower ζ, the MSE of REINFORCE is smaller for higher ζ. We consider this is because owing to the difficulty of MV control under this setting, naively maximizing the CR minimizes the MSE more than considering the MV trade-off for higher ζ.

D Details of experiments of portfolio management with a real-world dataset

We use the following portfolio models. An equally-weighted portfolio (EW) weights the financial assets equally (DeMiguel et al., 2009). A mean-variance portfolio (MV) computes the optimal variance under a mean constraint (Markowitz, 1952). To compute the mean vector and covariance matrix, we use the latest 10 years (120 months) data. A Kelly growth optimal portfolio with ensemble learning (EGO) is proposed by Shen et al. (2019). We set the number of resamples as

Table 3: The results of American-style option. The upper table shows the CRs and Vars over $100,000$ trials. The lower table shows The lower table shows the MSEs for ζ over $100,000$ trials.

	REINFORCE	EQUM			Tamar		Xie	
	$(\zeta = \infty)$	$\zeta = 1.6$	$\zeta = 0.8$	$\zeta = 0.4$	$\mathbb{V} = 1$	$\mathbb{V} = 0.1$	$\lambda = 100$	$\lambda = 0.1$
CR	0.353	0.352	0.341	0.322	0.352	0.349	0.335	0.333
Var	0.099	0.098	0.092	0.083	0.102	0.096	0.089	0.088

	REINFORCE	EQUM				
Target Value	$(\zeta = \infty)$	$\zeta = 2.2$	$\zeta = 1.8$	$\zeta = 1.4$	$\zeta = 1.0$	$\zeta = 0.6$
MSE from $\zeta = 2.2$	**3.512**	**3.512**	3.517	3.523	3.532	3.572
$\zeta = 1.8$	**2.194**	2.195	2.197	2.203	2.209	2.239
$\zeta = 1.4$	**1.197**	**1.197**	1.198	1.202	1.206	1.226
$\zeta = 1.0$	0.520	0.520	**0.519**	0.522	0.523	0.532
$\zeta = 0.6$	0.162	0.163	0.160	0.161	0.160	**0.159**

$m_1 = 50$, the size of each resample $m_2 = 5\tau$, the number of periods of return data $\tau = 60$, the number of resampled subsets $m_3 = 50$, and the size of each subset $m_4 = n^{0.7}$, where m is number of assets; that is, $m = 25$ in FF25, $m = 48$ in FF48 and $m = 100$ in FF100. Portfolio blending via Thompson sampling (BLD) is proposed by Shen & Wang (2016). We use the latest 10 years (120 months) data to compute for the sample covariance matrix and blending parameters. Denote a method proposed by Tamar et al. (2012) by Tamar, and choose the parameter var from $\{15, 30, 60\}$. Denote a method proposed by Xie et al. (2018) by Xie, and choose the parameter λ from $\{10, 100, 1000\}$. Denote the REINFORCE-based trajectory EQUMRL by EQUMRL, and choose the parameter ζ from $\{0.5, 0.75, 1.5\}$. To train the policies of MVRL methods, we assume the stationarity on time-series. Given historical datasets $(y_{a,t})_{t\in\{1,\dots,T\},a\in\{1,\dots,m\}}$, for each trajectory, we simulate portfolio management and generate $\{(S_t^k, A_t^k, r(S_t^k, A_t^k))\}_{t=1}^n$. Recall that state is given as $S_t = ((y_{a,t-1}, ..., y_{a,t-12}, w_{a,t-1})_{a\in\mathcal{A}}, \sum_{s=0}^{t-1} y_s)$, and the portfolio weight as given as $w_{a,t} = \pi(a, s_t)$ (see Section 6.2). Then, using the trajectory observations, we train the policies. We show the pseudo-code of the modified REINFORCE-based trajectory EQMRL in Algorithm 2.

Algorithm 2 REINFORCE-based EQUMRL in Section 6.2

Historical dataset $(y_{a,t})_{t\in\{1,\dots,T\},a\in\{1,\dots,m\}}$
Initialize the policy parameter θ_0;
for $k = 1, 2, \dots$ **do**
  Generate $\{(S_t^k, A_t^k, r(S_t^k, A_t^k))\}_{t=1}^n$ on π_θ by simulating the portfolio management using the historical dataset $(y_{a,t})_{t\in\{1,\dots,T\},a\in\{1,\dots,m\}}$;
  Update policy parameters $\theta_{k+1} \leftarrow \theta_k + \eta_i \widehat{\nabla} \mathbb{E}_{\pi_{\theta_k}} \left[u^{\text{trajectory}}(G; \alpha, \beta) \right]$.
end for

Table 4 shows the performance of each portfolio model without the penalty of the turnover $\Lambda = 0$. We report both results with and without the penalty of the turnover in the following. The average of real objective (minimizing variance with a penalty on return targeting) for Tamar, MVP and EQUMRL from July 2000 to June 2020 is shown in Table 5. We also divide the performance period into two for robustness checks. Table 6 and 7 shows the first-half results from July 2000 to June 2010 and the second-half results from July 2010 to June 2020. In almost all cases, the EQUMRL achieves the highest R/R.

We plot results of Tamar, Xie, and EQUMRL with various parameters in Figure 6. We choose the parameter of Tamar ($\mathbb{V}$) from $\{10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65\}$, Xie ($\lambda$) from $\{10, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$, and EQUMRL ($\zeta$) from $\{0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1, 1.125, 1.25, 1.375, 1.5\}$. We only annotate the points of $\mathbb{V} = 10, 25, 40, 55$, $\lambda = 10, 200, 500, 800$, and $\zeta = 0.125, 0.5, 0.875, 1.25$. Unlike Section 6.1, it is not easy to control the mean and variance owing to the difficulty of predicting the real-world financial markets. However, the EQUMRL tends to return more efficient results than the other methods.

Table 4: The performance of each portfolio model without the penalty of the turnover $\Lambda = 0$ during the out-of-sample period (from July 2000 to June 2020) for FF25 dataset (upper table) , FF48 (middle table), and FF100 (lower table). For each dataset, the best performance is highlighted in **bold**.

Method	EW	MV	EGO	BLD	Tamar			Xie			EQUM		
					$\mathbb{V}=15$	$\mathbb{V}=30$	$\mathbb{V}=60$	$\lambda=10$	$\lambda=100$	$\lambda=1000$	$\zeta=0.5$	$\zeta=0.75$	$\zeta=1.5$
FF25													
CR↑	0.80	0.11	0.87	0.56	**1.16**	1.00	0.80	0.96	1.01	0.90	1.11	1.12	1.02
Var↓	28.62	53.58	30.63	12.01	18.41	15.10	25.22	24.94	14.98	18.37	**11.77**	21.29	16.25
R/R↑	0.52	0.05	0.55	0.56	0.94	0.89	0.55	0.67	0.90	0.73	**1.13**	0.84	0.87
MaxDD↓	0.54	0.75	0.57	0.37	0.35	0.35	0.55	0.46	0.34	0.49	**0.27**	0.33	0.30
FF48													
CR↑	0.81	0.15	1.04	0.52	0.90	0.63	1.06	0.52	0.34	0.35	0.92	**1.49**	1.01
Var↓	22.91	76.89	31.87	9.65	40.13	15.81	18.07	11.79	6.72	**6.70**	15.11	30.96	27.28
R/R↑	0.59	0.06	0.64	0.58	0.49	0.55	0.86	0.52	0.46	0.47	0.82	**0.93**	0.67
MaxDD↓	0.30	0.48	0.31	0.18	0.32	0.19	0.21	0.20	0.17	**0.17**	0.20	0.27	0.28
FF100													
CR↑	0.81	0.14	0.86	0.53	1.07	**1.86**	1.36	1.33	1.05	1.07	1.35	1.30	1.38
Var↓	29.36	57.90	32.40	**11.79**	31.61	85.74	20.05	43.65	20.94	87.24	20.62	16.56	27.50
R/R↑	0.52	0.06	0.53	0.54	0.66	0.69	1.05	0.70	0.79	0.40	1.03	**1.10**	0.91
MaxDD↓	0.33	0.46	0.34	0.22	0.29	0.57	0.22	0.36	0.23	0.51	0.23	**0.20**	0.28

Table 5: The average of real objective (minimizing variance with a penalty on return targeting) for Tamar, MVP and EQUMRL for FF25 dataset (upper panel) , FF48 (middle panel) and FF100 (lower panel) from July 2020 to June 2020.

	Tamar			MVP			EQUM		
	$\mathbb{V}=15$	$\mathbb{V}=30$	$\mathbb{V}=60$	$\lambda=10$	$\lambda=100$	$\lambda=1000$	$\zeta=0.5$	$\zeta=0.75$	$\zeta=1.5$
With Turnover Penalty $\Lambda = 0.001$									
FF25	-5,466	-7,504	-7,646	-6,227	-6,398	-7,769	-7,853	-8,758	-9,027
FF48	-5,395	-7,399	-7,614	-6,647	-7,763	-7,677	-9,000	-8,436	-7,743
FF100	-6,711	-7,491	-6,233	-6,362	-5,354	-6,983	-8,681	-7,366	-8,362
Without Turnover Penalty $\Lambda = 0$									
FF25	-5,312	-7,486	-7,329	-6,130	-6,390	-7,340	-7,664	-7,660	-8,911
FF48	-5,464	-7,053	-7,505	-6,489	-9,763	-7,506	-8,729	-8,456	-7,421
FF100	-6,749	-7,287	-5,915	-6,047	-5,446	-7,008	-9,484	-7,306	-8,280

Table 6: The performance of each portfolio without turnover penalty ($Lambda = 0$) during first half out-of-sample period (from July 2000 to June 2010) and second half out-of-sample period (from July 2010 to June 2020) for FF25 dataset (upper panel) , FF48 (middle panel) and FF100 (lower panel). Among the comparisons of the various portfolios, the best performance within each dataset is highlighted in **bold**.

FF25	EW	MV	EGO	BLD	Tamar			Xie			EQUM		
					$\mathbb{V}=15$	$\mathbb{V}=30$	$\mathbb{V}=60$	$\lambda=10$	$\lambda=100$	$\lambda=1000$	$\zeta=0.5$	$\zeta=0.75$	$\zeta=1.5$
First-Half Period (from July 2000 to June 2010)													
CR↑	0.58	-0.42	0.64	0.41	1.35	1.31	1.18	0.96	1.32	1.15	**1.54**	1.34	1.22
Var↓	31.21	69.36	33.29	12.50	14.47	11.34	11.78	21.38	10.99	11.50	**9.96**	22.79	17.10
R/R↑	0.36	-0.17	0.38	0.41	1.23	1.35	1.19	0.72	1.38	1.18	**1.69**	0.97	1.02
MaxDD↓	0.54	0.75	0.58	0.37	0.21	0.16	0.19	0.32	0.15	**0.19**	0.10	0.33	0.21
Second-Half Period (from July 2000 to June 2010)													
CR↑	1.02	0.63	**1.03**	0.70	0.98	0.68	0.43	0.96	0.70	0.65	0.69	0.90	0.81
Var↓	25.92	37.24	27.92	**11.47**	22.30	18.66	38.39	28.50	18.77	25.10	13.21	19.69	15.32
R/R↑	0.69	0.36	0.67	0.72	0.72	0.55	0.24	0.62	0.56	0.45	0.66	0.70	**0.72**
MaxDD↓	0.31	0.50	0.31	**0.22**	0.35	0.35	0.55	0.46	0.34	0.49	0.27	0.32	0.30
FF48	EW	MV	EGO	BLD	Tamar			Xie			EQUM		
					$\mathbb{V}=15$	$\mathbb{V}=30$	$\mathbb{V}=60$	$\lambda=10$	$\lambda=100$	$\lambda=1000$	$\zeta=0.5$	$\zeta=0.75$	$\zeta=1.5$
First-Half Period (from July 2000 to June 2010)													
CR↑	0.60	0.41	0.75	0.34	1.57	0.68	1.35	1.03	-0.25	-0.26	1.23	**1.72**	1.09
Var↓	25.85	57.50	39.33	10.97	41.06	14.46	13.85	7.07	0.02	**0.02**	11.15	27.16	18.30
R/R↑	0.41	0.19	0.41	0.36	0.85	0.62	1.26	**1.34**	-6.36	-6.30	1.28	1.15	0.88
MaxDD↓	0.27	0.33	0.31	0.18	0.30	0.17	0.17	0.17	0.00	**0.00**	0.16	0.27	0.18
Second-Half Period (from July 2000 to June 2010)													
CR↑	1.02	-0.12	**1.33**	0.69	0.23	0.59	0.77	0.01	0.94	0.97	0.61	1.27	0.92
Var↓	19.88	96.14	24.24	**8.27**	38.29	17.16	22.13	16.00	12.72	12.63	18.87	34.66	36.24
R/R↑	0.79	-0.04	0.93	0.84	0.13	0.49	0.57	0.01	0.91	**0.94**	0.48	0.75	0.53
MaxDD↓	0.27	0.48	0.27	0.18	0.28	0.19	0.21	0.17	0.17	**0.17**	0.20	0.26	0.28
FF100	EW	MV	EGO	BLD	Tamar			MVP			EQUM		
					$\mathbb{V}=15$	$\mathbb{V}=30$	$\mathbb{V}=60$	$\lambda=10$	$\lambda=100$	$\lambda=1000$	$\zeta=0.5$	$\zeta=0.75$	$\zeta=1.5$
First-Half Period (from July 2000 to June 2010)													
CR↑	0.61	-0.37	0.73	0.41	1.08	1.88	1.67	1.42	1.29	0.73	1.43	1.50	**1.93**
Var↓	32.02	79.35	35.11	**12.43**	19.67	109.44	14.33	40.80	13.16	92.80	17.16	12.80	21.45
R/R↑	0.37	-0.15	0.43	0.40	0.84	0.62	**1.52**	0.77	1.23	0.26	1.20	1.45	1.44
MaxDD↓	0.28	0.46	0.30	0.17	0.22	0.52	0.18	0.30	**0.15**	0.51	0.20	0.17	0.22
Second-Half Period (from July 2000 to June 2010)													
CR↑	1.01	0.66	1.00	0.66	1.06	**1.83**	1.06	1.24	0.80	1.41	1.27	1.09	0.83
Var↓	26.61	35.92	29.64	**11.12**	43.56	62.05	25.58	46.48	28.61	81.45	24.07	20.23	32.94
R/R↑	0.68	0.38	0.63	0.68	0.56	0.81	0.73	0.63	0.52	0.54	**0.90**	0.84	0.50
MaxDD↓	0.32	0.33	0.32	0.22	0.27	0.34	0.22	0.33	0.23	0.45	0.22	**0.19**	0.27

Table 7: The performance of each portfolio with turnover penalty ($\Lambda = 0.001$) during first half out-of-sample period (from July 2000 to June 2010) and second half out-of-sample period (from July 2010 to June 2020) for FF25 dataset (upper panel) , FF48 (middle panel) and FF100 (lower panel). Among the comparisons of the various portfolios, the best performance within each dataset is highlighted in **bold**.

FF25	EW	MV	EGO	BLD	Tamar			Xie			EQUM		
					$\mathbb{V}=2.5$	$\mathbb{V}=15$	$\mathbb{V}=30$	$\lambda=60$	$\lambda=100$	$\lambda=1000$	$\zeta=0.5$	$\zeta=0.75$	$\zeta=1.5$
First-Half Period(from July 2000 to June 2010)													
CR↑	0.58	-0.43	0.64	0.56	1.32	1.31	1.19	0.93	1.17	1.17	**1.58**	1.55	1.54
Var↓	31.21	69.33	33.29	25.08	14.46	11.18	11.81	11.95	11.93	11.93	16.47	11.18	**9.96**
R/R↑	0.36	-0.18	0.38	0.39	1.21	1.36	1.20	0.93	1.17	1.17	1.35	1.61	**1.69**
MaxDD↓	0.54	0.75	0.58	0.52	0.21	0.15	0.19	0.20	0.17	0.17	0.26	0.13	**0.10**
Second-Half Period(from July 2000 to June 2010)													
CR↑	1.02	0.60	0.99	1.12	0.97	0.70	0.41	0.87	1.13	1.13	**1.48**	0.95	0.69
Var↓	25.92	37.28	27.95	19.09	22.54	18.72	38.46	39.97	39.61	39.61	32.08	18.67	**13.21**
R/R↑	0.69	0.34	0.65	0.89	0.71	0.56	0.23	0.48	0.62	0.62	**0.90**	0.76	0.66
MaxDD↓	0.31	0.52	0.32	**0.25**	0.35	0.35	0.56	0.54	0.51	0.51	0.36	0.31	0.27
FF48	EW	MV	EGO	BLD	Tamar			Xie			EQUM		
					$\mathbb{V}=2.5$	$\mathbb{V}=15$	$\mathbb{V}=30$	$\lambda=60$	$\lambda=100$	$\lambda=1000$	$\zeta=0.5$	$\zeta=0.75$	$\zeta=1.5$
First-Half Period(from July 2000 to June 2010)													
CR↑	0.60	0.36	0.70	0.49	1.36	0.41	1.11	0.58	1.21	0.82	1.35	1.18	**1.36**
Var↓	25.85	57.76	39.39	18.69	41.27	14.49	13.70	18.51	17.61	19.28	29.97	20.19	**6.04**
R/R↑	0.41	0.16	0.38	0.39	0.73	0.37	1.04	0.47	1.00	0.64	0.85	0.91	**1.92**
MaxDD↓	0.27	0.33	0.31	0.25	0.30	0.18	0.17	0.18	0.20	0.19	0.25	0.17	**0.10**
Second-Half Period(from July 2000 to June 2010)													
CR↑	1.02	-0.14	1.25	1.02	-0.00	0.36	0.52	0.43	0.94	1.20	**1.86**	0.92	0.45
Var↓	19.88	96.16	24.28	**13.14**	39.17	17.52	22.21	35.72	29.90	33.27	33.85	17.15	14.22
R/R↑	0.79	-0.05	0.88	0.98	-0.00	0.29	0.38	0.25	0.59	0.72	**1.11**	0.77	0.41
MaxDD↓	0.27	0.48	0.28	0.22	0.28	0.20	0.21	0.27	0.25	0.26	0.27	0.20	**0.16**
FF100	EW	MV	EGO	BLD	Tamar			MVP			EQUM		
					$\mathbb{V}=2.5$	$\mathbb{V}=15$	$\mathbb{V}=30$	$\lambda=60$	$\lambda=100$	$\lambda=1000$	$\zeta=0.5$	$\zeta=0.75$	$\zeta=1.5$
First-Half Period(from July 2000 to June 2010)													
CR↑	0.61	-0.41	0.67	0.59	1.07	0.98	1.23	0.24	1.26	**1.39**	1.19	1.17	1.22
Var↓	32.02	79.44	35.06	24.25	14.20	13.92	11.39	93.47	16.00	28.49	13.42	**9.16**	17.63
R/R↑	0.37	-0.16	0.39	0.41	0.99	0.91	1.26	0.09	1.09	0.90	1.12	**1.34**	1.01
MaxDD↓	0.28	0.46	0.30	0.25	0.17	0.18	0.18	0.51	0.21	0.21	0.17	**0.15**	0.22
Second-Half Period(from July 2000 to June 2010)													
CR↑	1.01	0.63	0.95	1.12	1.18	0.60	0.11	0.90	0.70	1.09	0.70	0.72	**1.64**
Var↓	26.61	35.96	29.61	19.27	36.80	26.69	20.99	82.15	41.40	53.59	**17.46**	19.26	48.45
R/R↑	0.68	0.37	0.60	**0.88**	0.67	0.40	0.08	0.34	0.38	0.52	0.58	0.57	0.82
MaxDD↓	0.32	0.33	0.32	0.26	0.25	0.23	0.22	0.45	0.32	0.31	**0.19**	0.26	0.31

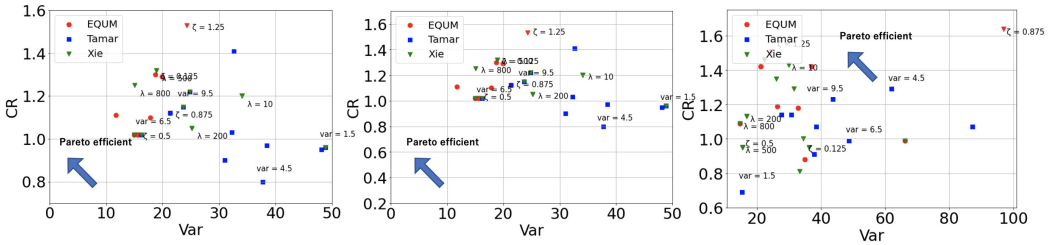

Figure 6: MV efficiency of the portfolio management experiment. Higher CRs and lower Vars methods are MV Pareto efficient. Left graph: FF25. Center graph: FF48. Right graph: FF100.

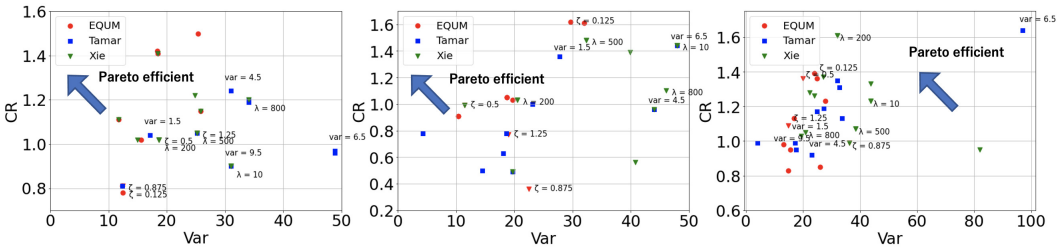

Figure 7: MV efficiency of the portfolio management experiment. Higher CRs and lower Vars methods are MV Pareto efficient. Left graph: FF25. Center graph: FF48. Right graph: FF100.