Solving dynamic portfolio selection problems via score-based diffusion models

Ahmad Aghapour
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109
aghapour@umich.edu

Erhan Bayraktar
Department of Mathematics
University of Michigan
Ann Arbor, MI 48109
erhan@umich.edu

Fengyi Yuan
School of Science and Engineering
The Chinese University of Hong Kong, Shenzhen
Shenzhen, Guangdong, China
yuanfengyi@cuhk.edu.cn

Abstract

In this work, we tackle the dynamic mean-variance portfolio selection problem in a *model-free* manner, based on (generative) diffusion models. We propose using data sampled from the real model $\mathbb{P}$ (which is unknown) with limited size to train a generative model $\mathbb{Q}$ (from which we can easily and adequately sample). With adaptive training and sampling methods that are tailor-made for time-series data, we obtain quantification bounds between $\mathbb{P}$ and $\mathbb{Q}$ in terms of the adapted Wasserstein metric $\mathcal{AW}_2$. We then propose a policy gradient algorithm based on the generative environment, in which our innovative adapted sampling method provides approximate scenario generators. We illustrate the performance of our algorithm on real data. The algorithm based on the generative environment produces portfolios that beat several important baselines, including the Markowitz portfolio, the equal weight (naive) portfolio, and S&P 500.

1 Introduction

Portfolio selection, a crucial and fundamental task in financial engineering, has been a subject of both theoretical and practical interest since the classical work Markowitz [1952]. With the ability to account for intertemporal hedging and rebalancing in a non-stationary market, formulating portfolio selection as *dynamic* multi-period problems has become increasingly appealing; see Mossin [1968], Merton [1971] for early works. Due to incompleteness of the information, estimation errors, or simply the infeasibility of fully modeling market uncertainty, we lack access to the real model, as opposed to model-based approaches. Instead, we are restricted to a limited set of financial data samples. Consequently, investigating portfolio selection problems in a *model-free* manner has been a long-standing topic in financial engineering; see discussions of 'universal portfolio' Cover [1991], stochastic portfolio theory Karatzas and Ruf [2017], Cuchiero et al. [2019], and the application of machine learning Ban et al. [2018], Gu et al. [2020]. However, dynamic portfolio selection problems are rather involved, even with model-based approaches, due to their inherent intertemporal structure Campbell and Viceira [1999], high dimensionality Brandt et al. [2005], and time inconsistency Basak and Chabakauri [2010], Björk et al. [2014]. Therefore, model-free dynamic portfolio selection problems are worth exploring.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Generative AI in Finance.

To address this challenge, we propose to leverage score-based diffusion models. These models can generate additional samples that exhibit distributional properties similar to the original data, but theoretical foundations for generating time-series data have not been established yet. We close this gap by proposing a adaptive sampling scheme (Algorithm 1), which outputs an alternative model $\mathbb{Q}$. The difference between $\mathbb{Q}$ and the original model $\mathbb{P}$ is assessed with adapted Wasserstein metric (Theorem 2.1). We then utilize the model $\mathbb{Q}$ as an oracle in the subsequent policy gradient algorithm: we sample a large number of paths $\tilde{s}^{1:T} \sim \mathbb{Q}$ and train a reinforcement learning (RL) agent using these simulated paths. The approximated model $\mathbb{Q}$ consists of a score network s_θ, which is the core of the score-matching technique used in generative models, and a Recurrent Neural Network (RNN) encoder R_θ, which encodes the current state of the market. These two networks are trained simultaneously using historical price data. The reinforcement learning agent is trained on the generated data from $\mathbb{Q}$, which are produced by evaluating s_θ and R_θ in an adaptive manner. Finally, the agent, fed with the encoded (and updated) market state, outputs portfolio selection actions whose performance is evaluated through real-world experiments.

2 Score-based diffusion model for time-series data

For simplicity, we take the OU process as the forward process of our diffusion model:

$$\mathrm{d}X_\tau = -X_\tau \mathrm{d}\tau + \sqrt{2}\mathrm{d}B_\tau. \tag{1}$$

The corresponding time-reversed SDE is

$$\mathrm{d}\bar{X}_\tau = [\bar{X}_\tau + 2\nabla \log p(\mathcal{T} - \tau, \bar{X}_\tau)]\mathrm{d}t + \sqrt{2}\mathrm{d}\bar{B}_\tau,$$

where p is the density function of the forward process (1) with certain initial conditions. For the purpose of conditional score-matching and sampling, we use $p_{t+1}(\tau, \cdot | x^{1:t})$ to denote the probability density function of the forward process, with random initial condition $\mathbb{P}_{x^{1:t}}$. Here, $x^{1:t}$ is the conditional input and can be used in practice as observations.

Suppose the real model of time-series data is $\mathbb{P}$, and for every t we have a pre-trained diffusion model s_θ^t with a score matching error $\varepsilon_{\text{score}}$ to the real (conditional) score function $\nabla_x p_t$ (details of the conditional score matching will be provided in Appendix B). We note that, fixing $x^{1:t}$, $s_\theta^{t+1}(\tau, x^{1:t}, \cdot)$, serves as an approximation of the *conditional score function* $\nabla_x \log p(\tau, \cdot | x^{1:t})$. This motives us to sample from the output distribution adaptively; see Algorithm 1.

Algorithm 1 Adaptive sampling

```
Inputs: pre-trained approximated score functions $s_\theta^t$, $t = 1, 2, \cdots, T$; backward SDE simulators $\{Y_s\}_{0 \le s \le \mathcal{T}}$.
Initialization: samples from the noise: $\{(Z_{(n)}^1, Z_{(n)}^2, \cdots, Z_{(n)}^T)\}_{n=1}^N \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^{dT}}, I_{dT \times dT})$.
for $n = 1$ to $N$ do
    With initial condition $\bar{Y}_0 = Z_{(n)}^1$, run the backward SDE with score matching function $s_1^\theta$, i.e.,
        $\mathrm{d}\bar{Y}_\tau = [\bar{Y}_\tau + 2s_\theta^1(\mathcal{T} - \tau, \bar{Y}_\tau)]\mathrm{d}\tau + \sqrt{2}\bar{B}_\tau.$
    $y_{(n)}^1 \leftarrow \bar{Y}_{\mathcal{T}}$.
    for $t = 1$ to $T - 1$ do
        With initial condition $\bar{Y}_0 = Z_{(n)}^{t+1}$, run the backward SDE with score matching function $(\tau, x) \mapsto s_\theta^{t+1}(\tau, y_{(n)}^{1:t}, x)$. , i.e.,
            $\mathrm{d}\bar{Y}_\tau = [\bar{Y}_\tau + 2s_\theta^{t+1}(\mathcal{T} - \tau, y_{(n)}^{1:t}, \bar{Y}_\tau)]\mathrm{d}\tau + \sqrt{2}\bar{B}_\tau.$
        $y_{(n)}^{t+1} \leftarrow \bar{Y}_{\mathcal{T}}$.
    end for
end for
Output: $\{(y_{(n)}^1, y_{(n)}^2, \cdots, y_{(n)}^T)\}_{n=1}^N$.
```

We now present our main theoretical result: an quantification error bound between the real model $\mathbb{P}$ and the output measure $\mathbb{Q}^{\mathcal{T}}$ of Algorithm 1.

Theorem 2.1. *For any $\varepsilon_{\text{score}} > 0$, there exists a family of score matching functions $\{s_\theta^t\}_{t=1}^T$ with score matching error $\varepsilon_{\text{score}}$, such that Algorithm 1 produces an approximated model $\mathbb{Q}^{\mathcal{T}}$ satisifying*

$$\mathcal{AW}_2^2(\mathbb{P}, \mathbb{Q}^{\mathcal{T}}) \leq C\Big(\mathcal{T}^{\frac{5T}{2}} \varepsilon_{\text{score}}^{1/2^{T-1}} + \mathcal{T}^{\frac{5(T-1)}{2}+\frac{1}{2^{T-2}}} e^{-c\mathcal{T}/2^{T-1}} \varepsilon_{\text{score}}^{-1/2^{T-1}}\Big). \tag{2}$$

Remark 1. Unlike usual Wasserstein bound in static data setting, the error is unbounded when we let $\varepsilon_{\text{score}} \to 0$ for a fixed $\mathcal{T}$. However, by taking $\varepsilon_{\text{score}} \to 0$ and $\mathcal{T} \to \infty$ in an appropriate scale (e.g., $\varepsilon_{\text{score}} = \mathcal{T}^{-52^{T-3}+1} e^{-c\mathcal{T}/2}$), we still obtain small error bounds. This would indeed be the case when we implement finite sample analysis, where $\varepsilon_{\text{score}}$ and $\mathcal{T}$ are in different scales of sample size n. Such an difference arises because we deviate from usual strongly log concavity assumption of data distribution. This assumption seems rather restrictive when imposed on all *conditional* distributions. See techinical appendices for details.

3 Application to portfolio selection

In this section, we apply score-based diffusion model, redesigned for time-series data in this work, to dynamic portfolio selection problem. More specifically, suppose the price data $S^{1:T} \sim \mathbb{P}$[1], we want to solve

$$v^*(\mathbb{P}) := \sup_{\vartheta} \Big\{ \mathbb{E}_{\mathbb{P}}[(\vartheta \cdot S)_T] - \frac{\gamma}{2} \mathrm{Var}_{\mathbb{P}}[(\vartheta \cdot S)_T] \Big\}, \tag{3}$$

where $(\vartheta \cdot S)_T := \sum_{l=1}^{T-1} \vartheta_l^{\mathbf{t}} (S^{l+1} - S^l)$. The idea is to solve (3) under $\mathbb{Q}$ (i.e., $v^*(\mathbb{Q})$) by a policy gradient algorithm, and use the output portfolio as an approximated solution to (3). This is feasible because by results in Section 2 we can sample from $\mathbb{Q}$ easily and the conditional sampling is also allowed.

To implement Algorithm 1, we use two neural networks to represent $\mathbb{Q}$: one score network s_θ and one RNN encoder R_θ. The RNN encoder is used to summarize the information from history by the recursive relation $h^{t+1} = R_\theta(s^t, h^{t-1})$, hence the conditional variables $s^{1:t}$ with varying length can be replaced by h^t on a single space. s_θ and R_θ are trained jointly by minimizing the following score-matching loss function:

$$\frac{1}{TM} \sum_{t,m} \Big| \sqrt{1 - e^{2\tau_{(m)}}} s_\theta \Big(\tau_{(m)}, h_{(m)}^t, s_{(m)}^t e^{-\tau_{(m)}} + \sqrt{1 - e^{-2\tau_{(m)}}} \mathbf{z}_{(m)} \Big) + \mathbf{z}_{(m)} \Big|^2. \tag{4}$$

Here, $\{s_{(m)}^{1:T}\}_{m=1}^M$ is a batch of data, and $\{h_{(m)}^{1:T}\}_{m=1}^M$ are the corresponding hidden variables obtained by feeding the data to the RNN encoder R_θ; $\{\tau_{(m)}\}_{k=1}^M$ are randomly sampled from $[0, \mathcal{T}]$; $\{\mathbf{z}_{(m)}\}_{m=1}^M$ are sampled from pure noise $\mathcal{N}(0, I)$. Although only s_θ explicitly appears in (4), R_θ is implicitly trained by backward propagation. In the sampling phase, we follow Algorithm 1, while replacing $y_{(n)}^{1:t}$ obtained from previous steps by $h_{(n)}^t = R_\theta(s_{(n)}^t, h_{(n)}^{t-1})$. We emphasize that the proposed sampling algorithm support *conditional sampling*: we can collect some contextual data $x^{-1:-T}$ and use it to obtain h^0. Inserting h^0 into s_θ enables sampling $s^{1:T}$ from some conditional distribution instead of a fixed distribution $\mathbb{P}$. This will be crucial for applications to portfolio selection. An implemented version of Algorithm 1, along with the training algorithm, is provided in Appendix E as pseudocode; see Algorithms 2 and 3.

As the next step, we train a reinforcement learning agent with a policy gradient algorithm with the *generative environment* $\mathbb{Q}$. The proposed algorithm is a revised version of the deterministic policy gradient algorithm (DDPG, c.f. Lillicrap et al. [2016]). We redesign part of the algorithm to accommodate the unique characteristics of portfolio selection problems. These adjustments include separating the scenario pool and replay buffer, which facilitates the convenient integration of the generative model. See Algorithm 4 in Appendix E for the pseudo code.

4 Experiments

We implement the proposed workflow to a set of real world data. Our empirical tests employ the value-weighted 10 Industry Portfolios from the Kenneth R. French Data Library, a monthly panel that aggregates all common stocks on the CRSP tape into ten broad industry groups—Non-Durables ('NoDur'),

[1] We change the notation to S as is frequently used in finance

Table 1: Performance of different portfolio strategies with $\gamma = 0.5$

	Return	Volatility	Sharpe	Sortino	Max Drawdown	Calmer
S&P 500	10.98%	0.1460	0.7522	0.6955	-0.2477	0.4433
EW	13.26%	0.1502	0.8831	0.8310	-0.2293	0.5783
HistMarkowitz	6.18%	0.2043	0.3025	0.2973	-0.4090	0.1511
GenMarkowitz	12.02%	0.2418	0.4970	0.5121	-0.5563	0.2160
GenTD3	**13.57%**	**0.1440**	**0.9428**	**0.9774**	**-0.2199**	**0.6405**

Durables ('Durbl'), Manufacturing ('Manuf'), Energy ('Enrgy'), HiTech ('HiTec'), Telecommunications ('Telcm'), Utilities ('Utils'), Shops ('Shops'), Health ('Hlth'), and Others ('Other'). The dataset that we use can be found and downloaded from `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. We provide it in the supplementary files along with the code.

For training, we use a *same-resolution UNet* which is inspired by Ronneberger et al. [2015] as the score network s_θ and an LSTM Hochreiter and Schmidhuber [1997] as the RNN encoder R_θ like Yan et al. [2021]. To improve the performnance, we employ variance-preserving stochastic differential equation (VPSDE) with a time varying noisy scheme $\{\beta(\tau)\}_{\tau\in[0,\mathcal{T}]}$, which generalizes the OU process in theoretical part. Details of training, including choices of hyperparameters, can be found in Appendix D. For the RL gradient, we use the *twin delayed* version of Algorithm 4 (Twin Delayed DDPG, **TD3**; c.f. Fujimoto et al. [2018]) to improve performance and stability. The training (along with validation) dataset consists of indices data from July 1926 through March 2009.

For testing, we set the time horizon $T = 12$ (i.e., one year) for TD3 agent, and use an extended test time horizon $T_{\text{test}} = 15T$ (i.e., 15 years) to show the robustness. That is to say, we use indices data spanning from April 2009 to March 2025 for the purpose of testing. We compare the market benchmark, the S&P 500 index, and four portfolio strategies: Equal Weight (EW), History-based Markotwitz (HistMarkowitz), Generative-model-based Markowitz (GenMarkowitz), and Generative-model-based TD3 (GenTD3). The results are presented in Table 1. Here, we set the risk aversion $\gamma = 0.5$. More experiment results, including those with varying risk aversions, are provided in Appendix D.

5 Limitations and future works

As a main theoretical limitations, the Lipschitiz continuity on the conditional variable $x^{1:t}$ in (10) seems restrictive. One could relax this assumption with a more refined approximation analysis, impose Lipschitz continuity only on x^{t+1}, which aligns with the usual Lipschitz condition for the score function. Moreover, the error bounds established here are unlikely to be optimal; tightening, or even optimizing, the rate remains an interesting direction. *Practically*, further fine-tuning of the proposed algorithms and large-scale experiments are highly motivated. Furthermore, to enhance the predictive ability we can incorporate more covariant variables into the input of RNN encoder, instead of just using the price (return) data. Quantitative analysis of the RL part of this work (e,g, the regret analysis) is also interesting. We leave these directions for future study.

References

Julio Backhoff, Mathias Beiglböck, Yiqing Lin, and Anastasiia Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 27(4):2528–2562, 2017.

Gah-Yi Ban, Noureddine El Karoui, and Andrew E. B. Lim. Machine learning and portfolio optimization. *Management Science*, 64(3):1136–1154, 2018.

Suleyman Basak and Georgy Chabakauri. Dynamic mean–variance asset allocation. *Review of Financial Studies*, 23(8):2970–3016, 2010.

Erhan Bayraktar and Bingyan Han. Fitted value iteration methods for bicausal optimal transport. *to appear in Applied Mathematics and Optimization*, 2025+. URL `https://arxiv.org/abs/2306.12658`. arXiv: 2306.12658.

Tomas Björk, Agatha Murgoci, and Xun Yu Zhou. Mean–variance portfolio optimization with state-dependent risk aversion. *Mathematical Finance*, 24(1):1–24, 2014.

Michael W. Brandt, Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud. A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *The Review of Financial Studies*, 18(3):831–873, 2005.

John Y. Campbell and Luis M. Viceira. Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics*, 114(2):433–495, 1999.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*, 2023.

Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

Christa Cuchiero, Walter Schachermayer, and Ting-Kam Leonard Wong. Cover's universal portfolio, stochastic portfolio theory and the numéraire portfolio. *Mathematical Finance*, 29(3):773–803, 2019.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953, 2009.

Scott Fujimoto, Herke van Hoof, and David Meger. Addressing function approximation error in actor-critic methods. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 1582–1591. PMLR, 2018.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6:695–709, 2005.

Ioannis Karatzas and Johannes Ruf. Trading strategies generated by Lyapunov functions. *Finance and Stochastics*, 21(3):753–787, 2017.

R. Z. Khasminskii. *Stochastic Stability of Differential Equations*, volume 66 of *Stochastic Modelling and Applied Probability*. Springer, Berlin, 2nd edition, 2012.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2016.

Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, 1971.

Jan Mossin. Optimal multiperiod portfolio policies. *Journal of Business*, 41(2):215–229, 1968.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. In *Advances in Neural Information Processing Systems*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b.

Cédric Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.

Tijin Yan, Hongwei Zhang, Tong Zhou, Yufeng Zhan, and Yuanqing Xia. Scoregrad: Multivariate probabilistic time series forecasting with continuous energy-based generative models. 2021. URL `https://arxiv.org/abs/2106.10121`. arXiv: 2106.10121.

A Frequently used notation

In this work, the data is in the form of time-series with length $T \in \mathbb{N}$ and dimension $d \in \mathbb{N}$. We denote by $t \in \{1, 2, \cdots, T\}$ the time index for the time-series. For general theoretical results, we use $X^{1:t} := (X^1, X^2, \cdots, X^t)$ ($Y^{1:t} := (Y^1, Y^2, \cdots, Y^t)$) to denote random variables sampled from data distribution (approximated model, respectively), and $x^{1:t}$ ($y^{1:t}$) the realization. For the diffusion model, we use $\tau \in [0, \mathcal{T}]$ to denote the *diffusive time*, i.e., the time flow of the forward and reversed SDE, with $\mathcal{T} > 0$ the maximal diffusive time.

For a probability $\mathbb{P}$ on the space of time-series data $\mathbb{R}^{dT}$, $\mathbb{P}_{1:t}$ is the joint distribution of $(X^1, X^2, \cdots X^t)$, thus $\mathbb{P}_{1:T} = \mathbb{P}$. Moreover, $\mathbb{P}_{x^{1:t}}$ is the conditional distribution of X^{t+1}, conditioning on $X^{1:t} = x^{1:t}$. By convention, we use $\mathbb{P}$ to denote the original (real) model, hence the distribution of $X^{1:T}$ and $\mathbb{Q}$ the alternative model, hence the distribution of $Y^{1:T}$.

To present the quantification error bound of the score-based diffusion model, we use the adapted Wasserstein metric. More specifically, for two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $\mathbb{R}^{dT}$, a coupling π between $\mathbb{P}$ and $\mathbb{Q}$ is said to be *causal in x* if for any $t \in \{1, 2, \cdots, T\}$,

$$\pi(Y^t \in \mathrm{d}y^t | X^{1:T} = x^{1:T}) = \pi(Y^t \in \mathrm{d}y^t | X^{1:t} = x^{1:t}).$$

A coupling π is said to be *bicausal* if it is causal in both x and y (c.f. Backhoff et al. [2017], Bayraktar and Han [2025+]). Define $\Pi_{\rm bc}(\mathbb{P}, \mathbb{Q})$ to be the set of all bicausal couplings between $\mathbb{P}$ and $\mathbb{Q}$. With the choice of transport cost $|x^{1:T} - y^{1:T}| := (\sum_{t=1}^{T} |x^t - y^t|^2)^{1/2}$, we define the *adapted Wasserstein metric* as follows:

$$\mathcal{AW}_2^2(\mathbb{P}, \mathbb{Q}) = \left(\inf_{\pi \in \Pi_{\rm bc}} \int_{\mathbb{R}^{dT} \times \mathbb{R}^{dT}} |x^{1:T} - y^{1:T}|^2 \pi(\mathrm{d}x^{1:T}, \mathrm{d}y^{1:T}) \right)^{1/2}.$$

We use K, C, c to denote generic constants which may vary from line to line.

B Details of the score matching

In Section 2, we assume that the approximated score functions $\{s_\theta^t\}_{t=1}^T$ have the score matching error $\varepsilon_{\text{score}}$. More specifically, we assume that they satisfy the following assumption:

Assumption 1. For any $\tau \in (0, \mathcal{T}]$ and $t = 1, 2, \cdots, T-1$, we have

$$\mathbb{E}_{X \sim p_1(\tau, \cdot)} |s_\theta^1(\tau, X) - \nabla_x \log p_1(\tau, X)|^2 \leq \varepsilon_{\text{score}}^2, \tag{5}$$

$$\mathbb{E}_{X^{1:t} \sim \mathbb{P}_{1:t}} \mathbb{E}_{X_\tau^{t+1} \sim p_{t+1}(\tau, \cdot | X^{1:t})} \left| s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1}) - \nabla_x \log p_{t+1}(\tau, X_\tau^{t+1} | X^{1:t}) \right|^2 \leq \varepsilon_{\text{score}}^2. \tag{6}$$

Here, $p_{t+1}(\tau, \cdot | x^{1:t})$ is the probability density function of the forward process, with random initial condition $\mathbb{P}_{x^{1:t}}$.

We remark that (5) is the classical score error bound for the first marginal of $\mathbb{P}$; see Song et al. [2021b]. (6) is crucial to our adaptive sampling scheme because it provides a proxy to $\nabla_x \log p_{t+1}(\tau, \cdot | X^{1:t})$, the score function conditioning on the realization $X^{1:t}$. However, (6) is not feasible directly, because we can not sample from the conditional probability, but rather, only have access to the joint distribution (data distribution). This issue can be resolved by the following *denoising score-matching techniques*, in which the error can be computed using only joint distribution. While the idea is borrowed from classical denoising score matching (e.g., Hyvärinen [2005]), we need to consider a conditional version and the proof is more complicated.

Proposition B.1. *For any $t = 1, 2, \cdots, T-1$, the following two score-matching problems are equivalent:*

$$\min_\theta \mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}} \mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})} \left| s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1}) - \nabla_x \log p_{t+1}(\tau, X_\tau^{t+1}|X^{1:t})\right|^2 \tag{7}$$

$$\min_\theta \mathbb{E}_{X^{1:t+1}\sim\mathbb{P}_{1:t+1}} \mathbb{E}_{X_\tau^{t+1}\sim \phi(\tau,\cdot|X^{t+1})} \left| s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1}) - \nabla_x \log \phi(\tau, X_\tau^{t+1}|X^{t+1})\right|^2 . \tag{8}$$

Here, $\phi(\tau, \cdot|x_0)$ is the probability density function of the forward process, with initial condition $X_0 = x_0$. In particular,

$$\phi(\tau, x|x_0) = (2\pi(1-e^{-2\tau}))^{-d/2} e^{-\frac{|x-x_0e^{-\tau}|^2}{2(1-e^{-2\tau})}} .$$

Because the explicit form of $\phi(\tau, x|x_0)$, $X_\tau^{t+1} \sim \phi(\tau, \cdot|X^{t+1})$ is equivalent to $X_\tau^{t+1} = X^{t+1}e^{-\tau} + \sqrt{1-e^{-2\tau}}\mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(0, I)$. Therefore, (8) is further equivalent to

$$\min_\theta \mathbb{E}_{X^{1:t+1}\sim\mathbb{P}_{1:t+1}} \mathbb{E}_{\mathbf{z}\sim\mathcal{N}(0,I)} \left| \sqrt{1-e^{-2\tau}} s_\theta^{t+1}\big(\tau, X^{1:t}, X^{t+1}e^{-\tau} + \sqrt{1-e^{-2\tau}}\mathbf{z}\big) + \mathbf{z}\right|^2 . \tag{9}$$

In the implementation, s_θ is trained so as to minimize the empirical version of (9), with an additional integration in τ to account for different diffusive time; see (4).

C Assumptions of Theorem 2.1

To establish Theorem 2.1, we need two technical assumptions, one on data distribution $\mathbb{P}$ and another on approximating score function s_θ. We present such assumptions in this appendix, and several remarks are in order.

Assumption 2 (Assumptions on the data distribution).

1. There exists a constant $L > 0$ such that for any $t \in \{1, 2, \cdots, T\}$, $\tau \in [0, \mathcal{T}]$ and $x^{1:t}, y^{1:t} \in \mathbb{R}^{dt}$,
$$\begin{aligned} &|\nabla_x \log p_1(\tau, x) - \nabla_x \log p_1(\tau, y)| \le L|x-y|, \\ &|\nabla_x \log p_{t+1}(\tau, x|x^{1:t}) - \nabla_x \log p_{t+1}(\tau, y|y^{1:t})| \le L(|x^{1:t} - y^{1:t}| + |x-y|). \end{aligned} \tag{10}$$
2. For some $c > 0$, $\mathbb{E}_{\mathbb{P}_1} e^{c|X_1|} < \infty$, and for $t = \{1, 2, \cdots, T-1\}$,
$$\sup_{x^{1:t}\in\mathbb{R}^{dt}} \mathbb{E}_{\mathbb{P}_{x^{1:t}}} e^{c|X^{t+1}|} < \infty.$$

Assumption 3 (Assumptions on the approximating network).

1. There exists a constant $L > 0$ such that for any $t \in \{1, 2, \cdots, T\}$, $\tau \in [0, \mathcal{T}]$ and $x^{1:t}, y^{1:t} \in \mathbb{R}^{dt}$,
$$\begin{aligned} &s_\theta^1(\tau, x) - s_\theta^1(\tau, y) \le L|x-y|, \\ &s_\theta^{t+1}(\tau, x^{1:t}, x) - s_\theta^{t+1}(\tau, y^{1:t}, y) \le L(|x^{1:t} - y^{1:t}| + |x-y|). \end{aligned}$$
2. There exist constants $M_{\text{disp}}, \delta > 0$ such that for any for any $t \in \{1, 2, \cdots, T\}$, $\tau \in [0, \mathcal{T}]$, $x^{1:t} \in \mathbb{R}^{dt}$ and $x \in \mathbb{R}^d$,
$$\begin{aligned} &2x \cdot s_\theta^1(\tau, x) \le -(1+\delta)|x|^2 + M_{\text{disp}}, \\ &2x \cdot s_\theta^{t+1}(\tau, x^{1:t}, x) \le -(1+\delta)|x|^2 + M_{\text{disp}}. \end{aligned}$$

Remark 2. Typically, it is advisable to refrain from making assumptions about the approximation network, as such assumptions may potentially compromise its approximating power. However, in this paper, we demonstrate that by adhering to Assumption 2, which are all assumptions pertaining to the data distribution, we can select the approximating network s_θ such that it satisfies Assumption 3 (with appropriately chosen M_{disp}), while maintaining the approximation errors (5) and (6). See Proposition C.1 below.

Proposition C.1. *Suppose Assumption 2 holds. For any $\varepsilon_{\text{score}} > 0$, there exists a s_θ satisfying Assumption 1 as well as Assumption 3 with $M_{\text{disp}} \sim \log(1/\varepsilon_{\text{score}})$.*

Remark 3. Even for static data ($T = 1$) our theoretical results are new from the perspective of assumptions needed. Indeed, when $T = 1$, Assumption 2 is reduced to the usual Lipschitz continuity of score functions, and we do not rely on any structural assumptions on $\mathbb{P}$, such as strongly log concavity. The price to pay is that the noise approximation error term $\mathcal{T}^2 e^{-c\mathcal{T}} \varepsilon_{\text{score}}^{-1}$ explodes when $\varepsilon_{\text{score}} \to 0$. See (2).

D Details and more results of the experiment

As is mentioned in Section 4, in the experiment we use the value-weighted 10 Industry Portfolios from the Kenneth R. French Data Library, spanning July 1926 through March 2025. The detailed split of the dataset is as follows:

1. The test dataset comprises data collected over the past 16 years, from April 2009 to March 2025. The initial year of the test dataset, spanning from April 2009 to March 2008, serves as the historical window and is utilized as the initial condition input into the diffusion model; see variable h in Algorithms 2-4. Consequently, the evaluation period is from March 2010 to March 2025.
2. The validation dataset consists of data from August 1992 to March 2009, encompassing 500 months prior to the test dataset.
3. The remaining dataset is employed in Algorithm 2 to train the diffusion model s_θ and R_θ.

For training, we use a *same-resolution UNet* which is inspired by Ronneberger et al. [2015] as the score network s_θ and an LSTM Hochreiter and Schmidhuber [1997] as the RNN encoder R_θ. To improve the performnance, e employ the refined version of OU process: the variance-preserving stochastic differential equation (VPSDE):

$$\mathrm{d}X_\tau = -\tfrac{1}{2}\,\beta(\tau)\,X_\tau\,\mathrm{d}\tau + \sqrt{\beta(t)}\,\mathrm{d}B_\tau, \qquad \beta(\tau) = \beta_{\min} + (\beta_{\max} - \beta_{\min})\,\tau,$$

discretised into N uniform timesteps $\{t_n\}_{n=1}^{N}$. Default hyperparameters are $\beta_{\min} = 0.01$, $\beta_{\max} = 10.0$, and $N = 1000$. Then, s_θ and R_θ are trained according to Algorithm 2. Detailed model architectures and training configurations are avaible in our github repository. Optimisation uses the AdamW optimiser with learning rate 10^{-3}, $(\beta_1, \beta_2) = (0.9, 0.999)$, and weight decay 10^{-2}. A cosine-annealing scheduler runs for 400 epochs with a 5 epoch warm-up. Gradients are clipped to a maximum norm of 1, and mixed precision training (bfloat16) is enabled via `torch.cuda.amp`. We maintain an exponential moving average (EMA) of model weights with decay 0.999 and evaluate validation metrics on the EMA weights. With these settings, training on a single NVIDIA T-40 (On Google Colab) is completed in about 10 minutes.

For testing, we set the time horizon $T = 12$ (i.e., one year) for TD3 agent, and use an extended test time horizon $T_{\text{test}} = 15T$ (i.e., 15 years) to show the robustness. We compare the market benchmark, the S&P 500 index, and four portfolio strategies with details described below:

1. **Equal Weight (EW)**: $a \equiv (1/d, 1/d, \cdots, 1/d)$;
2. **History-based Markotwitz (HistMarkowitz)**: For each $t = 1, 2, \cdots, T_{\text{test}}$, action a is determined by solving the Markowitz problem with no short selling, no borrowing and fully investment constraints. The mean and covariance are estimated by the sample mean and sample covariance of the most recent 60 months data (as in DeMiguel et al. [2009]). In other words, this is a monthly re-balancing Markowitz strategy in which we use historical data as estimators;
3. **Generative-model-based Markowitz (GenMarkowitz)**: For each $t = 1, 2, \cdots, T_{\text{test}}$, we collect the most recent T months (i.e., one year) data as the context window, and use it as the inital feature h^0 in the diffusion model (see Algorithm 3). Then, we sample 500 samples of *one month predictions* for s^t. Then, we use this prediction to solve constrained Markowitz problem. In other words, this is a monthly re-balancing Markowitz strategy based on generative model;
4. **Generative-model-based TD3 (GenTD3)**: We set the first year returns from the test dataset as the context window $s^{-T:-1}$, and then adaptively sample 500 paths of one-year prediction $s^{1:T}$ based on Algorithm 3. We use these samples as the scenario pool $\mathcal{S}$ to train the TD3 agent. The trained agent is used repeatedly for the first 7.5 years, outputing the portfolio with *updated feature* h^t for any $t = 1, 2, \cdots, T_{\text{Test}}$. Then, at the midway of the testing period, we use the most recent one-year data as a new context window and generate another set of 500 paths. The agent is retrained using these paths and then used for the rest 7.5 years[2]. The training of the TD3 policy network costs about 30min (including retraining at 7.5 years).

[2] We only retrain the model once due to the limited computation resources. Ideally, retraining the RL agent *each year*, which is aligned with the length of test period, leads the agent better utilized the up-to-date market

Table 2: Performance of different portfolio strategies with $\gamma = 3$

	Return	Volatility	Sharpe	Sortino	Max Drawdown	Calmer
S&P 500	10.98%	0.1460	0.7522	0.695519	-0.2477	0.4433
EW	13.26%	0.1502	0.8831	0.8310	**-0.2293**	**0.5783**
HistMarkowitz	8.08%	0.1553	0.5200	0.5226	-0.3014	0.2679
GenMarkowitz	**13.59%**	0.1536	0.8850	0.8526	-0.2690	0.5053
GenTD3	12.80%	**0.1398**	**0.9158**	**0.8966**	-0.2325	0.5507

Table 3: Performance of different portfolio strategies with $\gamma = 5$

	Return	Volatility	Sharpe	Sortino	Max Drawdown	Calmer
S&P 500	10.98%	0.1460	0.7522	0.6955	-0.2477	0.4433
EW	13.26%	0.1502	0.8831	0.8310	-0.2293	0.5783
HistMarkowitz	9.26%	0.1375	0.6737	0.6909	-0.2605	0.3555
GenMarkowitz	**14.17%**	**0.1360**	**1.0418**	**1.0543**	**-0.1838**	**0.7711**
GenTD3	12.73%	0.1392	0.9150	0.8916	-0.2326	0.5474

In Section 4 we present the portfolio performance when $\gamma = 0.5$. In this appendix we also present the same type of results with $\gamma = 3$ and $\gamma = 5$ to showcase the robustness.

We observe from Tables 2 and 3 that, although **GenMarkwotiz** performs exceptional for $\gamma = 5$, it crashes for smaller $\gamma = 0.5$, even underperforming **EW** and S&P 500. **GenTD3**, in contrast, is stable in varying risk aversion and achieves the best results for $\gamma = 0.5$. It also has a performance comparable to **GenMarkowitz** for $\gamma = 5$, better than the benchmarks. This sensitivity analysis indicates the importance of the proposed adaptive sampling scheme. We also conclude that, the two strategies based on diffusion model are suitable for different risk aversions. In particular, choosing the appropriate risk aversion (which is not an easy task, though), **GenMarkowitz** may also have satisfactory performance, justifying the applicability of diffusion models in financial data.

E Algorithms

Algorithm 2 Training the score network and the RNN encoder

```
Initialization: Initialize $s_\theta$ and $R_\theta$; collect the training data.
for epochs do
    Sample a batch of data $\{s^{1:T}_{(m)}\}_{m=1}^{M}$.
    Compute $\{h^{1:T}_{(m)}\}_{m=1}^{M}$ using $R_\theta$.
    Sample $\{\tau_{(m)}\}_{m=1}^{M}$ from Uniform$[0, \mathcal{T}]$.
    Sample $\{\mathbf{z}_{(m)}\}_{m=1}^{M}$ from $\mathcal{N}(0, I)$.
    Update $\theta$ by minimizing (4).
end for
Outputs: Trained score network $s_\theta$ and RNN encoder $R_\theta$.
```

F Proofs

Proof of Proposition B.1. By definition, it is not hard to show that

$$p_{t+1}(\tau, x|x^{1:t}) = \int_{\mathbb{R}^d} \phi(\tau, x|x_0)\mathbb{P}_{x^{1:t}}(\mathrm{d}x_0).$$

model. However, this simple retraining scheme presented here has been sufficient to illustrate the application of our theories.

Algorithm 3 Adaptive sampling: implementation

Inputs: Trained score network s_θ and RNN encoder R_θ; number time discretization (predictor) steps $\mathrm{N}_{\mathrm{pre}}$; number of corrector steps $\mathrm{N}_{\mathrm{cor}}$; predictor steps $\{\tau_k\}_{k=1}^{\mathrm{N}_{\mathrm{pre}}}$; corrector step sizes $\{\mathbf{e}_k\}_{k=1}^{\mathrm{N}_{\mathrm{pre}}}$; initial feature h^0

```
$h \leftarrow h^0$.
Sample $s^{1:T} \sim \mathcal{N}(0_{\mathbb{R}^{dT}}, I_{dT\times dT})$
for $t = 1, \cdots, T$ do
    for $k = 0, 1, \cdots, \mathrm{N}_{\mathrm{pre}} - 1$ do
        $s^t \leftarrow s^t + \Big(s^t + 2s_\theta\big(\mathcal{T} - \tau_k, h, s^t\big)\Big) \cdot (\tau_{k+1} - \tau_k) + \sqrt{2}\varepsilon$, where $\varepsilon \sim \mathcal{N}(0, t_{k+1} - t_k)$.
        # The predictor step.
        for $l = 0, 1, \cdots, \mathrm{N}_{\mathrm{cor}} - 1$ do
            $s^t \leftarrow s^t + \mathbf{e}_k s_\theta\big(\tau_k, h, s^t\big) + \sqrt{2}\varepsilon'$, where $\varepsilon' \sim \mathcal{N}(0, \mathbf{e}_k)$.
            # The corrector step.
        end for
    end for
    $h \leftarrow R_\theta(s^t, h)$.   #Compute $h^t$ from $h^{t-1}$.
end for
```

Outputs: $s^{1:T}$.

The term to be minimized in (7) can thus be calculated by

$$
\begin{aligned}
&\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})}\Big|s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1}) - \nabla_x \log p_{t+1}(\tau, X_\tau^{t+1}|X^{1:t})\Big|^2\\
=&\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})}\big|s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1})\big|^2\\
&+\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})}\big|\nabla_x \log p_{t+1}(\tau, X_\tau^{t+1}|X^{1:t})\big|^2\\
&-2\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})}\big[s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1})\cdot\nabla_x \log p_{t+1}(\tau, X_\tau^{t+1}|X^{1:t})\big]\\
=&\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})}\big|s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1})\big|^2 + C\\
&-2\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\int_{\mathbb{R}^d} s_\theta^{t+1}(\tau, X^{1:t}, x)\cdot\nabla_x p_{t+1}(\tau, x|X^{1:t})\mathrm{d}x\\
=&\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})}\big|s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1})\big|^2 + C\\
&-2\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\int_{\mathbb{R}^d} s_\theta^{t+1}(\tau, X^{1:t}, x)\cdot\nabla_x\int_{\mathbb{R}^d}\phi(\tau, x|x_0)\mathbb{P}_{X^{1:t}}(\mathrm{d}x_0)\mathrm{d}x\\
=&\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})}\big|s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1})\big|^2 + C\\
&-2\mathbb{E}_{X^{1:t}\sim\mathbb{P}_{1:t}}\mathbb{E}_{X^{t+1}\sim\mathbb{P}_{X^{1:t}}}\mathbb{E}_{X_\tau^{t+1}\sim\phi(\tau,\cdot|X^{t+1})}\big[s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1})\cdot\nabla_x\log\phi(\tau, X_\tau^{t+1}|X^{t+1})\big]\\
=&\mathbb{E}_{X^{1:t+1}\sim\mathbb{P}_{1:t+1}}\mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|X^{1:t})}\big|s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1})\big|^2 + C\\
&-2\mathbb{E}_{X^{1:t+1}\sim\mathbb{P}_{1:t+1}}\mathbb{E}_{X_\tau^{t+1}\sim\phi(\tau,\cdot|X^{t+1})}\big[s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1})\cdot\nabla_x\log\phi(\tau, X_\tau^{t+1}|X^{t+1})\big]\\
=&\mathbb{E}_{X^{1:t+1}\sim\mathbb{P}_{1:t+1}}\mathbb{E}_{X_\tau^{t+1}\sim\phi(\tau,\cdot|X^{t+1})}\Big|s_\theta^{t+1}(\tau, X^{1:t}, X_\tau^{t+1}) - \nabla_x\log\phi(\tau, X_\tau^{t+1}|X^{t+1})\Big|^2 + C.
\end{aligned}
$$

Here, C is a generic constant independent from θ. The proof is completed because (7) and (8) attain the same minimizer. □

To obtain the Wasserstein bounds from total variation bounds, we first provide a uniform-in-time moment estimate for SDE under *strong dissipative condition*; see Section 1.2 of Khasminskii [2012]. Suppose that for the function $f : [0, \mathcal{T}] \times \mathbb{R}^d \to \mathbb{R}^d$, there exists $M_{\mathrm{disp}}, \delta > 0$ such that for any t, x,

$$x \cdot f(t, x) \leq -(1 + \delta)|x|^2 + M_{\mathrm{disp}}.$$

Algorithm 4 Policy gradient with generative environment

```
Initialization: Initialize Q-network $Q_\alpha$ and policy network $\pi_\beta$; initialize target networks: $\alpha_{\text{targ}} \leftarrow \alpha$, $\beta_{\text{targ}} \leftarrow \beta$; set up an empty replay buffer $\mathcal{D}$ and an empty scenario pool $\mathcal{S}$, both with size $L$; train the score network $s_\theta$ and RNN encoder $R_\theta$ using Algorithm 2. Collect context window data $s^{-T:-1}$.
Use $s^{-T:-1}$ to compute initial feature $h^0$.
Sample $L$ paths prediction $s^{1:T}$ and compute corresponding feature paths $h^{1:T}$ using Algorithm 3; store them in $\mathcal{S}$
for steps do
    Initialize $w = 1$; sample one path $(s^{1:T}, h^{1:T})$ from $\mathcal{S}$; sample $c$ from Exp(10).
    for $t = 1$ to $T - 1$ do
        if warm-up then
            $a \sim$ Uniform(K)
        else
            $a \leftarrow \mathcal{P}_K(\pi_\beta(t, w, h^t, c) + \varepsilon)$.
        end if
        Store $(w, a, c)$ in $\mathcal{D}$.
        $w \leftarrow w + a \cdot (s^{t+1} - s^t)$.
        if not warm-up then
            Sample a batch $\{(s^{1:t+1}_{(j)}, h^{1:t+1}_{(j)}\}_{j=1}^B$ from $\mathcal{S}$ and $\{(w_{(j)}, a_{(j)}, c_{(j)}\}_{j=1}^B$ from $\mathcal{D}$.
            $w'_{(j)} = w_{(j)} + a_{(j)} \cdot (s^{t+1}_{(j)} - s^t_{(j)})$.
            if $t \leq T - 1$ then
                Update $\alpha$ by minimizing
                    $\frac{1}{B}\sum_j \Big|Q_\alpha(t, w_{(j)}, h^t_{(j)}, a_{(j)}, c_{(j)}) - Q_{\alpha_{\text{targ}}}\Big(t{+}1, w'_{(j)}, h^{t+1}_{(j)}, \pi_{\beta_{\text{targ}}}(t{+}1, w'_{(j)}, h^{t+1}_j, c_{(j)}), c_{(j)}\Big)\Big|^2.$
            else
                Update $\alpha$ by minimizing
                    $\frac{1}{B}\sum_j \Big|Q_\alpha(t, w_{(j)}, h^t_{(j)}, a_{(j)}, c_{(j)}) - |w_{(j)} - c_{(j)}|^2\Big|^2.$
            end if
            Update $\beta$ by gradient ascent via
                $\frac{1}{B}\sum_j \nabla_a Q_\alpha\Big(t, w_{(j)}, h^t_{(j)}, \pi_\beta(t, w_{(j)}, h^t_{(j)}, c_{(j)}), c_{(j)}\Big)\nabla_\beta \pi_\beta(t, w_{(j)}, h^t_{(j)}, c_{(j)})$
            Update target network:
                $\alpha_{\text{targ}} \leftarrow \rho\alpha_{\text{targ}} + (1 - \rho)\alpha,$
                $\beta_{\text{targ}} \leftarrow \rho\beta_{\text{targ}} + (1 - \rho)\beta.$
        end if
    end for
end for
Find the optimal multiplier by maximizing
    $-\frac{\gamma}{2} Q_\alpha\big(1, 0, h^1, \pi_\beta(1, 0, h^1, c), c\big) + c.$
Output: Policy network $\pi_\beta$
```

We consider the following SDE:

$$\begin{aligned} \mathrm{d}X_t &= (X_t + f(t, X_t))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \\ X_0 &\sim N(0, I). \end{aligned}$$

Lemma F.1. *There exist constants $c_1, c_2, c_3 > 0$, which are independent from t and M_{disp}, such that*

$$\sup_{t\geq 0} \mathbb{E}[e^{c_1|X_t|^2}] \leq c_2 e^{c_3 M_{\text{disp}}}. \tag{11}$$

Proof. Let $V(x) = e^{\theta|x|^2}$, where $\theta > 0$ will be determined later. Applying Itô's formula to $V(X_s)$, we have

$$\begin{aligned} V(X_t) - V(X_0) = \int_0^t \Big(2\theta\Big(V(X_s)X_s \cdot \big(X_s + f(s, X_s)\big)\Big) + 2\theta d V(X_s) + 4\theta^2|X_s|^2 V(X_s)\Big) ds \\ + 2\sqrt{2}\theta \int_0^t X_s V(X_s) dB_s. \end{aligned} \tag{12}$$

Considering a sequence of localization stopping time $\tau_n \to \infty$ and taking expectation on the above identity (12), we obtain

$$\begin{aligned} \mathbb{E}[V(X_{t\wedge\tau_n})] - \mathbb{E}[V(X_0)] &= \int_0^{t\wedge\tau_n} \Big(2\theta\Big(\mathbb{E}\big[V(X_s)X_s \cdot \big(X_s + f(s, X_s)\big)\big]\Big) \\ &\quad + 2\theta d\mathbb{E}[V(X_s)] + 4\theta^2\mathbb{E}[|X_s|^2 V(X_s)]\Big) ds \\ &= \int_0^t \Big(2\theta\Big(\mathbb{E}\big[V(X_{s\wedge\tau_n})X_{s\wedge\tau_n} \cdot \big(X_{s\wedge\tau_n} + f(s\wedge\tau_n, X_{s\wedge\tau_n})\big)\big]\Big) \\ &\quad + 2\theta d\mathbb{E}[V(X_{s\wedge\tau_n})] + 4\theta^2\mathbb{E}[|X_{s\wedge\tau_n}|^2 V(X_{s\wedge\tau_n})]\Big) ds \end{aligned}$$

The differential form yields

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[V(X_{t\wedge\tau_n})] =& 2\theta\mathbb{E}\big[V(X_{s\wedge\tau_n})X_{s\wedge\tau_n} \cdot \big(X_{s\wedge\tau_n} + f(s\wedge\tau_n, X_{s\wedge\tau_n})\big)\big]\Big) \\ &+ 2\theta d\mathbb{E}[V(X_{t\wedge\tau_n})] + 4\theta^2\mathbb{E}[|X_{t\wedge\tau_n}|^2 V(X_{t\wedge\tau_n})] \\ \leq& (4\theta^2 - 2\theta\delta)\mathbb{E}[|X_{t\wedge\tau_n}|^2 V(X_{t\wedge\tau_n})] + 2\theta(M_{\text{disp}} + d)\mathbb{E}[V(X_{t\wedge\tau_n})]. \end{aligned}$$

Choose $\theta < \delta/2$, $R^2 > 2(M_{\text{disp}} + d)/(\delta - 2\theta)$ and denote $I_t = \big((4\theta^2 - 2\theta\delta)|X_{t\wedge\tau_n}|^2 + 2\theta(M_{\text{disp}} + d)\big)V(X_{t\wedge\tau_n})$. On the one hand, with $\gamma := 2\theta(M_{\text{disp}} + d)$, it is clear that $I_t \leq -\gamma V(X_{t\wedge\tau_n})$ when $|X_{t\wedge\tau_n}| \geq R$. On the other hand, when $|X_{t\wedge\tau_n}| \leq R$,

$$\begin{aligned} I_t \leq& 2\theta(M_{\text{disp}} + d)V(X_{t\wedge\tau_n}) \\ \leq& 2\theta(M_{\text{disp}} + d)V(R) \\ =& 2\theta(M_{\text{disp}} + d)e^{\frac{2\theta(M_{\text{disp}}+d)}{\delta - 2\theta}} \\ \leq& c_1' e^{c_2' M_{\text{disp}}}, \end{aligned}$$

where c_1', c_2' are some constants independent of t and M_{disp}. We thus obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[V(X_{t\wedge\tau_n})\big] \leq& -\gamma\mathbb{E}[V(X_{t\wedge\tau_n})I_{|X_{t\wedge\tau_n}|>R}] + c_1' e^{c_2' M_{\text{disp}}} \\ =& -\gamma\mathbb{E}[V(X_{t\wedge\tau_n})] + \gamma\mathbb{E}[V(X_{t\wedge\tau_n})I_{|X_{t\wedge\tau_n}|\leq R}] + c_1' e^{c_2' M_{\text{disp}}} \\ \leq& -\gamma\mathbb{E}[V(X_{t\wedge\tau_n})] + c_3' e^{c_4' M_{\text{disp}}} \end{aligned}$$

By Grownwall's inequality,

$$\mathbb{E}\big[V(X_{t\wedge\tau_n})\big] \leq e^{-\gamma t}\mathbb{E}[V(X_0)I_{|X_0|>R}] + \frac{c_3' e^{c_4' M_{\text{disp}}}}{\gamma} \leq c_5' e^{c_6' M_{\text{disp}}},$$

Here, constants $c_1' \sim c_6'$ are all independent of t, n and M_{disp}. Thus, letting $n \to \infty$, (11) is established with $c_1 = \theta$, $c_2 = c_5'$, and $c_3 = c_6'$, all independent of t and M_{disp}. □

Corollary F.2. *For any $t, R > 0$, there exists $c_1, c_2 > 0$, which are independent of t and R, such that*

$$\mathbb{E}[|X_t|^2 I_{\{|X_t|\geq R\}}] \leq e^{-c_1 R^2 + c_2 M_{\text{disp}}}.$$

Proof of Theorem 2.1. We first prove the following bound between conditional distributions: There exists a constant $c > 0$ such that, for $t = 1, 2, \cdots, T-1$ we have

$$\mathcal{W}_2^2\big(\mathbb{P}_{x^{1:t}}, \mathbb{Q}^{\mathcal{T}}_{y^{1:t}}\big) \leq C\Big(\alpha(\mathcal{T}) + \mathcal{T}^2 \mathcal{E}(x^{1:t})^{1/2} + \mathcal{T}^{5/2}|x^{1:t} - y^{1:t}|\Big), \tag{13}$$

where

$$\mathcal{E}(x^{1:t}) := \int_0^{\mathcal{T}} \mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|x^{1:t})}\big|s_\theta^{t+1}(\tau, x^{1:t}, X_\tau^{t+1}) - \nabla_x \log p_{t+1}(\tau, X_\tau^{t+1}|x^{1:t})\big|^2 \mathrm{d}\tau,$$
$$\alpha(\mathcal{T}) := \mathcal{T}^2 e^{-\mathcal{T}} + e^{-c\mathcal{T} + cM_{\text{disp}}}.$$

To this end, consider the following two SDEs with initial conditions $\bar{X}_0 \sim p(\mathcal{T}, \cdot|x^{1:t})$ and $\bar{Y}_0 \sim \mathcal{N}(0, I)$, :

$$\mathrm{d}\bar{X}_\tau = [\bar{X}_\tau + 2\nabla_x \log p_{t+1}(\mathcal{T} - \tau, \bar{X}_\tau | x^{1:t})]\mathrm{d}\tau + \sqrt{2}\mathrm{d}B_\tau,$$
$$\mathrm{d}\bar{Y}_\tau = [\bar{Y}_\tau + 2s_\theta^{t+1}(\mathcal{T} - \tau, y^{1:t}, \bar{Y}_\tau)]\mathrm{d}\tau + \sqrt{2}\mathrm{d}B_\tau.$$

By the proof of total variation bounds for continuous time DDPM (see, e.g., Song et al. [2021a] and Chen et al. [2023]), we have

$$\begin{aligned}\mathrm{TV}(\mathbb{P}_{x^{1:t}}, \mathbb{Q}^{\mathcal{T}}_{y^{1:t}})^2 \leq & C\bigg(e^{-2\mathcal{T}} + \int_0^{\mathcal{T}} \mathbb{E}_{X_\tau^{t+1}\sim p_{t+1}(\tau,\cdot|x^{1:t})}\big|s_\theta^{t+1}(\tau, y^{1:t}, X_\tau^{t+1}) - \nabla_x \log p_{t+1}(\tau, X_\tau^{t+1}|x^{1:t})\big|^2 \mathrm{d}\tau\bigg) \\ \leq & C\big(e^{-2\mathcal{T}} + \mathcal{E}(x^{1:t}) + \mathcal{T}|x^{1:t} - y^{1:t}|^2\big)\end{aligned}$$

Therefore, recalling that $\sqrt{a+b+c} \leq \sqrt{a} + \sqrt{b} + \sqrt{c}$ for $a, b, c \geq 0$, we have

$$\mathrm{TV}(\mathbb{P}_{x^{1:t}}, \mathbb{Q}^{\mathcal{T}}_{y^{1:t}}) \leq C\big(e^{-\mathcal{T}} + \mathcal{E}(x^{1:t})^{1/2} + \mathcal{T}^{1/2}|x^{1:t} - y^{1:t}|\big). \tag{14}$$

To bound $\mathcal{W}(\mathbb{P}_{x^{1:t}}, \mathbb{Q}_{y^{1:t}})$ with their total variation metric, we use the following classical results regarding these two different metrics (see Villani [2008], Theorem 6.15): for any two probability measures μ, ν and any $R > 0$,

$$\mathcal{W}_2^2(\mu, \nu) \leq R^2 \mathrm{TV}(\mu, \nu) + \int_{|x|\geq R} |x|^2 (\mu + \nu)(\mathrm{d}x).$$

Taking $\mu = \mathbb{P}_{x^{1:t}}$, $\nu = \mathbb{Q}_{y^{1:t}}$, we have

$$\mathcal{W}_2^2(\mathbb{P}_{x^{1:t}}, \mathbb{Q}^{\mathcal{T}}_{y^{1:t}}) \leq C\bigg(R^2 \mathrm{TV}(\mathbb{P}_{x^{1:t}}, \mathbb{Q}^{\mathcal{T}}_{y^{1:t}}) + \mathbb{E}_{\mathbb{P}_{x^{1:t}}}\big[|X^{t+1}|^2 I_{\{|X^{t+1}|\geq R\}}\big] + \mathbb{E}_{\mathbb{Q}^{\mathcal{T}}_{y^{1:t}}}\big[|Y^{t+1}|^2 I_{\{|Y^{t+1}|\geq R\}}\big]\bigg). \tag{15}$$

We now estimate three terms on the right-hand side of (15). By Assumption 2,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_{x^{1:t}}}\big[|X^{t+1}|^2 I_{\{|X^{t+1}|\geq R\}}\big] \leq & C\mathbb{E}_{\mathbb{P}_{x^{1:t}}}\big[e^{\frac{c}{2}|X^{t+1}|} I_{\{|X^{t+1}|\geq R\}}\big] \\ \leq & C\Big\{\mathbb{E}_{\mathbb{P}_{x^{1:t}}}\big[e^{c|X^{t+1}|}\big]\Big\}^{1/2}\Big\{\mathbb{P}_{x^{1:t}}\big(|X^{t+1}| \geq R\big)\Big\}^{1/2} \\ \leq & Ce^{-\frac{c}{2}R}\mathbb{E}_{\mathbb{P}_{x^{1:t}}}\big[e^{c|X^{t+1}|}\big]. \\ \leq & Ce^{-\frac{c}{2}R}\end{aligned} \tag{16}$$

By Corollary F.2,

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}_{y^{1:t}}}\big[|Y^{t+1}|^2 I_{\{|Y^{t+1}|\geq R\}}\big] = & \mathbb{E}_{\mathbb{Q}^{\mathcal{T}}}\big[|\bar{Y}_{\mathcal{T}}|^2 I_{\{|\bar{Y}_{\mathcal{T}}|\geq R\}}\big] \\ \leq & e^{-c_1 R^2 + c_2 M_{\text{disp}}}.\end{aligned} \tag{17}$$

Plugging (14), (16) and (17) into (15), we have

$$\begin{aligned}\mathcal{W}_2(\mathbb{P}_{x^{1:t}}, \mathbb{Q}^{\mathcal{T}}_{y^{1:t}})^2 \leq & C(R^2 e^{-\mathcal{T}} + R^2\mathcal{E}(x^{1:t})^{1/2} + R^2\mathcal{T}^{1/2}|x^{1:t} - y^{1:t}| + Ce^{-\frac{c}{2}R} + e^{-c_1R^2 + c_2M_{\text{disp}}}) \\ \leq & C(R^2 e^{-\mathcal{T}} + R^2\mathcal{E}(x^{1:t})^{1/2} + R^2\mathcal{T}^{1/2}|x^{1:t} - y^{1:t}| + e^{-cR + cM_{\text{disp}}}).\end{aligned}$$

Taking $R = \mathcal{T}$, we conclude (13).

Next, we prove the main result (2). Because $\pi \in \Pi_{\text{bc}}(\mathbb{P}, \mathbb{Q}^{\mathcal{T}})$ if and only if $\pi_{x^{1:t}, y^{1:t}} \in \Pi(\mathbb{P}_{x^{1:t}}, \mathbb{Q}^{\mathcal{T}}_{y^{1:t}})$ (Backhoff et al. [2017], Proposition 5.1), we construct a specific coupling $\pi^\varepsilon \in \Pi_{\text{bc}}(\mathbb{P}, \mathbb{Q})$ as follows:

(a) $\int_{\mathbb{R}^d \times \mathbb{R}^d} |x^1 - y^1|^2 \pi^\varepsilon_1(\mathrm{d}x^1, \mathrm{d}y^1) \leq \mathcal{W}_2^2(\mathbb{P}_1, \mathbb{Q}^{\mathcal{T}}_1) + \varepsilon$;

(b) For $t = 1, 2, \cdots, T-1$, $\int_{\mathbb{R}^d \times \mathbb{R}^d} |x^{t+1} - y^{t+1}|^2 \pi^\varepsilon_{x^{1:t}, y^{1:t}}(\mathrm{d}x^{t+1}, \mathrm{d}y^{t+1}) \leq \mathcal{W}_2^2(\mathbb{P}_{x^{1:t}}, \mathbb{Q}_{y^{1:t}}) + \varepsilon$;

(c) $\pi^\varepsilon(\mathrm{d}x^{1:T}, \mathrm{d}y^{1:T}) := \pi^\varepsilon_1(\mathrm{d}x^1, \mathrm{d}y^1)\pi^\varepsilon_{x^1, y^1}(\mathrm{d}x^2, \mathrm{d}y^2) \cdots \pi^\varepsilon_{x^{1:T-1}, y^{1:T-1}}(\mathrm{d}x^T, \mathrm{d}y^T)$.

We first use induction to show that, for $t = 1, 2, \cdots, T$,

$$\mathbb{E}_{\pi^\varepsilon_{1:t}}|X^{1:t} - Y^{1:t}|^2 \leq C\Big(\mathcal{T}^{\frac{5t}{2}}\varepsilon_{\text{score}}^{1/2^{t-1}} + \mathcal{T}^{\frac{5(t-1)}{2}}(\alpha(\mathcal{T}) + \varepsilon)^{1/2^{t-1}}\Big). \tag{18}$$

By repeating the proof of (1), with $\mathbb{P}_{x^{1:t}}$, $\mathbb{Q}_{y^{1:t}}$ replaced by $\mathbb{P}_1$, $\mathbb{Q}_1$, we can prove that (18) holds for $t = 1$. This is because when considering non-conditional probability $\mathbb{P}_1$ and $\mathbb{Q}_1$, the term $|x^{1:t} - y^{1:t}|$ disappears and the term $\mathcal{E}(x^{1:t})$ is upper bounded by $\mathcal{T}\varepsilon^2_{\text{score}}$.

Suppose now that (18) is true for t, we aim to prove it for $t+1$. Indeed, by using (13), we have

$$\begin{aligned}\mathbb{E}_{\pi^\varepsilon_{1:t+1}}[|X^{1:t+1} - Y^{1:t+1}|^2] \leq & C\Bigg(\mathbb{E}_{\pi^\varepsilon_{1:t}}[|X^{1:t} - Y^{1:t}|^2] + \mathbb{E}_{\pi^\varepsilon_{1:t}}\mathbb{E}_{\pi^\varepsilon_{X^{1:t}, Y^{1:t}}}[|X^{t+1} - Y^{t+1}|^2]\Bigg) \\ \leq & C\Bigg(\mathbb{E}_{\pi^\varepsilon_{1:t}}[|X^{1:t} - Y^{1:t}|^2] + \mathbb{E}_{\pi^\varepsilon_{1:t}}\big[\mathcal{W}_2^2\big(\mathbb{P}_{X^{1:t}}, \mathbb{Q}^{\mathcal{T}}_{Y^{1:t}}\big)\big] + \varepsilon\Bigg) \\ \leq & C\Bigg(\mathbb{E}_{\pi^\varepsilon_{1:t}}|X^{1:t} - Y^{1:t}|^2 + \mathcal{T}^2\mathbb{E}_{\pi^\varepsilon_{1:t}}\mathcal{E}(X^{1:t})^{1/2} + \mathcal{T}^{\frac{5}{2}}\big\{\mathbb{E}_{\pi^\varepsilon_{1:t}}|X^{1:t} - Y^{1:t}|^2\big\}^{1/2} \\ & + \alpha(\mathcal{T}) + \varepsilon\Bigg)\end{aligned} \tag{19}$$

With (6), we have

$$\mathbb{E}_{\pi^\varepsilon_{1:t}}[\mathcal{E}(X^{1:t})^{1/2}] = \mathbb{E}_{\mathbb{P}_{1:t}}[\mathcal{E}(X^{1:t})^{1/2}] \leq \big\{\mathbb{E}_{\mathbb{P}_{1:t}}[\mathcal{E}(X^{1:t})]\big\}^{1/2} \leq \mathcal{T}^{1/2}\varepsilon_{\text{score}}.$$

If $\mathbb{E}_{\pi^\varepsilon_{1:t}}|X^{1:t} - Y^{1:t}|^2 > 1$, then from (19) we know

$$\begin{aligned}\mathbb{E}_{\pi^\varepsilon_{1:t+1}}|X^{1:t+1} - Y^{1:t+1}|^2 \leq & C\Big(\mathcal{T}^{\frac{5}{2}}\mathbb{E}_{\pi^\varepsilon_{1:t}}|X^{1:t} - Y^{1:t}|^2 + \mathcal{T}^{\frac{5}{2}}\varepsilon_{\text{score}} + \alpha(\mathcal{T}) + \varepsilon\Big) \\ \leq & C\Big(\mathcal{T}^{\frac{5(t+1)}{2}}\varepsilon_{\text{score}}^{1/2^{t-1}} + \mathcal{T}^{\frac{5t}{2}}(\alpha(\mathcal{T}) + \varepsilon)^{1/2^{t-1}} \\ & + \mathcal{T}^{\frac{5}{2}}\varepsilon_{\text{score}} + \alpha(\mathcal{T}) + \varepsilon\Big) \\ \leq & C\Big(\mathcal{T}^{\frac{5(t+1)}{2}}\varepsilon_{\text{score}}^{1/2^t} + \mathcal{T}^{\frac{5t}{2}}(\alpha(\mathcal{T}) + \varepsilon)^{1/2^t}\Big).\end{aligned}$$

Here, we use $\mathcal{T} > 1$, $\varepsilon_{\text{score}} < 1$ and choose a small ε such that $\alpha(\mathcal{T}) + \varepsilon < 1$. Similarly, if $\mathbb{E}_{\pi^{\varepsilon}_{1:t}}|X^{1:t} - Y^{1:t}|^2 \leq 1$,

$$\begin{aligned}\mathbb{E}_{\pi^{\varepsilon}_{1:t+1}}|X^{1:t+1} - Y^{1:t+1}|^2 \leq & C\Big(\mathcal{T}^{\frac{5}{2}}\big\{\mathbb{E}_{\pi^{\varepsilon}_{1:t}}|X^{1:t} - Y^{1:t}|^2\big\}^{1/2} + \mathcal{T}^{\frac{5}{2}}\varepsilon_{\text{score}} + \alpha(\mathcal{T}) + \varepsilon\Big) \\ \leq & C\Big(\mathcal{T}^{\frac{5}{2}\left(\frac{t}{2}+1\right)}\varepsilon_{\text{score}}^{1/2^t} + \mathcal{T}^{\frac{5}{2}\left(\frac{t-1}{2}+1\right)}(\alpha(\mathcal{T}) + \varepsilon)^{1/2^t} \\ & + \mathcal{T}^{\frac{5}{2}}\varepsilon_{\text{score}} + \alpha(\mathcal{T}) + \varepsilon\Big) \\ \leq & C\Big(\mathcal{T}^{\frac{5(t+1)}{2}}\varepsilon_{\text{score}}^{1/2^t} + \mathcal{T}^{\frac{5t}{2}}(\alpha(\mathcal{T}) + \varepsilon)^{1/2^t}\Big).\end{aligned}$$

In either case, we establish that (18) holds for $t+1$, hence for all $t \in \{1, 2, \cdots, T\}$. Taking $t = T$ in (18) and using the definition of adapted Wasserstein metric, we obtain

$$\begin{aligned}\mathcal{AW}_2^2(\mathbb{P}, \mathbb{Q}^{\mathcal{T}}) \leq & \mathbb{E}_{\pi^{\varepsilon}}|X^{1:T} - Y^{1:T}|^2 \\ \leq & C\Big(\mathcal{T}^{\frac{5T}{2}}\varepsilon_{\text{score}}^{1/2^{T-1}} + \mathcal{T}^{\frac{5(T-1)}{2}}(\alpha(\mathcal{T}) + \varepsilon)^{1/2^{T-1}}\Big). \end{aligned} \tag{20}$$

Letting $\varepsilon \to 0$ in (20) gives

$$\begin{aligned}\mathcal{AW}_2^2(\mathbb{P}, \mathbb{Q}^{\mathcal{T}}) \leq & \mathbb{E}_{\pi^{\varepsilon}}|X^{1:T} - Y^{1:T}|^2 \\ \leq & C\Big(\mathcal{T}^{\frac{5T}{2}}\varepsilon_{\text{score}}^{1/2^{T-1}} + \mathcal{T}^{\frac{5(T-1)}{2}}(\mathcal{T}^2 e^{-\mathcal{T}} + e^{-c\mathcal{T} + cM_{\text{disp}}})^{1/2^{T-1}}\Big).\end{aligned}$$

Using Proposition C.1 we take $M_{\text{disp}} \sim \log(1/\varepsilon_{\text{score}})$ and obtain (2). □

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Claims match delivered contributions: adaptive diffusion with conditional sampling, an adapted-Wasserstein error bound, and a generative-environment RL portfolio method showing empirical gains on French 10 Industry; assumptions/limits are stated.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The Limitations and future works section explicitly notes strong assumptions (conditional Lipschitz), potentially suboptimal rates, dataset/scope constraints, and practical/compute needs (larger-scale experiments, RL regret analysis).

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Assumptions are clearly stated and numbered (Assumptions 2, 3), and results are cross-referenced (Theorem 2.1, Propositions). Proofs are included in Appendix F.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper details datasets/splits, baselines, training procedures, and pseudocode, and states that full implementation specifics (architectures) are available in the authors' GitHub repository, enabling reproduction.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The anonymized code submission includes step-by-step instructions that regenerate every table and figure using the public Kenneth R. French dataset. Both the code and the data are included in a .zip file within the supplemental materials of submission.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies the dataset and splits (train/val/test windows), model choices (UNet + LSTM), optimizer and VPSDE hyperparameters, RL setup (TD3 variant, scenario pool size, horizon/retraining), baselines and constraints (60-month window, no shorting/borrowing, fully invested), risk-aversion values, metrics, and monthly rebalancing—with further details in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

Justification: Experiments of this research does not involve statistical significance test. We justify the results of experiments by directly evaluate the output portfolios via classical metrics such as Sharpe, Sortino, or Calmer. We also showcase the robustness by choosing a relative long time-horizon (15 years).

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We state that experiments ran on Google Colab with an NVIDIA T4, and we report per-stage runtimes (e.g., diffusion training 10 min, TD3 training 33 min (including retraining) and total compute in the submission/supplement, which is sufficient to reproduce resource needs.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes] .

   Justification: The research does not involve human subjects or participants, uses open-sourced data, and does not produce any harmful social consequences.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: The submission does not include a Broader Impacts/ethics discussion and does not address positive impacts or potential harms.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The work uses a public financial dataset (Kenneth French 10 Industry Portfolios) and does not release general-purpose pretrained generative models or scraped datasets. No sensitive or personal data are involved, so high-risk misuse and corresponding safeguards are not applicable.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes] .

Justification: The dataset we use is open-sourced and the URL to the data is provided in the paper. We also acknowledge other open-sourced code that we referred to in the README file of supplementary materials.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The submission introduces a codebase (and optional checkpoints) and includes structured documentation: a README with install/run steps and configs and a brief model card outlining training data, evaluation, limitations, and intended use. Assets are anonymized for review and include data-prep scripts and notes on consent (not applicable for the public dataset).

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The work uses public financial time-series data and does not involve crowdsourcing or human-subject experiments, so these items are not applicable.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The work uses public financial time-series data and does not involve human participants or crowdsourcing; IRB approval is not applicable.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA] .

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.