Sample Average Approximation for Conditional Stochastic Optimization with Dependent Data

Yafei Wang [1] **Bo Pan** [1] **Mei Li** [2] **Jianya Lu** [3] **Lingcheng Kong** [4] **Bei Jiang** [1] **Linglong Kong** [1]

Abstract

Conditional Stochastic Optimization (CSO) is a powerful modelling paradigm for optimization under uncertainty. The existing literature on CSO is mainly based on the independence assumption of data, which shows that the solution of CSO is asymptotically consistent and enjoys a finite sample guarantee. The independence assumption, however, does not typically hold in many important applications with dependence patterns, such as time series analysis, operational control, and reinforcement learning. In this paper, we aim to fill this gap and consider a Sample Average Approximation (SAA) for CSO with dependent data. Leveraging covariance inequalities and independent block sampling technique, we provide theoretical guarantees of SAA for CSO with dependent data. In particular, we show that SAA for CSO retains asymptotic consistency and a finite sample guarantee under mild conditions. In addition, we establish the sample complexity $O(d/\varepsilon^4)$ of SAA for CSO, which is shown to be of the same order as independent cases. Through experiments on several applications, we verify the theoretical results and demonstrate that dependence does not degrade the performance of the SAA approach in real data applications.

1. Introduction

In this paper, we study a specific class of stochastic optimization problems, called Conditional Stochastic Optimization (CSO) (Hu et al., 2020a;b; Wu et al., 2023), in the form of,

$$\begin{aligned} F^\star &:= \min_{x\in\mathbb{X}} \mathbb{E}_\xi \left[f_\xi \left(\mathbb{E}_{\eta|\xi} \left[g_\eta(x,\xi) \right] \right) \right], \\ x^\star &:= \arg\min_{x\in\mathbb{X}} \mathbb{E}_\xi \left[f_\xi \left(\mathbb{E}_{\eta|\xi} \left[g_\eta(x,\xi) \right] \right) \right], \end{aligned} \tag{1}$$

where $\mathbb{X} \in \mathbb{R}^d$ is the set of decision variables, $f_\xi(\cdot) : \mathbb{R}^m \to \mathbb{R}$ is a loss function depending on the random vector ξ and $g_\eta(\cdot,\xi) : \mathbb{R}^d \to \mathbb{R}^m$ is a vector-valued loss function depending on both random vectors $\xi \in \Xi$ and $\eta \in H$. Throughout, we only assume that we have samples from the distribution $\mathbb{P}(\xi)$ and the conditional distribution $\mathbb{P}(\eta \mid \xi)$. The goal of the problem (1) is to find an optimal solution of decision $x \in \mathbb{X}$ that minimizes a given composition expected loss, where the inner expectation with respect to (w.r.t.) to an underlying population distribution η given ξ, $\mathbb{P}(\eta \mid \xi)$, and the outer expectation is w.r.t. the distribution $\mathbb{P}(\xi)$. Here we define $F^\star$ and $x^\star$ as the optimal value and solution for the problem (1), respectively.

CSO is a powerful modelling paradigm that is widely used in a variety of applications, such as meta-learning in deep networks (Finn et al., 2017), optimal control in reinforcement learning (Dai et al., 2017; Hambly et al., 2021), policy estimation in linear quadratic adaptive control (Wang & Janson, 2021), instrumental variable regression in causal inference (Hu et al., 2020b; Goda & Kitade, 2023). It can be regarded as an intermediate class of optimization between classical Stochastic Optimization (SO) (Fouskakis & Draper, 2002) and Multistage Stochastic Optimization (MSO) (Zhang & Xiao, 2021). Specifically, CSO is much more general than SO and includes the classical SO as a special case when $g_\eta(\cdot)$ is an identity function. In addition, it can model and account for dynamic randomness and involves conditional construction, while SO works with the cumulative distribution functions. CSO is also less complicated than MSO which aims at a static choice, especially when it comes to calculation and uncertainty quantification.

A variety of optimization methods in solving problem (1) have been studied in the existing literature. However, the construction of optimal solutions in large-scale, data-driven settings remains a major challenge. In particular, since the true distributions $\mathbb{P}(\xi)$ and $\mathbb{P}(\eta \mid \xi)$ in problem (1) are unknown in most cases of practical interest, it could be

[1]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada [2]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China [3]School of Mathematics, Statistics and Actuarial Science, University of Essex, Colchester, UK [4]Department of Mathematics and Statistics, Beijing Jiaotong University, Beijing, China. Correspondence to: Linglong Kong <lkong@ualberta.ca>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

extremely difficult to evaluate the original problem (1) accurately: we lack essential information to solve problems directly. Even if the true distributions are known, the learning procedure could be computationally excruciating, since evaluating the corresponding expectation for a fixed x involves computing a multivariate integral, which could be high-dimensional and intractable. Despite the possibility that the given loss functions are affine functions, the evaluation of the integral is still NP-hard. To address these problems, in this paper, we consider a Sample Average Approximation (SAA)-based modelling paradigm (Wang et al., 2022) for CSO, which embeds problem (1) into a surrogate problem by replacing the unknown (true) probability measure with an approximation. That is, we reformulate problem (1) into a surrogate problem constructed entirely from training samples that can be solved efficiently to find estimators of $J^\star$ and $x^\star$.

Most of the existing work of SAA for CSO focuses mainly on the learning algorithms for optimal solution with convergence guarantees (Hu et al., 2020b; 2021; Goda & Kitade, 2023), while some works explore theoretical properties (Hu et al., 2020a), including asymptotics, sample complexity, etc. Despite these favourable theoretical guarantees and utility in applications, the established results depend critically on the assumption that random samples are independent and identically distributed (IID) (Ermoliev & Norkin, 2013; Kim et al., 2015; Bertsimas et al., 2018). This assumption, however, may be difficult to justify in practice or may be completely invalid. In fact, in most of the aforementioned applications of CSO problems, the training samples of ξ from $\mathbb{P}(\xi)$ and/or η from the conditional distribution of $\mathbb{P}(\eta \mid \xi)$ for each ξ are naturally dependent. For applications where the trajectories of random variables have special dynamic structures, such as in optimal system control and reinforcement learning, the IID assumption of data fails: the observed state-action trajectories are naturally dependent and conditional on the initial state (Wang et al., 2016b; Tucker et al., 2018). This is also especially true for data sources, such as time series, Markov decision processes and various autoregressive processes, with general dependence patterns (Borrelli & Keviczky, 2008; Mcdonald et al., 2011; Wang et al., 2022). Unfortunately, the literature on CSO with dependent data is remarkably sparse. Ignoring the existence of such dependencies can severely degrade the performance of the model and cause it to deviate from traditional statistical guarantees. Therefore, it is necessary to study the impact of dependency within samples on the SAA of problem (1).

Relaxations of the independence assumption have long been an active area of research in both statistics and machine learning communities. A widely used assumption to relax the assumption of IID is the "mixing" of data generation, where the dependence of the future on the past is made explicit through the quantification of the decay of the dependence as the future moves away from the past (Doukhan, 2012). There are many definitions of mixing of different strengths. Most results in the learning literature focus on β-mixing (Mcdonald et al., 2011), the notion we will use in this paper. The popularity of the study of learning under mixed conditions is partly due to the fact that many temporally dependent data generation processes are mixed. For example, the geometric ergodicity of a strictly stationary Markov chain implies that the β-mixing coefficients are exponentially decaying in time (Lu et al., 2022). Likewise, mixing stochastic processes can be modelled for various types of autoregressive processes (Mokkadem, 1988). More generally, stable dynamical systems subject to finite-variable state noise give rise to mixing Markov processes, a class that encompasses many dynamical systems in system identification, adaptive control, and reinforcement learning (Vidyasagar & Karandikar, 2006).

We summarize the main contributions of this paper as follows.

- Leveraging covariance inequalities and independent block sampling technique, we generalize SAA for CSO to the scenario where training samples are assumed to satisfy β-mixing and establish exponential deviation bounds that hold uniformly over the decision set.
- We prove that the SAA solution is asymptotically consistent and has a finite sample guarantee by establishing a $1-\alpha$ confidence bound on the out-of-sample risk.
- We also establish the sample complexity of $O(d/\varepsilon^4)$ for CSO in the dependent setting. It is shown to be of the same order as independent cases. In addition, we study the impact of the conditional structure of CSO on the sample complexity. The sample complexity decreases to $O(d/\varepsilon^2)$ when the inner and outer randomness are independent.
- The theoretical properties are evaluated in numerical experiments with data generated from a stationary β-mixing stochastic process, which are in support of our theoretical results and show that the dependence of β-mixing data does not degrade the performance of SAA for CSO.

2. Related Work

The properties of the solutions of SAA to CSO are well understood: related work on convergence analysis can be found in Hu et al. (2020a) and Ermoliev & Norkin (2013); the sample complexity of CSO is studied in Shapiro (2006) and Hu et al. (2020a). Dai et al. (2017), Hu et al. (2020b) and Zhang & Xiao (2019) develop a stochastic gradient-based algorithm for the solution of SAA for CSO. Although CSO shares similarities with stochastic composition optimization (SCO) (Wang et al., 2016a; Yu & Huang, 2017; Wang et al., 2017) or multistage stochastic optimization (MSO) (Dupačová, 1995; Swamy & Shmoys, 2012; Pflug & Pichler, 2014), they fundamentally differ: SCO is a composition of two deterministic functions, whereas for CSO the internal randomness is conditional on the external randomness, and the internal expectation is taken over the conditional distribution given the external randomness. CSO is a class of optimization problems situated between stochastic optimization and MSO, while the latter is not constrained regarding the dependence. These differences lead to a drastic difference in the construction of the SAA and in the complexity of the sampling. Our work is most similar to that of Dai et al. (2017); Hu et al. (2020a;b). However, these works focus on sample complexity or tractability where the data are assumed to be IID. Results on performance guarantees, such as asymptotic properties and finite-sample properties of SAA for CSO with unknown probability distributions and dependent training data, are not established and are of particular interest.

Our work has the same strain as a series of papers, such as Wang et al. (2022); Farden (1981); Li et al. (2021); Liu et al. (2023); Pan (2023), which consider the scenario data is dependent. In particular, Pan (2023) establishes a central limit theorem results for the SAA for SO and Wang et al. (2022) shows that the SAA retains consistency and finite sample guarantee when the data is dependent. Under the setup where data are dependent, Liu et al. (2023) develops inferential tools for constructing confidence intervals. Despite these developments, properties, such as asymptotic properties and finite sample performance, of estimators of optimal value and solutions obtained from the well-known SAA approach for CSO with dependent data are still unexplored. In this paper, we aim to fill this gap.

3. Motivating Applications

CSO can be used to model a variety of applications, including robust supervised learning, operational control, portfolio allocation, reinforcement learning, etc. Some of these examples are discussed in detail below.

Reinforcement Learning Consider policy evaluation for a Markov decision process characterized by a tuple $\mathcal{M} := (\mathcal{S}, \mathcal{A}, P, r, \gamma)$, where $\mathcal{S}$ is a state space, $\mathcal{A}$ is an action space, $P(s, a, s')$ represents the transition probability of the state from s to s' given the action a, $r(s,a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is a reward function, and $\gamma \in (0,1)$ is a discount factor. Given a stochastic policy $\pi(a|s)$, the goal of policy evaluation is to estimate the value function $V^\pi(s) := \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r(s_k, a_k) \mid s_0 = s\right]$ under the given policy π. To estimate the value function $V^\pi(s)$, one can minimize the mean squared Bellman error,

$$\min_{V(\cdot):\mathcal{S}\to\mathbb{R}} \mathbb{E}_{s\sim\mu(\cdot), a\sim\pi(\cdot|s)}[(r(s,a) - \mathbb{E}_{s'|a,s}[V(s) - \gamma V(s')])^2],$$

where $\mu(\cdot)$ is the stationary distribution. This minimization problem clearly can be viewed as a special case of CSO with a dependent state-action trajectory where $\xi = (s,a)$, $\eta = s'$, $f_\xi(y) = (r(s,a) - y)^2$, $g_\eta(x,\xi) = V(s) - \gamma V(s')$. A similar finding is also discussed in Dai et al. (2017); Hu et al. (2020a).

Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017) MAML learns a meta-initialization parameter using metadata from similar learning tasks such that taking one or multiple gradient steps on a small training data would generalize well on a new task. It can be framed into the following CSO problem,

$$\min_x \mathbb{E}_{i\sim\mathbb{P}, a\sim D^i_{\text{query}}}\left[\ell_i\left(\mathbb{E}_{b\sim D^i_{\text{support}}}\left(x - \alpha\nabla\ell_i(x,b)\right), a\right)\right],$$

where $\mathbb{P}$ represents the distribution of different tasks, D^i_{support} and D^i_{query} correspond to support (training) data and query (testing) data of the task i, $\ell_i\left(\cdot, D^i\right)$ is the loss function on data D^i from task i, and α is a fixed meta step size. Setting $\xi = (i,a)$ and $\eta = b$, MAML is clearly a special case of CSO for which multiple samples can be drawn from the conditional distribution of $\mathbb{P}(\eta \mid \xi)$.

Noisy Linear Quadratic Regulator (Hambly et al., 2021) The Linear Quadratic Regulator (LQR) problem is one of the most fundamental problems in optimal control theory. The LQR problem is concerned with finding a controller for a linear dynamic system, i.e., a system where the dynamics of the state are described by a linear function of the current state and the inputs, with quadratic cost. Consider the following LQR problem over a finite time horizon T (Hambly et al., 2021),

$$\min_{\{u_t\}_{t=0}^{T-1}} \mathbb{E}_{x_0}[\mathbb{E}_{w_t|x_0}[\sum_{t=0}^{T-1}(x_t^\top Q_t x_t + u_t^\top R_t u_t) + x_T^\top Q_T x_T]],$$

such that for $t = 0, 1, \cdots, T-1$,

$$x_{t+1} = Ax_t + Bu_t + w_t, x_0 \sim \mathbb{P}, \tag{2}$$

where $A \in \mathbb{R}^{d\times d}$ and $B \in \mathbb{R}^{d\times k}$ are referred to as system (transition) matrices; $Q_t \in \mathbb{R}^{d\times d}$ ($\forall t = 0, 1, \cdots, T$) and $R_t \in \mathbb{R}^{k\times k}$ ($\forall t = 0, 1, \cdots, T-1$) are matrices that parameterize the quadratic costs; $x_t \in \mathbb{R}^d$ is the state of the system with the initial state x_0 drawn from a distribution $\mathbb{P}$; $u_t \in \mathbb{R}^p$ is the action taken at time t and $\{w_t\}_{t=1}^{T-1}$ are random noise from distribution $\mathbb{Q}$, which could be dependent from x_0.

Such applications are special cases of the CSO framework. We emphasize that CSO is versatile enough to be generalized to diverse domains such as portfolio selection, peer-to-peer optimization, and model-agnostic meta-learning. Some other motivating examples are presented in the Appendix.

4. Sample Average Approximation

In this paper, we analyze the theoretical properties of the corresponding SAA approach for solving CSO with dependent samples. Let $\{\xi_i\}_{i=1}^N$, $\{\eta_{ij}\}_{j=1}^{M_i}$ be samples that can be viewed as a realization of random variables ξ, η governed by the distribution $\mathbb{P}(\xi)$, $\mathbb{P}(\eta|\xi)$ supported on Ξ, H, separately. An empirical version of problem (1) is the search for an optimal decision $\widehat{x}^\star_{N,M} \in \mathbb{X}$ based on samples $\{\xi_i\}_{i=1}^N$, $\{\eta_{ij}\}_{j=1}^{M_i}$. Formally, we embed problem (1) with SAA to generate optimal solution $\widehat{x}^\star_{N,M}$ by approximating $\mathbb{P}(\xi)$ and $\mathbb{P}(\eta|\xi)$ with the empirical probability distribution $\hat{\mathbb{P}}(\xi) = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_i}$ and $\hat{\mathbb{P}}(\eta|\xi_i) = \frac{1}{M_i}\sum_{j=1}^{M_i} \delta_{\eta_{ij}}$—the distribution that places an equal probability mass at each sample point. Let $F(x) = \mathbb{E}[f_\xi(\mathbb{E}g_{\eta|\xi}(x,\xi))]$ and $\hat{F}_{N,M}(x) = \frac{1}{N}\sum_{i=1}^N f_{\xi_i}\left(\frac{1}{M_i}\sum_{j=1}^{M_i} g_{\eta_{ij}}(x,\xi_i)\right)$. The resultant surrogate optimization by replacement of $\mathbb{P}(\xi)$ and $\mathbb{P}(\eta|\xi)$ with their corresponding empirical distributions can be concisely expressed as

$$\widehat{F}_{N,M} = \min_{x\in\mathbb{X}} \left\{ \frac{1}{N}\sum_{i=1}^N f_{\xi_i}\left(\frac{1}{M_i}\sum_{j=1}^{M_i} g_{\eta_{ij}}(x,\xi_i)\right)\right\}$$
$$\widehat{x}_{N,M} = \arg\min_{x\in\mathbb{X}} \left\{ \frac{1}{N}\sum_{i=1}^N f_{\xi_i}\left(\frac{1}{M_i}\sum_{j=1}^{M_i} g_{\eta_{ij}}(x,\xi_i)\right)\right\}, \tag{3}$$

which approximates $F^\star$ and $x^\star$ in problem (1), where both are functions of the finite samples.

5. Preliminaries

In this section, we first collect notations and definitions that will be used throughout the rest of the paper. We also introduce some of the mathematical tools and propositions that will be necessary for the following discussion. Given the set A, we define an indicator function $\mathbb{I}_A(x) = 1$ if $x \in A$, 0 otherwise. For any $x \in \mathbb{R}^d$ we denote $x_+ := \max\{x, 0\}$, $\|x\|_1 := \sum_{i=1}^d |x_i|$, $\|x\| := \|x\|_2 = (\sum_{i=1}^d x_i^2)^{\frac{1}{2}}$, and $\|x_+\|_0 := \sum_{i=1}^d \mathbb{I}_{[0,\infty)}(x_i)$. Define cl$A$, int$A$ as the closure and interior, respectively, of a set A. A function $f: \mathbb{X} \to \mathbb{R}\cup\infty$ is said to be L-Lipschitz continuous if there exists a constant $L > 0$, such that $|f(x_1) - f(x_2)| \le L\|x_1 - x_2\|_2$, $\forall x_1, x_2 \in \mathbb{X}$. Similar to classical stochastic optimization, we assume no knowledge of the true distribution of $\mathbb{P}(\xi)$ or the conditional distribution of $\mathbb{P}(\eta|\xi)$.

β-mixing We next introduce some definitions for dependent observations in mixing theory (Doukhan, 2012), which provides a unified perspective that allows us to characterize the dependence of data, and then briefly discuss the learning scenarios in the non-IID case.

Definition 5.1 (Stationary). A sequence of random variables $\mathbf{Z} = \{Z_t\}_{t=-\infty}^\infty$ is said to be stationary if for any t and non-negative integers m and k, the random vectors $(Z_t, \ldots, Z_{t+m})$ and $(Z_{t+k}, \ldots, Z_{t+m+k})$ have the same distribution.

In a stationary sequence, the distribution of a variable Z_t is consistent over time. However, this property does not imply independence. In particular, for $i < j < k$, $\mathbb{P}(Z_j \mid Z_i)$ cannot be equal to $\mathbb{P}(Z_k \mid Z_i)$. In the following, we present the definition of β-mixing, a standard measure of the dependence of the random variables within a stationary sequence.

Definition 5.2 (β-mixing). Let $\sigma_l = \sigma(Z_1, \ldots, Z_l)$ and $\sigma'_{l+k} = \sigma(Z_{l+k}, Z_{l+k+1}, \ldots)$, where $\sigma(Z_{i_1}, Z_{i_2}, \ldots, Z_{i_k})$ is the σ-algebra for the collection $(Z_{i_1}, Z_{i_2}, \ldots, Z_{i_k})$. The k^{th} β-mixing coefficient for $(Z_t)_{t=1,2,\ldots}$ is defined as

$$\beta(k) := \sup_{l\ge1} \mathbb{E}\left[\sup_{B\in\sigma'_{l+k}} |\mathbb{P}(B \mid \sigma_l) - \mathbb{P}(B)|\right].$$

The process $(Z_t)_{t=1,2,\ldots}$ is said to be β-mixing if $\beta(k)\to 0$ as $k\to\infty$. Further, we say that $(Z_t)_{t=1,2,\ldots,}$ is algebraically β-mixing if there exist real numbers $\beta_0 > 0$ and $r > 0$ such that $\beta(k) \le \beta_0/k^r$ for all k and is exponentially β-mixing if for some constants $\bar{\beta}_0 \ge 0$ and $\bar{\beta}_1 > 0$, we have $\beta(k) \le \bar{\beta}_0 \exp(-\bar{\beta}_1 k)$.

β-mixing is a weaker assumption than other mixing, such as ϕ-mixing, and thus covers a more general non-IID scenario. Roughly speaking, the coefficient of $\beta(k)$ measures the dependence of an event on those that occurred more than k units of time in the past. Many results in the probability literature rely on β mixing coefficients. The application of β-mixing data to the results of statistical machine learning is also highly desirable in applied work. Many common time series models are known to be β-mixing. For extending the results of IID to dependent data, β-mixing is considered "just right" (Vidyasagar, 1997). Meir (2000) estimates generalization error bounds for nonparametric methods based on model selection using structural risk minimization. Lu et al. (2022) proved an almost sure invariance principle for stationary β-mixing stochastic processes defined in Hilbert space.

Independent Block Sampling Our main result requires the method of independent block sampling as used by Yu (1993). The idea of this method is to transform a sequence of dependent variables into some subsequence that can be viewed as nearly IID. In detail, let r and p be non-negative integers such that $n/(2pr) \to 1$ as $n \to \infty$. We split a dependent sequence $\{Z_1 \ldots, Z_n\}$ into three sequences, $\{U_i\}_{i=1}^r$, $\{V_i\}_{i=1}^r$ with consisting of r blocks and W_n containing the rest. Each block for U_i and V_i contain a p consecutive points. $\{U_i\}_{i=1}^r$, $\{V_i\}_{i=1}^r$, and W_n are defined as follows,

$$\begin{aligned} U_i &= \{Z_j : 2(i-1)p + 1 \leq j \leq (2i-1)p\}, \\ V_i &= \{Z_j : (2i-1)p + 1 \leq j \leq 2ip\}, \\ W_n &= \{Z_j : j = 2rp + 1, \ldots, n\}. \end{aligned}$$

The sequence $\{U_1, \ldots, U_r\}$ is now an IID block sequence, so we can apply standard results of IID to explore the theoretical properties related to $\{Z_i\}$.

6. Probability Deviation Bound of SAA

We will establish the deviation bound of SAA of CSO under the scenario training samples are dependent. Before formalizing it, we first introduce some inequalities (Lemma 6.1, 6.2) that will be used to prove the main results.

Lemma 6.1. *Let $\{Z_j\}_{j=1}^n$ be a β-mixing sequence and $Z_j, Z_{j+k} \in \mathcal{Z}$ be measurable w.r.t. $\sigma(Z_1, \ldots, Z_j)$ and $\sigma(Z_{j+k}, Z_{j+k+1}, \ldots)$ respectively. Suppose a function $h : \mathcal{Z} \to \mathbb{R}$ is Borel measurable. We then have that $|\operatorname{cov}(h(Z_j), h(Z_{j+k}))| \leq 4\|h\|^2\beta(k)$, where $\|h\|^2$ is a bound of function h.*

Lemma 6.2. *(Arab & Oliveira, 2019) Assume that the sequence $\{Z_i\}_{i=1}^n$ is stationary, there exists some $c > 0$ such that for every $i \geq 1, |Z_i| \leq c$ almost surely, and $\frac{1}{n}\operatorname{Var}\left(\sum_{i=1}^n Z_i\right) \leq \sigma^2$. Denote $U_j = \sum_{k=2(j-1)p+1}^{(2j-1)p} Z_k$ with p being consecutive points in each block. Let $d_n > 1$, be a sequence of real numbers. Then, for every $t \leq \frac{d_n - 1}{d_n}\frac{1}{cp}$ and n large enough, $\mathbb{E}\left[e^{tU_j}\right] \leq \exp\left(t^2\sigma^2 p d_n\right)$.*

Lemma 6.1 shows a covariance inequality bound for β-mixing sequence. Lemma 6.2 shows an exponential bound for the Laplace transformation of the blocks U_j mentioned in Section 4. We next introduce commonly used assumptions for CSO for theoretical analysis (Hu et al., 2020a).

Assumption 6.3. (a) The decision set $\mathbb{X} \subseteq \mathbb{R}^d$ is a compact set with a finite diameter $D_{\mathbb{X}} > 0$.
(b) $f_\xi(\cdot) : \mathbb{R}^m \to \mathbb{R}$, $g_\eta(\cdot, \xi) : \mathbb{R}^d \to \mathbb{R}^m$ are Lipschitz continuous, with Lipschitz constants L_f, L_g, respectively, for any given ξ and η.
(c) For any $x \in \mathbb{X}$, $f_\xi(x)$ is Borel measurable in ξ, and $g_\eta(x, \xi)$ is Borel measurable in η for all ξ.
(d) $\sigma_g^2 = \max_{x,\xi} \mathbb{E}\left\|g_\eta(x,\xi) - \mathbb{E}[g_\eta(x,\xi)]\right\|^2 < \infty$.
(e) $|f_\xi(\cdot)| \leq M_f$, $\|g_\eta(\cdot,\xi)\|_2 \leq M_g$ for any ξ and η.

Theorem 6.4 (Deviation bound of SAA)**.** *Assume that $\{\xi_i\}_{i=1}^N, \{\eta_{ij}\}_{j=1}^{M_i}$ are stationary and β-mixing sequence where the mixing coefficient satisfies $\sum_{k=l}^\infty \beta(k) = O(\rho^l)$ for some $\rho \in (0,1)$. Let $d_N > 1$ be a sequence of real numbers, $M = \min_i M_i$. For any $\varepsilon > 0$, under Assumption 6.3, when $L_f^2\left\{M\sigma_g^2 + 8M_g^2 \sum_{j=1}^{M-1}(M-j)\beta(j)\right\} \leq \frac{M^2\varepsilon^2}{16}$, we have that,*

$$\mathbb{P}\left(\sup_{x\in\mathbb{X}}\left|\widehat{F}_{N,M}(x) - F(x)\right| > \varepsilon\right) \leq \tau,$$

where
$\tau = O(1)\left(\frac{4L_f L_g D_{\mathbb{X}}}{\varepsilon}\right)^d \exp\left(-\frac{N\varepsilon^2}{32 d_N M_f^2\left(1+\sum_{i>0}\beta(i)\right)}\right)$.

Theorem 6.4 establishes a deviation bound uniformly in $x \in \mathbb{X}$ for SAA, which shows that the bound is critically dependent on the Lipschitz parameters, the inner/outer dependent structure, and the dimension of the decision set.

The general proof step is similar to the IID case: we first construct a net to get rid of the supreme over x; secondly, we show the required sample size $M = \min_i M_i$ to guarantee $\|\mathbb{E}[f_\xi(\frac{1}{M}\sum_{j=1}^M g_{\eta_j}(x,\xi)) - f_\xi(\mathbb{E}g_\eta(x,\xi))]\|$ is sufficiently small; we thirdly establish the result based on deviation inequality for sums of random variables. However, the price we need to pay under the dependent scenario is that we need to establish the inequality for the measurable function of β-mixing sequence as shown in Lemma 6.1 for the second step. Compared with the well-known deviation inequality for the sums of IID random variables, we need to extend it to the β-mixing sequence which is established in this paper by using the independent block sampling technique and Laplace transformation of the blocks given in Lemma 6.2. From Theorem 6.4, we immediately have the following result.

Corollary 6.5 (ε-optimal solution)**.** *Under Assumption Theorem 6.4, for any $\varepsilon > 0$, we have*

$$\mathbb{P}\left(F(\widehat{x}^\star_{N,M}) - F(x^\star) > \varepsilon\right) \leq \tau,$$

where
$\tau = O(1)\left(\frac{4L_f L_g D_{\mathbb{X}}}{\varepsilon}\right)^d \exp\left(-\frac{N\varepsilon^2}{128 d_N M_f^2\left(1+\sum_{i>0}\beta(i)\right)}\right)$.

These upper bounds may not be practical unless the parameters, e.g. L_f, L_g, are known. Our main concern here is that, despite the constraint, we can still infer the exponential decay rate of the SAA, and that $\widehat{x}^\star_{N,M}$ is a ε-optimal solution.

7. Performance Guarantee

Denote ξ_{test} and $\eta_{\text{test}}|\xi_{\text{test}}$ as test samples that are assumed to be drawn from $\mathbb{P}(\xi)$ and $\mathbb{P}(\eta|\xi)$ and are independent of the training data. Since the true distributions $\mathbb{P}(\xi)$, $\mathbb{P}(\eta|\xi)$ are unknown, the out-of-sample risk defined as $\mathbb{E}\left[f_\xi\left(\mathbb{E}g_{\eta_{\text{test}}}(\widehat{x}^\star_{N,M}, \xi_{\text{test}})\right)\right]$, which measures the validation of the model performance on testing data, cannot yet be evaluated in practice. It is more practical to determine its bounds. It can be seen directly that $F^\star \leq$

$\mathbb{E}\left[f_\xi\left(\mathbb{E}g_{\eta_{\text{test}}}(\widehat{x}^\star_{N,M}, \xi_{\text{test}})\right)\right]$, but this lower bound is impractical unless the true distributions are known. Alternatively, we study the upper bound of the SAA for CSO.

Theorem 7.1 (Finite sample guarantee)**.** *Assume that $\{\xi_i\}_{i=1}^N, \{\eta_{ij}\}_{j=1}^{M_i}$ are stationary and β-mixing sequence where the mixing coefficient satisfies $\sum_{k=l}^\infty \beta(k) = O(\rho^l)$ for some $\rho \in (0,1)$. Under Assumption 6.3, for any $\alpha \in (0,1)$,*

$$\mathbb{P}\{\mathbb{E}\left[f_\xi\left(\mathbb{E}g_{\eta_{test}}(\widehat{x}^\star_{N,M}, \xi_{test})\right)\right] \le \widehat{F}^\star_{N,M} + \varepsilon(\alpha, N, M)\} \ge 1-\alpha,$$

where $\varepsilon(\alpha, N, M)$ is a function of α, N and M and tends to 0 as $N, M \to \infty$ for any fixed α.

Theorem 7.1 shows that the out-of-sample risk is bounded by a sphere of $F^\star_{N,M}$ with radius $\varepsilon(\alpha, N, M)$ with probability $1-\alpha$. This provides a guarantee for prediction by using SAA to obtain an estimator of $x^\star$ with dependent samples. Furthermore, it can be shown that SAA is asymptotically optimal for CSO. That is, the optimal solution $\widehat{x}^\star_{N,M}$ and the optimal value $\widehat{F}^\star_{N,M}$ converge almost surely to their true counterparts $x^\star$, $F^\star$.

Theorem 7.2 (Asymptotic consistency)**.** *Assume that $f_\xi(\cdot) : \mathbb{R}^m \to \mathbb{R}$, $g_\eta(\cdot, \xi) : \mathbb{R}^d \to \mathbb{R}^m$ are convex, closed, proper functions for any given ξ and η. Under Assumption of Theorem 6.4, we have*

$$\lim_{N,M\to\infty} \widehat{F}^\star_{N,M} = F^\star, \quad \lim_{N,M\to\infty} \|\widehat{x}^\star_{N,M} - x^\star\|_2 = 0, \ a.s.$$

In summary, the results established above indicate that SAA for CSO retains the performance guarantee when the data source is β-mixing. Similarly, results can be extended to other suitable mixing conditions with modification.

8. Sample Complexity of SAA for CSO

The above theorems, the deviation-error bound, the finite-sample guarantee, and the asymptotic consistency, depend on the sample size N, M, motivating us to derive the sample complexity of SAA for CSO with dependent data. In this section, we analyze the number of samples required for the SAA to be ε-optimal for solving the CSO problem with high probability.

Theorem 8.1 (Sample complexity)**.** *With probability $1-\alpha$, the solution to the SAA problem is ε-optimal to the original CSO problem if the sample size $M = \min_{i=1,\cdots,N} M_i$ satisfy that $L_f^2\left\{M\sigma_g^2 + 8M_g^2\sum_{j=1}^{M-1}(M-j)\beta(j)\right\} \le \frac{M^2\varepsilon^2}{16}$, and*

$$N \ge O(1)\left(\frac{d_N M_f^2(1+\sum_{i=1}^\infty \beta(i))}{\varepsilon^2}\right)\left[d\log\left(\frac{8l_f L_g D_{\mathbb{X}}}{\varepsilon}\right) + \log\left(\frac{1}{\alpha}\right)\right].$$

From Theorem 8.1 it can be seen that the total sample complexity of SAA for CSO in order to obtain a ε-optimal solution is $O(d/\varepsilon^4)$ if we ignore the log factor. Hu et al. (2020a) obtain the same result for the CSO problem in the independent setting. This indicates that the SAA does not seem to sacrifice accuracy to achieve ε-optimality under the assumption of β-mixing. We emphasize that the lower bound on the complexity of SAA is tight: the order can not be improved without additional assumptions on loss functions (e.g., smoothness). The complexity is tight since our complexity results match the asymptotic rate established in Dentcheva et al. (2017) and achieve the same order of complexity established for CSO under IID cases described in Hu et al. (2020a).

For the sake of comparison with bi-level MSO, we also establish sample complexity under the scenario that η is independent of ξ,

$$F^\star := \min_{x\in\mathbb{X}} \mathbb{E}_\xi\left[f_\xi\left(\mathbb{E}_\eta\left[g_\eta(x,\xi)\right]\right)\right],$$
$$x^\star := \arg\min_{x\in\mathbb{X}} \mathbb{E}_\xi\left[f_\xi\left(\mathbb{E}_\eta\left[g_\eta(x,\xi)\right]\right)\right].$$

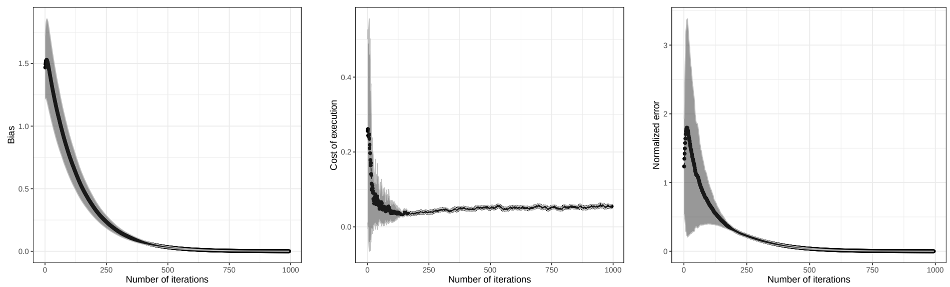
Theorem 8.2 (Sample complexity with independent scheme)**.** *Assume $\{\xi_i\}_{i=1}^N$, $\{\eta_j\}_{j=1}^M$ generated from distribution $\mathbb{P}(\xi)$, $\mathbb{P}(\eta)$ respectively are stationary and β-mixing sequence, and $\{\xi_i\}$, $\{\eta_j\}$ are independent with each other. Under Assumption 6.3, with probability $1-\alpha$, the solution to SAA problem is ε-optimal to the original problem if the sample size N and M satisfy*

$$\begin{cases} N \ge O(1)\frac{1+\sum_{i=1}^\infty \beta(i)}{\varepsilon^2}\left(\log\frac{1}{\alpha} + d\log\frac{4L_f L_g D_{\mathbb{X}}}{\varepsilon}\right), \\ M \ge O(1)\frac{1+\sum_{i=1}^\infty \beta(i)}{\varepsilon^2}\left(\log(\frac{Nm}{\alpha}) + d\log\left(\frac{4L_f L_g D_{\mathbb{X}}}{\varepsilon}\right)\right) \end{cases}.$$

In contrast to the sample complexity of $O(d/\varepsilon^4)$ for CSO, in the independent case the sample complexity drops to $O(d/\varepsilon^2)$.

9. Numerical Experiments

To verify the theoretical results for SAA for CSO with dependent data and its applications in real problems, in this section, we conduct numerical experiments, including Model-Agnostic Meta-Learning (MAML) Linear Quadratic Regulator (LQR), invariant regression, and real application of risk-averse portfolio allocation.



(a) (b) (c)

Figure 1. Bias, value of loss function and normalized error for the learned policy for MAML-LQR.

MAML-LQR In this section, we examine the convergence of SAA for CSO with MAML LQR problem where the initial state is assumed to be dependent. Our data generation procedure is as follows: the initial state of the multi-agent is generated from the autoregressive process, $x_{i+1,0} = 0.9x_{i,0} + e$, $\forall i \in [1, N-1]$, where e is normal noise, N is the number of agents. This means there is a dependence between the state variables among the N agents. The optimal policy in the linear feedback form at time $t \geq 0$ is parameterized by a matrix $K = (K_1, \ldots, K_T)$ given by $u_t = -K_t x_t$. The cost can be expressed as to optimize the policy K:

$$C(K) = \min_{\{u_t\}_{t=0}^{T-1}} \mathbb{E}_{x_0}[\mathbb{E}_{w_t|x_0}[\sum_{t=0}^{T-1}(x_t^\top Q_t x_t + u_t^\top R_t u_t) + x_T^\top Q_T x_T]].$$

We then use the K to take the action ($u_t = -K_t x_t$) and move to the next state: $x_{t+1} = Ax_t + Bu_t + w_t$, $w_t \sim N(0,1)$, $t = 0, 1, \ldots, T-1$. We consider the setting,

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = (-0.6, -1)^\top$$

$$Q_T = \begin{pmatrix} \epsilon & 0 \\ 0 & \delta + \phi\sigma^2 \end{pmatrix}, Q_t = \begin{pmatrix} \epsilon & 0 \\ 0 & \phi\sigma^2 \end{pmatrix}, R_t = \delta.$$

where $\phi = 5 \times 10^{-6}$, $\epsilon = 10^{-8}, T = 10$. We replicate 50 times with initial policy $K^0 \in \mathbb{R}^{1\times 2T}$ with $\left\{K^0\right\}_{ij} = -0.2$ for all i, j.

We plot the bias of the estimate of K against the ground truth $K^\star$, defined as ($|vec(K^t) - vec(K^\star)|$) and is shown in Figure 1 (a). Figure 1 (b) is the value of the loss function based on K^t, and Figure 1 (c) is the normalized error $\frac{|C(K^t) - C(K^\star)|}{C(K^\star)}$, defined as the deviation between the losses based on K^t and $K^\star$.

Figure 1 indicates that although the dependence among the training samples, under β-mixing conditions, SAA for CSO retains convergence guarantee against the ground truth.

Robust Regression via Peer-to-Peer (P2P) Network We consider the problem of robust regression with training samples ξ generated from P2P network

$$\min_x F(x) = \mathbb{E}_{\xi=(a,b)} \left|\mathbb{E}_{\eta|\xi}(\eta^\top x - b)\right|,$$

where $\xi = (a, b)$, $a \in \mathbb{R}^d$ is a random feature vector, and $b \in \mathbb{R}$ is the response. $\eta = \mathcal{N}\left(a, \sigma_\eta^2 I_d\right)$ is a perturbed noisy observation of the input feature vector a with $\sigma_\eta^2 = 1$. Let $d = 10$, and the data is generated as follows, with 10 servers (each with 20000 samples stored), $a_i \sim \mathcal{N}\left(\mu_i, \sigma_\xi^2 I_d\right)$, $b_i = a_i^T x^\star$ with $\sigma_\xi^2 = 1$ and pre-specified $\mu_i, x^\star$. Here, μ_i is allowed to be different from server to server, but the same for σ_ξ^2. For a given sample budget T, ranging from 10^3 to 10^6, we adopt different sample allocation strategies for $N = \{O(T^{\frac{1}{2}}), O(T^{\frac{1}{3}}), O(T^{\frac{1}{4}})\}$, with $M = O(T^{\frac{1}{2}})$ and repeat 30 runs for each sample allocation to report the average performance.

The performance measures include the bias, the in-sample risk to the true values $x^\star$ and $F(x^\star)$, and the probability guarantee, which is defined as $\mathbb{P}(\widehat{x}_{N,M}^\star \in \mathbb{B}_{\varepsilon=0.5}(x^\star))$. Additionally, to compare the performance of the two sampling schemes by choosing different inner samples, we also consider the special case with independent inner and outer randomness. In the independent sampling scheme, $\eta_{ij} = \eta_{1j}$ for all $i > 1$. The results are shown in Figure 2.

Figure 2 (a) shows that the setting $N = O(T^{\frac{1}{2}})$ has the best performance, smallest bias, and highest probability guarantee for robust regression. This is consistent with our sample complexity results. We can see from Figure 2 (b) that the probability guarantee increases exponentially with the increase of the sample size. This is also in support of our theoretical findings of the exponential decay rate of the probability error bound of the bias. The performance of the probabilistic guarantee and the in-sample risk are visualized simultaneously in Figure 2 (c). For two sampling schemes, Figure 2 (d) shows that the in-sample risk is smaller for the independent sampling scheme, and the gap is gradually reduced with increasing sample size, which are consistent

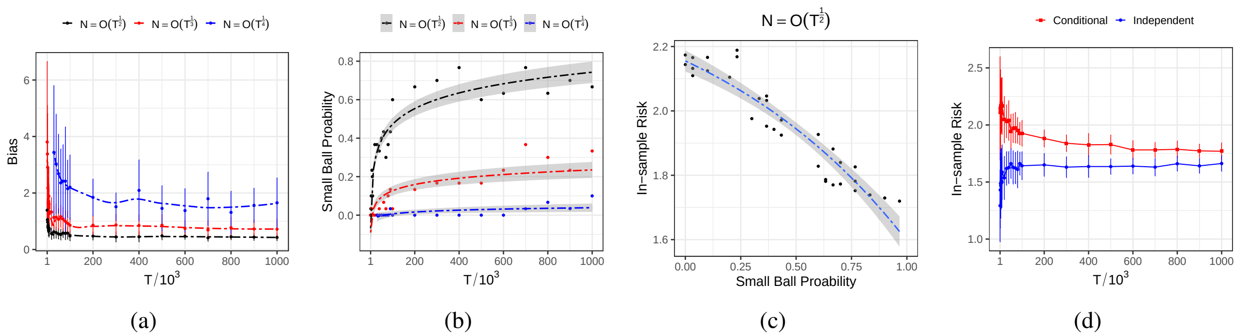


Figure 2. Robust regression problem via P2P network.

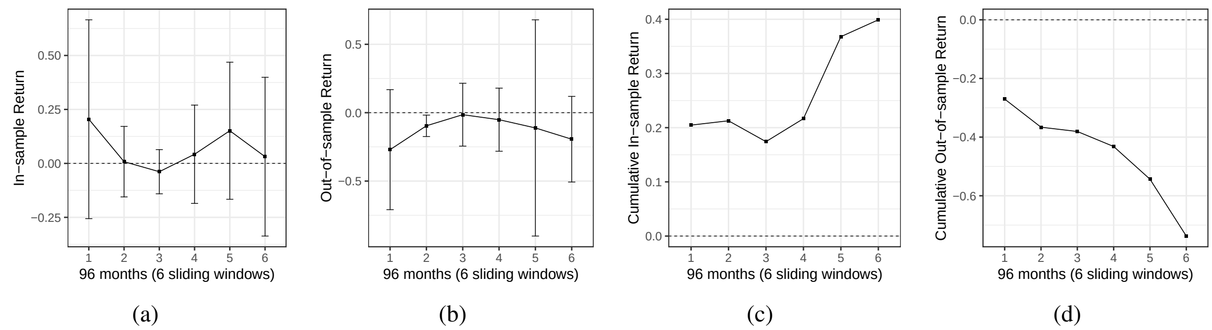


Figure 3. Incremental learning from Feb 2014 to Feb 2022

with our theoretical analysis.

Risk-Averse Portfolio Optimization Problem In this subsection, we study a risk-averse portfolio model by analyzing a portfolio dataset from the Keneth R. French Data Library[1]. The data are collected from Feb 2014 to Feb 2022 with daily data. We would like to seek an optimal decision, x, in portfolio allocation to maximum monthly return $\xi \in \mathbb{R}^d$ for d assets which was invested during N training periods. Let η_{ij} be the j-th daily return at time $i \in \{1, \cdots, N\}$, and $x \in \mathbb{R}^d$ be the decision variable, where each component x_k represents investment percentage allocated to asset k, and satisfies simplex constraint $\mathbb{X} := \{x : x_k \geq 0, \sum_{k=1}^d x_k = 1\}$. We consider the risk-averse problem,

$$\min_{x \in \mathbb{X}} \left\{ -\mathbb{E}_{\widehat{\mathbb{P}}(\xi)} \left[x^\top \mathbb{E}_{\widehat{\mathbb{P}}(\eta|\xi)}[\eta] \right] + \lambda \mathbb{V}_{\widehat{\mathbb{P}}(\xi)} \left[x^\top \mathbb{E}_{\widehat{\mathbb{P}}(\eta|\xi)}[\eta] \right] \right\}$$

where $\widehat{\mathbb{P}}(\xi) = \sum_{i=1}^N \delta_{\xi_i}$, $\widehat{\mathbb{P}}(\eta \mid \xi) = \sum_{j=1}^{M_i} \delta_{\eta_{ij}}$. The λ and $c \in (0, 1)$ are the penalized parameters. In our experiment, we also introduce the penalty term of $r(x) = 0.1\|x\|_1$ and set the parameter, $\lambda = 0.5$ and $N = 24$. Let K months be a window. The incremental learning is implemented in this process: we predict the total return of a window by using the latest 24 months with $K = 12$.

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

The risk-averse portfolio problem is one of the real applications of CSO, where the assets of individuals, firms, or countries are typically dependent, and their monthly unit returns are a kind of time series data. Figure 3 shows the in-sample return, loss function value based on training samples, and out-of-sample return, loss function value based on testing samples, of SAA. As shown in Figure 3 (a), the mean return of 6 sliding windows are above zero, indicating a positive yield. As a result, the cumulative return increases significantly over time (Figure 3 (c)). (Some more results can be found in the Appendix)

10. Conclusion

In this paper, we study the SAA for CSO when the data are dependent. We establish exponential deviation bounds for the SAA and show that the SAA retains asymptotic consistency and the finite sample guarantee when samples form a β-mixing sequence. We also show that the sample complexity with dependent data is of the same order as for IID cases. These results indicate the reasonableness of using SAA for CSO in practice and are verified through numerical experiments and real data applications. Although the performance guarantee under independent cases holds with dependent data under mild conditions, the established results in this paper show that the dependent structure affects the learning efficiency which is captured by the mixing coefficient $\beta(k)$. Studying the effect of $\beta(k)$ on specific models and applications, the sensitivity of loss with different degrees of dependency would be interesting. It is also interesting to find a method to estimate $\beta(k)$ for practical implications to evaluate model performance in applications. We leave these works for the future.

Acknowledgements

We thank the four reviewers for their thoughtful comments. Yafei Wang was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Start-up from the University of Alberta. Lingchen Kong was supported by the 111 Project of China (B16002) and the National Natural Science Foundation of China (12071022,12371322). Bei Jiang and Linglong Kong were partially supported by grants from the Canada CIFAR AI Chairs program, the Alberta Machine Intelligence Institute (AMII), and NSERC, and Linglong Kong was also partially supported by grants from the Canada Research Chair program from NSERC.

Impact Statement

This paper provides an insight into the theoretical guarantee of the SAA for the CSO with dependent data. The results obtained are expected to benefit the mathematics community, statisticians and data scientists. There are no ethical implications and expected societal impacts that we feel need to be emphasized here, due to the theoretical nature of the work.

References

Arab, I. and Oliveira, P. Asymptotic results for certain weak dependent variables. *Theory of Probability and Mathematical Statistics*, 99:19–37, 2019.

Bertsekas, D. *Dynamic programming and optimal control: Volume I*, volume 4. Athena scientific, 2012.

Bertsimas, D., Gupta, V., and Kallus, N. Robust sample average approximation. *Mathematical Programming*, 171: 217–282, 2018.

Borrelli, F. and Keviczky, T. Distributed lqr design for identical dynamically decoupled systems. *IEEE Transactions on Automatic Control*, 53(8):1901–1912, 2008.

Dai, B., He, N., Pan, Y., Boots, B., and Song, L. Learning from conditional distributions via dual embeddings. In *Artificial Intelligence and Statistics*, pp. 1458–1467. PMLR, 2017.

Dentcheva, D., Penev, S., and Ruszczyński, A. Statistical estimation of composite risk functionals and risk optimization problems. *Annals of the Institute of Statistical Mathematics*, 69:737–760, 2017.

Doukhan, P. *Mixing: properties and examples*, volume 85. Springer Science & Business Media, 2012.

Dupačová, J. Multistage stochastic programs: The state-of-the-art and selected bibliography. *Kybernetika*, 31(2): 151–174, 1995.

Ermoliev, Y. M. and Norkin, V. I. Sample average approximation method for compound stochastic optimization problems. *SIAM Journal on Optimization*, 23(4):2231–2263, 2013.

Farden, D. Stochastic approximation with correlated data. *IEEE Transactions on Information Theory*, 27(1):105–113, 1981.

Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pp. 1126–1135. PMLR, 2017.

Fouskakis, D. and Draper, D. Stochastic optimization: a review. *International Statistical Review*, 70(3):315–349, 2002.

Goda, T. and Kitade, W. Constructing unbiased gradient estimators with finite variance for conditional stochastic optimization. *Mathematics and Computers in Simulation*, 204:743–763, 2023.

Hambly, B., Xu, R., and Yang, H. Policy gradient methods for the noisy linear quadratic regulator over a finite horizon. *SIAM Journal on Control and Optimization*, 59(5): 3359–3391, 2021.

Hu, Y., Chen, X., and He, N. Sample complexity of sample average approximation for conditional stochastic optimization. *SIAM Journal on Optimization*, 30(3):2103–2133, 2020a.

Hu, Y., Zhang, S., Chen, X., and He, N. Biased stochastic first-order methods for conditional stochastic optimization and applications in meta learning. *Advances in Neural Information Processing Systems*, 33:2759–2770, 2020b.

Hu, Y., Chen, X., and He, N. On the bias-variance-cost tradeoff of stochastic optimization. *Advances in Neural Information Processing Systems*, 34:22119–22131, 2021.

Kim, S., Pasupathy, R., and Henderson, S. G. A guide to sample average approximation. *Handbook of Simulation Optimization*, pp. 207–243, 2015.

Li, M., Sutter, T., and Kuhn, D. Distributionally robust optimization with markovian data. In *International Conference on Machine Learning*, pp. 6493–6503. PMLR, 2021.

Liu, R., Chen, X., and Shang, Z. Statistical inference with stochastic gradient methods under phi-mixing data. *arXiv preprint arXiv:2302.12717*, 2023.

Lu, J., Wu, W. B., Xiao, Z., and Xu, L. Almost sure invariance principle of $beta$- mixing time series in hilbert space. *arXiv preprint arXiv:2209.12535*, 2022.

Mcdonald, D., Shalizi, C., and Schervish, M. Estimating beta-mixing coefficients. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 516–524. JMLR Workshop and Conference Proceedings, 2011.

Meir, R. Nonparametric time series prediction through adaptive model selection. *Machine Learning*, 39:5–34, 2000.

Mokkadem, A. Mixing properties of arma processes. *Stochastic Processes and Their Applications*, 29(2):309–315, 1988.

Mroueh, Y., Voinea, S., and Poggio, T. A. Learning with group invariant features: A kernel perspective. *Advances in Neural Information Processing Systems*, 28, 2015.

Pan, B. A general framework of optimal stochastic optimization with dependent data: Multiple optimality guarantee, sample complexity, and tractability. 2023.

Pflug, G. C. and Pichler, A. *Multistage stochastic optimization*, volume 1104. Springer, 2014.

Rio, E. Inequalities and limit theorems for weakly dependent sequences. 2013.

Shapiro, A. On complexity of multistage stochastic programs. *Operations Research Letters*, 34(1):1–8, 2006.

Swamy, C. and Shmoys, D. B. Sampling-based approximation algorithms for multistage stochastic optimization. *SIAM Journal on Computing*, 41(4):975–1004, 2012.

Tucker, G., Bhupatiraju, S., Gu, S., Turner, R., Ghahramani, Z., and Levine, S. The mirage of action-dependent baselines in reinforcement learning. In *International Conference on Machine Learning*, pp. 5015–5024. PMLR, 2018.

Vidyasagar, M. *A theory of learning and generalization: with applications to neural networks and control systems*. Springer, 1997.

Vidyasagar, M. and Karandikar, R. L. A learning theory approach to system identification and stochastic adaptive control. *Probabilistic and Randomized Methods for Design under Uncertainty*, pp. 265–302, 2006.

Wang, F. and Janson, L. Exact asymptotics for linear quadratic adaptive control. *The Journal of Machine Learning Research*, 22(1):12136–12247, 2021.

Wang, M., Liu, J., and Fang, E. Accelerating stochastic composition optimization. *Advances in Neural Information Processing Systems*, 29, 2016a.

Wang, M., Fang, E. X., and Liu, H. Stochastic compositional gradient descent: algorithms for minimizing compositions of expected-value functions. *Mathematical Programming*, 161(1):419–449, 2017.

Wang, Y., Pan, B., Tu, W., Liu, P., Jiang, B., Gao, C., Lu, W., Jui, S., and Kong, L. Sample average approximation for stochastic optimization with dependent data: Performance guarantees and tractability. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 3859–3867, 2022.

Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., and Freitas, N. Dueling network architectures for deep reinforcement learning. In *International Conference on Machine Learning*, pp. 1995–2003. PMLR, 2016b.

Wu, X., Sun, J., Hu, Z., Li, J., Zhang, A., and Huang, H. Federated conditional stochastic optimization. *arXiv preprint arXiv:2310.02524*, 2023.

Yang, Z., Chen, Y., Hong, M., and Wang, Z. Provably global convergence of actor-critic: A case for linear quadratic regulator with ergodic cost. *Advances in Neural Information Processing Systems*, 32, 2019.

Yu, B. Density estimation in the $l\infty$ norm for dependent data with applications to the gibbs sampler. *The Annals of Statistics*, pp. 711–735, 1993.

Yu, Y. and Huang, L. Fast stochastic variance reduced admm for stochastic composition optimization. *arXiv preprint arXiv:1705.04138*, 2017.

Zhang, J. and Xiao, L. A stochastic composite gradient method with incremental variance reduction. *Advances in Neural Information Processing Systems*, 32, 2019.

Zhang, J. and Xiao, L. Multilevel composite stochastic optimization via nested variance reduction. *SIAM Journal on Optimization*, 31(2):1131–1157, 2021.

A. Motivated Applications

Robust Invariant Learning (Mroueh et al., 2015) Taking kernel learning as an example, we would like to build an estimator that minimizes the expected risk while preserving consistency over a group of data transformations. Let $\xi_1 = (a_1, b_1), \ldots, \xi_N = (a_N, b_N)$ be a set of input data, where a_i is the feature vector and b_i is the label. One plausible way to achieve such consistency is to consider the class of robust linear classifiers. The robust invariant estimator is then established by averaging functions over all possible kernel transformation $\eta = \sigma(a)$,

$$\min_{(x,x_0)} \mathbb{E}_{\xi=(a,b)} \left[\ell\left(b, \mathbb{E}_{\eta|\xi}\left[\sigma(a)^T x + x_0\right]\right)\right].$$

Here $\ell(\cdot)$ is some loss function and $\mathbb{P}(\eta|\xi)$ is a conditional distribution over all possible kernel transformations. Corresponding to problem (1), $g_\eta(x, \xi) = \eta^T x + x_0$.

Risk-Averse Portfolio Optimization Consider a general robust formulation of mean-variance trade-off in risk-averse portfolio optimization

$$\min_{x\in\mathbb{X}} \left\{-\mathbb{E}^{P(\xi)}\left[x^\top \mathbb{E}^{P(\eta|\xi)}[\eta]\right] + \lambda \mathbb{V}^{P(\xi)}\left[x^\top \mathbb{E}^{P(\eta|\xi)}[\eta]\right]\right\},$$

where ξ, η are monthly and daily returns respectively. This optimization problem can be seen as a special case of CSO with $g(x) = (x^\top \mathbb{E}^{P(\eta|\xi)}[\eta], (\mathbb{E}^{P(\xi)}[x^\top \mathbb{E}^{P(\eta|\xi)}[\eta]])^2)$, $f(y, z) = -y + \lambda y^2 - \lambda z$. The returns are clearly dependent.

B. Proof of Main Results

Lemma B.1 (Rio (2013), Corollary 1.4)**.** *Let $\{\xi_i\}_{i=1}^N$ be a strictly stationary sequence of r.v.S with values in some polish space Ξ. Set $\beta_i = \beta(\alpha(\xi_0), \sigma(\xi_i))$. For any numerical function f, Let $S_N(f) = f(\xi_1) + \cdots + f(\xi_N)$. There exists a sequence $(b_i)_{i\in Z}$ of measurable functions form Ξ into $[0, 1]$, satisfying $\int_\Xi b_i d\mathbb{P} = \beta_i$, $\operatorname{Var}(S_N(f)) \leq N \cdot \int_\Xi \left(1 + 4\sum_{i>0} b_i\right) f^2 d\mathbb{P}$.*

Proof of Lemma 6.1. Denote $\mathcal{F}_1^j = \sigma(Z_1, \ldots, Z_j)$, $\mathcal{F}_{j+k}^\infty = \sigma(Z_{j+k}, Z_{j+k+1}, \ldots)$. Since Z_j and Z_{j+k} are $\mathcal{F}_1^k$ measurable and $\mathcal{F}_{n+k}^\infty$ measurable respectively, we have that

$$\begin{aligned}
|\mathbb{E}[h(Z_j)h(Z_{j+k})] - (\mathbb{E}[h(Z_j)])(\mathbb{E}[h(Z_{j+k})])| &= \left|\mathbb{E}\left[\mathbb{E}[h(Z_j)h(Z_{j+k}) \mid \mathcal{F}_1^k]\right] - (\mathbb{E}[h(Z_j)])(\mathbb{E}[h(Z_{j+k})])\right| \\
&= \left|\mathbb{E}\left[h(Z_j)\cdot\mathbb{E}[h(Z_{j+k}) \mid \mathcal{F}_1^k]\right] - (\mathbb{E}[h(Z_j)])(\mathbb{E}[h(Z_{j+k})])\right| \\
&\leq \|h\| \cdot \mathbb{E}\left|\mathbb{E}[h(Z_{j+k}) \mid \mathcal{F}_1^k] - (\mathbb{E}[h(Z_{j+k})])\right| \\
&= \|h\| \cdot \mathbb{E}\left[\tilde{\xi}\left(\mathbb{E}[h(Z_{j+k}) \mid \mathcal{F}_1^k] - (\mathbb{E}[h(Z_{j+k})])\right)\right] \\
&= \|h\| \cdot \mathbb{E}\left[\tilde{\xi}\mathbb{E}[h(Z_{j+k}) \mid \mathcal{F}_1^k] - \tilde{\xi}(\mathbb{E}[h(Z_{j+k})])\right] \\
&= \|h\| \cdot \left[\mathbb{E}\left[\tilde{\xi}h(Z_{j+k})\right] - \left(\mathbb{E}[\tilde{\xi}]\right)(\mathbb{E}[h(Z_{j+k})])\right]
\end{aligned}$$

where $\tilde{\xi} = \operatorname{sign}\left(\mathbb{E}[h(Z_{j+k}) \mid \mathcal{F}_1^k] - (\mathbb{E}[h(Z_{j+k})])\right)$ is $\mathcal{F}_1^k$ measurable.

$$\begin{aligned}
\left|\mathbb{E}\left[\tilde{\xi}h(Z_{j+k})\right] - \left(\mathbb{E}[\tilde{\xi}]\right)(\mathbb{E}[h(Z_{j+k})])\right| &= \left|\mathbb{E}\left[\mathbb{E}\left[\tilde{\xi}h(Z_{j+k}) \mid \mathcal{F}_{n+k}^\infty\right]\right] - \left(\mathbb{E}[\tilde{\xi}]\right)(\mathbb{E}[h(Z_{j+k})])\right| \\
&= \left|\mathbb{E}\left[\mathbb{E}h(Z_{j+k})\left[\tilde{\xi} \mid \mathcal{F}_{n+k}^\infty\right]\right] - \left(\mathbb{E}[\tilde{\xi}]\right)(\mathbb{E}[h(Z_{j+k})])\right| \\
&\leq \|h\| \cdot \left|\mathbb{E}\left[\mathbb{E}\left[\tilde{\xi} \mid \mathcal{F}_{n+k}^\infty\right]\right] - \left(\mathbb{E}[\tilde{\xi}]\right)\right| \\
&= \|h\| \cdot \mathbb{E}\left[\tilde{\eta}\left(\mathbb{E}\left[\tilde{\xi} \mid \mathcal{F}_{n+k}^\infty\right] - \left(\mathbb{E}[\tilde{\xi}]\right)\right)\right] \\
&= \|h\| \cdot \left[\mathbb{E}[\tilde{\eta}\tilde{\xi}] - (\mathbb{E}[\tilde{\eta}])\left(\mathbb{E}[\tilde{\xi}]\right)\right],
\end{aligned}$$

where $\tilde{\eta} = \operatorname{sign}\left(\mathbb{E}\left[\tilde{\xi} \mid \mathcal{F}_{n+k}^\infty\right] - \left(\mathbb{E}[\tilde{\xi}]\right)\right)$ is $\mathcal{F}_{n+k}^\infty$ measurable. It follows that

$$|\mathbb{E}[h(Z_j)h(Z_{j+k})] - (\mathbb{E}[h(Z_{j+k})]\mathbb{E}[h(Z_j)])| \leq \|h\|^2 \cdot \left|\mathbb{E}[\tilde{\eta}\tilde{\xi}] - (\mathbb{E}[\tilde{\eta}])\left(\mathbb{E}[\tilde{\xi}]\right)\right|.$$

Let $B=\{\tilde{\eta}=1\}\in\mathcal{F}_{n+k}^{\infty}$, $A=\{\tilde{\xi}=1\}\in\mathcal{F}_1^k$. Therefore,

$$
\begin{aligned}
&|\mathbb{E}[h(Z_j)h(Z_{j+k})]-(\mathbb{E}[h(Z_j)]\mathbb{E}[h(Z_{j+k})])| \\
&\leq \|h\|^2\cdot\left|\mathbb{P}(AB)+\mathbb{P}(\bar{A}\bar{B})-\mathbb{P}(A\bar{B})-\mathbb{P}(\bar{A}B)-\left[\mathbb{P}(A)\mathbb{P}(B)-\mathbb{P}(A)\mathbb{P}(\bar{B})-\mathbb{P}(\bar{A})\mathbb{P}(B)+\mathbb{P}(\bar{A})\mathbb{P}(\bar{B})\right]\right| \\
&=\|h\|^2\cdot\left|[\mathbb{P}(AB)-\mathbb{P}(A)\mathbb{P}(B)]+\left[\mathbb{P}(\bar{A}\bar{B})-\mathbb{P}(\bar{A})\mathbb{P}(\bar{B})\right]-\left[\mathbb{P}(\bar{A}B)-\mathbb{P}(\bar{A})\mathbb{P}(B)\right]-\left[\mathbb{P}(A\bar{B})-\mathbb{P}(A)\mathbb{P}(\bar{B})\right]\right| \\
&\leq \|h\|^2\cdot 4\alpha(k)\leq 4\|h\|^2\cdot\beta(k).
\end{aligned}
$$

Then, we have that

$$
|\mathrm{Cov}(h(Z_j),h(Z_{j+k}))|=|\mathbb{E}[h(Z_j)h(Z_{j+k})]-(\mathbb{E}[h(Z_j)])\,(\mathbb{E}[h(Z_{j+k})])\leq 4\|h\|^2\beta(k).
$$

Proof of Theorem 6.4. We first construct a ν-net $\{x_l\}_{l=1}^K$ on the set $\mathbb{X}$ with $L_fL_g\nu=\frac{\varepsilon}{4}$ to get rid of the $\sup$ over x, which implies $K\leq O(1)\left(\frac{4L_fL_gD_{\mathbb{X}}}{\varepsilon}\right)^d$. Based on the Lipschitz continuity of $f_\xi(\cdot)$, $g_\eta(\cdot)$, for any $x\in\mathbb{X}$, there exist $x_{l(x)}$ such that

$$
\begin{aligned}
\left|\hat{F}_{N,M}(x)-\hat{F}_{N,M}\left(x_{l(x)}\right)\right| &\leq L_fL_g\left\|x-x_{l(x)}\right\|_2\leq\frac{\varepsilon}{4}, \\
\left|F(x)-F\left(x_{l(x)}\right)\right| &\leq L_fL_g\left\|x-x_{l(x)}\right\|_2\leq\frac{\varepsilon}{4}.
\end{aligned}
$$

Therefore, for any $x\in\mathbb{X}$, we have that,

$$
\begin{aligned}
\left|\hat{F}_{N,M}(x)-F(x)\right| &\leq\left|\hat{F}_{N,M}(x)-\hat{F}_{N,M}\left(x_{l(x)}\right)\right|+\left|\hat{F}_{N,M}\left(x_{l(x)}\right)-F(x_{l(x)})\right| \\
&\quad+\left|F\left(x_{l(x)}\right)-F(x)\right| \\
&\leq\frac{\varepsilon}{2}+\left|\hat{F}_{N,M}\left(x_{l(x)}\right)-F\left(x_{l(x)}\right)\right| \\
&\leq\frac{\varepsilon}{2}+\max_{l\in\{1,\cdots,K\}}\left|\hat{F}_{N,M}\left(x_{l(x)}\right)-F\left(x_{l(x)}\right)\right|, \qquad \text{(A1)}
\end{aligned}
$$

which implies

$$
\mathbb{P}\left(\sup_{x\in\mathbb{X}}\left|\hat{F}_{N,M}(x)-F(x)\right|>\varepsilon\right)\leq\sum_{l=1}^K\mathbb{P}\left(\left|\hat{F}_{N,M}\left(x_{l(x)}\right)-F\left(x_{l(x)}\right)\right|>\frac{\varepsilon}{2}\right).
$$

Let $\tilde{Z}_i=f_{\xi_i}\left(\frac{1}{M}\sum_{j=1}^m g_{\eta_{ij}}\left(x,\xi_i\right)\right)-\mathbb{E}\left[f_\xi\left(\mathbb{E}\left[g_\eta(x,\xi)\right]\right)\right]$, where $\{\tilde{Z}_i\}_{i=1}^N$ is a β-mixing sequence since f_ξ is a measurable function.We first show that $|\mathbb{E}(\tilde{Z}_i)|\leq\frac{\varepsilon}{4}$,

$$
\begin{aligned}
\left|\mathbb{E}\left[f_\xi\left(\frac{1}{M}\sum_{j=1}^M g_{\eta_j}\left(x,\xi\right)\right)-f_\xi\left(\mathbb{E}\left[g_\eta(x,\xi)\right]\right)\right]\right| &\leq L_f\mathbb{E}\left\|\frac{1}{M}\sum_{j=1}^M g_{\eta_j}(x,\xi)-\mathbb{E}\left[g_\eta(x,\xi)\right]\right\| \\
&\leq L_f\left\{\mathbb{E}\left\|\frac{1}{M}\sum_{j=1}^M g_{\eta_j}(x,\xi)-\mathbb{E}\left[g_\eta(x,\xi)\right]\right\|^2\right\}^{1/2} \\
&=L_f\left\{\left\|\mathrm{Cov}\left(\frac{1}{M}\sum_{j=1}^M g_{\eta_j}(x,\xi),\quad\frac{1}{M}\sum_{j=1}^M g_{\eta_j}(x,\xi)\right)\right\|\right\}^{1/2} \\
&=L_f\left\{\frac{1}{M^2}\sum_{j,l=1}^M\left\|\mathrm{Cov}\left(g_{\eta_j}(x,\xi),\quad g_{\eta_l}(x,\xi)\right)\right\|\right\}^{1/2} \\
&\leq L_f\left\{\frac{\sigma_g^2}{M}+\frac{8M_g^2}{M^2}\sum_{j=1}^{M-1}(M-j)\beta(j)\right\}^{1/2}\leq\frac{\varepsilon}{4}.
\end{aligned}
$$

Then, when $|\mathbb{E}(\tilde{Z}_i)| \leq \frac{\varepsilon}{4}$, we have

$$\mathbb{P}\left(|F_{N,M}(x) - F(x)| > \frac{\varepsilon}{2}\right) = \mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} Z_i\right| > \frac{\varepsilon}{2}\right) \leq \mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N}\left(\tilde{Z}_i - \mathbb{E}\left[\tilde{Z}_i\right]\right)\right| > \frac{\varepsilon}{4}\right).$$

We next explore the property of $\mathbb{P}\left(\left|\frac{1}{N}\sum_{i}^{N}\left(\tilde{Z}_i - \mathbb{E}[\tilde{Z}_i]\right)\right| > \frac{\varepsilon}{4}\right)$. We first introduce some of the notations that are essential for the theoretical development. Let $Z_i := \tilde{Z}_i - \mathbb{E}[\tilde{Z}_i]$, $U_k = Z_{2(k-1)p+1} + \cdots + Z_{(2k-1)p}$, $V_k = Z_{(2k-1)p+1} + \cdots + Z_{2kp}$, and $W_N = Z_{2pr+1} + \cdots + Z_N$, for $k = 1, \cdots, r$. Note that $|Z_i| \leq 2M_f$ and $|U_k| \leq 2pM_f$. We denote $U_N = \sum_{k=1}^{r} U_k$, $V_N = \sum_{k=1}^{r} V_k$. Thus, we have $\sum_{i=1}^{N} Z_i = U_N + V_N + W_N$.

To show the bound of $\mathbb{P}\left(\left|\sum_{i=1}^{N} Z_i\right| > N\varepsilon\right)$, we only need to analyze the term $\mathbb{P}\left(|U_N| > N\varepsilon\right)$ and $\mathbb{P}\left(|U_N| > N\varepsilon\right) \leq e^{-N\varepsilon t}\mathbb{E}[e^{tU_N}]$. Let $C_Z := 2M_f$.

$$\begin{aligned}
\mathbb{E}\left[e^{tU_N}\right] &= \mathbb{E}\left[e^{t\sum_{k=1}^{r} U_k}\right] = \mathrm{Cov}\left(e^{t\sum_{k=1}^{r-1} U_k}, e^{tU_r}\right) + \mathbb{E}\left[e^{t\sum_{k=1}^{r-1} U_k}\right]\mathbb{E}\left[e^{tU_r}\right] \\
&= \mathrm{Cov}\left(e^{t\sum_{k=1}^{r-1} U_k}, e^{tU_r}\right) + \mathrm{Cov}\left(e^{t\sum_{k=1}^{r-2} U_k}, e^{tU_{r-1}}\right)\left(\mathbb{E}\left[e^{tU_r}\right]\right) \\
&\quad + \left(\mathbb{E}\left[e^{t\sum_{k=1}^{r-2} U_k}\right]\right)\left(\mathbb{E}\left[e^{tU_{r-1}}\right]\right)\left(\mathbb{E}\left[e^{tU_r}\right]\right) \\
&= \cdots = \sum_{l=1}^{r-1} \mathrm{Cov}\left(e^{t\sum_{k=1}^{r-l} U_k}, e^{tU_{r-l+1}}\right)\left(\mathbb{E}\left[e^{tU_1}\right]\right)^{l-1} + \left(\mathbb{E}\left[e^{tU_1}\right]\right)^{r}.
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}\left(e^{t\sum_{k=1}^{r-l} U_k}, e^{tU_{r-l+1}}\right) &\leq t^2 \cdot \exp\left\{t(r-l+1)pC_Z\right\}\sum_{k=1}^{r-l} \mathrm{Cov}\left(U_k, U_{r-l+1}\right) \\
&= t^2 \cdot \exp\left\{t(r-l+1)pC_Z\right\}\sum_{k\in A_1}\sum_{l\in A_2} \mathrm{Cov}\left(U_k, U_l\right),
\end{aligned}$$

where

$$\begin{cases}
A_1 = \{\underbrace{1, \cdots, p}_{A_{11}}, \quad \underbrace{2p+1, \cdots, 3p}_{A_{12}}, \cdots, \underbrace{2(r-k-1)p+1, \cdots, (2(r-k)-1)p}_{A_1(r-k)}, \\
A_2 = \{2(r-k)p+1, \cdots, (2(r-k)+1)p\}, \\
A_{1i} = \{2(i-1)p+1, \cdots, (2i-1)p\}, i = 1, \cdots, r-k.
\end{cases}$$

$$\begin{aligned}
\sum_{k\in A_{11}}\sum_{l\in A_2} \mathrm{Cov}\left(U_k, U_l\right) &\leq p \cdot \mathrm{Cov}\left(Z_1, Z_{2(r-k)p+1}\right) + \cdots + p\,\mathrm{Cov}\left(Z_p, Z_{2(r-k)p+1}\right) \\
&\leq p^2 \cdot \mathrm{Cov}\left(Z_p, Z_{2(r-k)p+1}\right) = p^2 \cdot \mathrm{Cov}\left(Z_1, Z_{2(r-k)p-p+1}\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{k\in A_{1i}}\sum_{l\in A_2} \mathrm{Cov}\left(U_k, U_l\right) &\leq p \cdot \mathrm{Cov}\left(Z_{2(i-1)p+1}, Z_{2(r-k)p+1}\right) + \cdots + p \cdot \mathrm{Cov}\left(Z_{(2i-1)p}, Z_{2(r-k)p+1}\right) \\
&= p \cdot \sum_{l=(2r-2k-2i+2)p-p+1}^{(2r-2k-2i+2)p} \mathrm{Cov}(Z_1, Z_l).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{k=1}^{r-l} \mathrm{Cov}\left(U_k, U_{r-l+1}\right) &\leq \sum_{k=1}^{r-l} p \cdot \sum_{i=(2r-2l-2k+2)p-p+1}^{(2r-2l-2k+2)p} \mathrm{cov}\left(Z_1, Z_i\right) \\
&\leq \sum_{k=1}^{r-l} p \cdot 4C_Z^2 \cdot \sum_{i} \beta(i) \leq 4pC_Z^2 \sum_{i=p}^{\infty} \beta(i).
\end{aligned}$$

This inequality together with Lemma 6.2, we have that

$$
\begin{aligned}
\mathbb{E}\left[e^{tU_N}\right] &\leq \sum_{l=1}^{r-1} t^2 \exp\left\{tpC_Z(r-l+1)\right\}\left(4pC_Z^2\right)\left(\sum_{i=p}^{\infty}\beta(i)\right)\cdot\left\{\left(\mathbb{E}\left[e^{tU_1}\right]\right)^{l-1}\right\}+\left(\mathbb{E}[e^{tU_1}]\right)^r \\
&\leq 4pC_Z^2\left(\sum_{i=p}^{\infty}\beta(i)\right)\cdot\sum_{l=1}^{r-1} t^2\exp\left\{tpC_Z(r-(l-1))\right\}\cdot\exp\left\{t^2\sigma^2 pd_N(l-1)\right\}+\exp\left\{t^2\sigma^2 pd_N r\right\} \\
&= 4pC_Z^2\left(\sum_{i=p}^{\infty}\beta(i)\right)\cdot\sum_{l=0}^{r-2} t^2\exp\left\{tpC_Z r - tpC_Z l + t^2\sigma^2 pd_N l\right\}+\exp\left\{t^2\sigma^2 pd_N r\right\} \\
&= 4pC_Z^2\left(\sum_{i=p}^{\infty}\beta(i)\right) t^2\exp\left\{tpC_Z r\right\}\cdot\sum_{l=0}^{r-2}\exp\left\{tpl\left(t\sigma^2 d_N - C_Z\right)\right\}+\exp\left\{t^2\sigma^2 N d_N/2\right\} \\
&= 4pC_Z^2 t^2\left(\sum_{i=p}^{\infty}\beta(i)\right)\exp\left\{\frac{tC_Z N}{2}\right\}\cdot\sum_{l=0}^{r-2}\exp\left\{tpl\left(t\sigma^2 d_N - C_Z\right)\right\}+\exp\left\{\frac{t^2\sigma^2 N d_N}{2}\right\}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\mathbb{P}(|U_N|>N\varepsilon) \leq e^{-N\varepsilon t}4pC_Z^2 t^2&\left(\sum_{i=p}^{\infty}\beta(i)\right)\exp\left\{\frac{tC_Z N}{2}\right\}\cdot\sum_{l=0}^{r-2}\exp\left\{tpl(t\sigma^2 d_N - C_Z)\right\} \\
&+\exp\left\{\frac{t^2\sigma^2 N d_N}{2}-N\varepsilon t\right\}.
\end{aligned}
$$

Let $t=\frac{\varepsilon}{\sigma^2 d_N}$, we have $\frac{t^2\sigma^2 N d_N}{2}-N\varepsilon t = -\frac{N\varepsilon^2}{2\sigma^2 d_N}$ and $t\sigma^2 d_N - C_Z = \varepsilon - C_Z < 0$ for small ε. This indicates that $\sum_{l=0}^{r-2}\exp\left\{tpl\left(t\sigma^2 d_N - C_Z\right)\right\}$ converges. In addition, $N\varepsilon t = \frac{N\varepsilon^2}{\sigma^2 d_N}$, so, $e^{-N\varepsilon t} = \exp\left(-\frac{N\varepsilon^2}{\sigma^2 d_N}\right) \leq \exp\left(-\frac{N\varepsilon^2}{2\sigma^2 d_N}\right)$. Therefore, there exists a constant $C'>0$, such that

$$
\begin{aligned}
\mathbb{P}\left(|U_N|>N\varepsilon\right) &\leq \left\{\left(\frac{4pC_Z^2\varepsilon^2}{\sigma^4 d_N^2}\right)\left(\sum_{i=p}^{\infty}\beta(i)\right)\exp\left(\frac{\varepsilon C_Z N}{2\sigma^2 d_N}\right)\cdot C' + 1\right\}\exp\left(-\frac{N\varepsilon^2}{2\sigma^2 d_N}\right) \\
&= \left\{\frac{C'\varepsilon^2}{\sigma^4 d_N^2}e^{\frac{C_Z N\varepsilon}{2\sigma^2 d_N}}\cdot p\cdot\left(\sum_{i=p}^{\infty}\beta(i)\right)+1\right\}\exp\left(-\frac{N\varepsilon^2}{2\sigma^2 d_N}\right).
\end{aligned}
$$

Note that $t=\frac{\varepsilon}{\sigma^2 d_N} \leq \frac{d_N-1}{pC_Z d_N}$ implies $\varepsilon \leq \frac{\sigma^2(d_N-1)}{pC_Z}$ under the condition $d_N>1$. Therefore, with $\frac{pd_N}{N}>0$,

$$
\exp\left\{\frac{\varepsilon C_Z N}{2\sigma^2 d_N}\right\} \leq \exp\left\{\frac{N}{2pd_N}\left(d_N-1\right)\right\} = \exp\{r\}\cdot\exp\left\{-\frac{N}{2pd_N}\right\}.
$$

Then, there exists a constant $C>0$, such that, $\mathbb{P}\left(|U_N|>N\varepsilon\right) \leq C\cdot\exp\left(-\frac{N\varepsilon^2}{2\sigma^2 d_N}\right)$.

$$
\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N}Z_i\right|>\varepsilon\right) &\leq \mathbb{P}\left(\left|\frac{U_N}{N}\right|+\left|\frac{V_N}{N}\right|+\left|\frac{W_N}{N}\right|>\varepsilon\right) \leq \mathbb{P}\left(\left|\frac{U_N}{N}\right|+\left|\frac{V_N}{N}\right|+\frac{2pC_Z}{N}>\varepsilon\right) \\
&\leq \mathbb{P}\left(\left|\frac{U_N}{N}\right|+\left|\frac{V_N}{N}\right|>\frac{\varepsilon}{2}\right) \leq 2\cdot\mathbb{P}\left(\left|\frac{U_N}{N}\right|>\frac{\varepsilon}{4}\right) \leq C\cdot\exp\left(-\frac{N\varepsilon^2}{32\sigma^2 d_N}\right).
\end{aligned}
$$

In addition, Lemma B.1 indicates that

$$
\mathrm{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i\right) = \frac{1}{N}\mathrm{Var}\left(\sum_{i=1}^{N}Z_i\right) \leq \frac{1}{N}\cdot N\cdot M_f^2\left(1+\sum_{i>0}\beta(i)\right) = M_f^2\left(1+\sum_{i>0}\beta(i)\right).
$$

That is,

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} Z_i\right| > \varepsilon\right) \leq C \cdot \exp\left(-\frac{N\varepsilon^2}{32 d_N M_f^2 \left(1 + \sum_{i>0}\beta(i)\right)}\right).$$

The results of the theorem hold.

Proof of Corollary 6.5. Note that

$$\begin{aligned}
&\mathbb{P}(F(\hat{x}^\star_{N,M}) - F(x^\star) > \varepsilon) \\
&= \mathbb{P}(F(\hat{x}^\star_{N,M}) - F_{N,M}(\hat{x}^\star_{N,M}) + F_{N,M}(\hat{x}^\star_{N,M}) - F_{N,M}(x^\star) + F_{N,M}(x^\star) - F(x^\star) > \varepsilon) \\
&\leq \mathbb{P}(F(\hat{x}^\star_{N,M}) - F_{N,M}(\hat{x}^\star_{N,M}) > \varepsilon/2) + \mathbb{P}(F_{N,M}(x^\star) - F(x^\star) > \varepsilon/2),
\end{aligned}$$

where the last inequality holds because $F_{N,M}(\hat{x}^\star_{N,M}) - F_{N,M}(x^\star) < 0$. Based on the result of Theorem 6.4, the result of Corollary 6.5 holds.

Proof of Theorem 7.1. Let $\alpha = O(1)\left(\frac{4L_f L_g D_{\mathbb{X}}}{\varepsilon}\right)^d \exp\left(-\frac{N\varepsilon^2}{32 d_N M_f^2\left(1+\sum_{i>0}\beta(i)\right)}\right)$, we get the result of Theorem 7.1 from Theorem 6.4.

Proof of Theorem 7.2. Let $\hat{F}_{N,M}(x,\xi,\eta) = \frac{1}{N}\sum_{i=1}^N f_{\xi_i}\left(\frac{1}{M_i}\sum_{j=1}^{M_i} g_{\eta_{ij}}(x,\xi_i)\right)$, $F(x) = \mathbb{E}[f_\xi\left(\mathbb{E}g_{\eta|\xi}(x,\xi)\right)]$. Define a set of approximate solutions as

$$x^\star_\varepsilon = \{x \in \mathbb{X} : F(x) \leq F^\star + \varepsilon\}.$$

Let us rewrite the approximation problem in a parametric form

$$\Phi(x, y(x,\xi,\eta)) = F(x) + \hat{F}_{NM}(x,\xi,\eta) - F(x)$$

where $y(x,\xi,\eta) = \hat{F}_{NM}(x,\xi,\eta) - F(x)$ and consider a parametric problem

$$\Phi(x, y(x)) = F(x) + y(x).$$

Theorem 6.4 indicates that with probability 1, $\hat{F}_{NM}(x,\xi,\eta) \to F(x)$ uniformly in $x \in \mathbb{X}$, which implies $y(x) \to 0$ uniformly in $x \in \mathbb{X}$. Denote

$$\begin{aligned}
\Phi^\star(y) &= \min_x\{\Phi(x, y(x))\}, \quad y : \mathbb{X} \to \mathbb{R}, \\
x^\star(y) &= \{x \in \mathbb{X} : \Phi(x, y(x)) = \Phi^\star(y)\}.
\end{aligned}$$

Since the functions f, g are convex, bounded functions and the set $\mathbb{X}$ is compact, we have $\|x^\star(y) - x^\star\|_2 \to 0$ as $y \to 0$. Therefore,

$$\begin{aligned}
\hat{F}^\star_{NM} &= \Phi^\star(\cdot, y(x,\xi,\eta)) \to \Phi^\star(0) = F^\star, \quad a.s. \\
\|\hat{x}^\star_{NM} - x^\star\|_2 &= \|x^\star(y(\cdot,\xi,\eta)) - x^\star\|_2 \to 0, \quad a.s..
\end{aligned}$$

Proof of Theorem 8.1. We get the result of Theorem 8.1 from the ε-optimal solution given in Corollary 6.5.

Proof of Theorem 8.2. From formula (A1), one can see that it only needs to show the boundedness of

$\mathbb{P}\left(\left|\hat{F}_{N,M}(x) - F(x)\right| > \varepsilon\right)$ for fix $x \in \mathbb{X}$. Note that

$$
\begin{aligned}
&\mathbb{P}\left(\left|\hat{F}_{N,M}(x) - F(x)\right| > \varepsilon\right) \\
&\le \mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f_{\xi_i}\left(\frac{1}{M}\sum_{j=1}^{M} g_{\eta_j}(x,\xi_i)\right) - \frac{1}{N}\sum_{i=1}^{N} f_{\xi_i}\left(\mathbb{E}\left(g_\eta(x,\xi_i)\right)\right)\right| > \frac{\varepsilon}{2}\right) \\
&\quad + \mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f_{\xi_i}\left(\mathbb{E}_\eta\left(g_\eta(x,\xi)\right)\right) - \mathbb{E}_\xi\left[f_\xi\left(\mathbb{E}_\eta\left(g_\eta(x,\xi)\right)\right)\right]\right| > \frac{\varepsilon}{2}\right) \\
&\le \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N}\left| f_{\xi_i}\left(\frac{1}{M}\sum_{j=1}^{M} g_{\eta_j}(x,\xi_i)\right) - f_{\xi_i}\left(\mathbb{E}_\eta\left(g_\eta(x,\xi_i)\right)\right)\right| > \frac{\varepsilon}{2}\right) \\
&\quad + \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N} f_{\xi_i}\left(\mathbb{E}_\eta\left(g_\eta(x,\xi_i)\right) - \mathbb{E}_\eta\left[f_\xi\left(\mathbb{E}_\eta\left(g_\eta(x,\xi)\right)\right)\right] \right| > \frac{\varepsilon}{2}\right) \\
&\le \sum_{i=1}^{N}\mathbb{P}\left(\left\|\frac{1}{M}\sum_{j=1}^{M} g_{\eta_j}(x,\xi_i) - E_\eta\left(g_\eta(x,\xi_i)\right)\right\| > \frac{\varepsilon}{2L_f}\right) + C\cdot\exp\left(-\frac{N\varepsilon^2}{32 d_N M_f^2\left(1+\sum_{i>0}\beta(i)\right)}\right) \\
&\le N\cdot m\cdot\exp\left(-\frac{M\varepsilon^2}{32 L_f^2 d_M M_g^2\left(1+\sum_{i>0}\beta(i)\right)}\right) + C\cdot\exp\left(-\frac{N\varepsilon^2}{32 d_N M_f^2\left(1+\sum_{i>0}\beta(i)\right)}\right).
\end{aligned}
$$

This inequality combining with formula (A1) indicates that

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{x\in\mathbb{X}}\left|\frac{1}{N}\sum_{i=1}^{N} f_{\xi_i}\left(\frac{1}{M_i}\sum_{j=1}^{M_i} g_{\eta_j}(x,\xi_i)\right) - \mathbb{E}_\xi\left[f_\xi\left(E_\eta g(x;\xi)\right)\right]\right| > \varepsilon\right) \\
&\le O(1)\left(\frac{4L_f L_g D_{\mathbb{X}}}{\varepsilon}\right)^d\left(\exp\left(-\frac{N\varepsilon^2}{64 d_N M_f^2\left(1+\sum_{i>0}\beta(i)\right)}\right) + Nm\exp\left(-\frac{M\varepsilon^2}{64 d_M M_g^2\left(1+\sum_{i>0}\beta(i)\right)}\right)\right).
\end{aligned}
$$

We then have the result of Theorem 8.2.

C. Linear Quadratic Regulator

We consider the Linear Quadratic Regulator (LQR) problem over a finite time horizon T (Hambly et al., 2021; Yang et al., 2019),

$$
\min_{\{u_t\}_{t=0}^{T-1}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_t^\top Q_t x_t + u_t^\top R_t u_t\right) + x_T^\top Q_T x_T\right] \tag{A2}
$$

such that for $t = 0, 1, \cdots, T-1$,

$$
x_{t+1} = Ax_t + Bu_t + w_t,
$$

where $A \in \mathbb{R}^{d\times d}$ and $B \in \mathbb{R}^{d\times k}$ are referred to as system (transition) matrices, $Q_t \in \mathbb{R}^{d\times d}$ ($\forall t = 0, 1, \cdots, T$) and $R_t \in \mathbb{R}^{k\times k}$ ($\forall t = 0, 1, \cdots, T-1$) are matrices that parameterize the quadratic costs. $x_t \in \mathbb{R}^d$ is the state of the system with the initial state x_0. $u_t \in \mathbb{R}^p$ is the action taken at time t and $\{w_t\}_{t=1}^{T-1}$ are random noise assuming to be independent from x_0.

Let $P_t^\star$ be the solution to the discrete algebraic Riccati equation

$$
P_t^\star = Q_t + A^\top P_{t+1}^\star A - A^\top P_{t+1}^\star B\left(B^\top P_{t+1}^\star B + R_t\right)^{-1} B^\top P_{t+1}^\star A.
$$

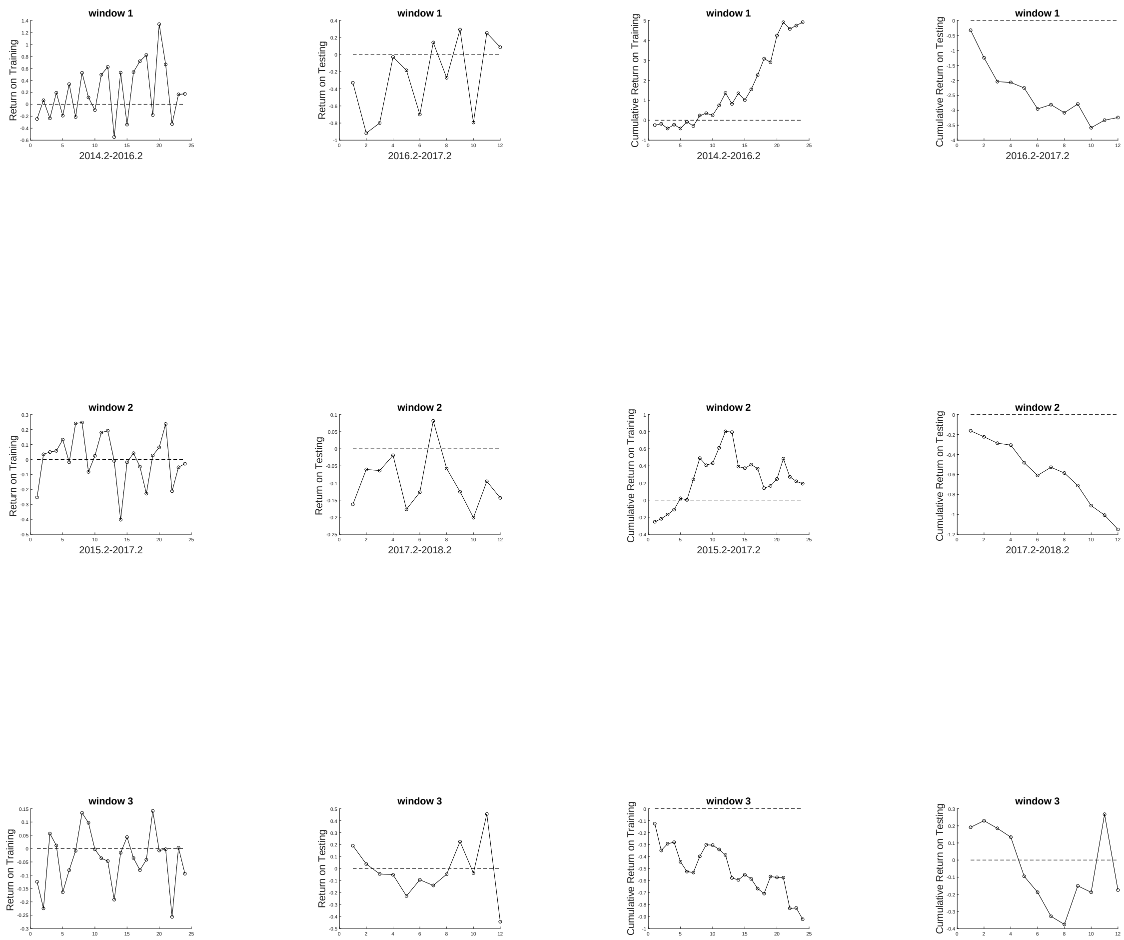
Under mild assumptions, Bertsekas (2012) shows that the optimal control sequence $\{u_t\}_{t=0}^{T-1}$ is given by

$$
u_t = -K_t^\star x_t, \qquad \text{where } K_t^\star = \left(B^\top P_{t+1}^\star B + R_t\right)^{-1} B^\top P_{t+1}^\star A.
$$

We then use policy gradient method to find an optimal policy $K_t, t = 1, \cdots, T-1$, through minimizing the loss function (A2).

D. Numerical Experiments: Risk-Averse Portfolio Optimization Problem

We predict the total return of a window $(K = 12)$ using the latest 24 months. The results of the experiment on six windows are shown as follows.



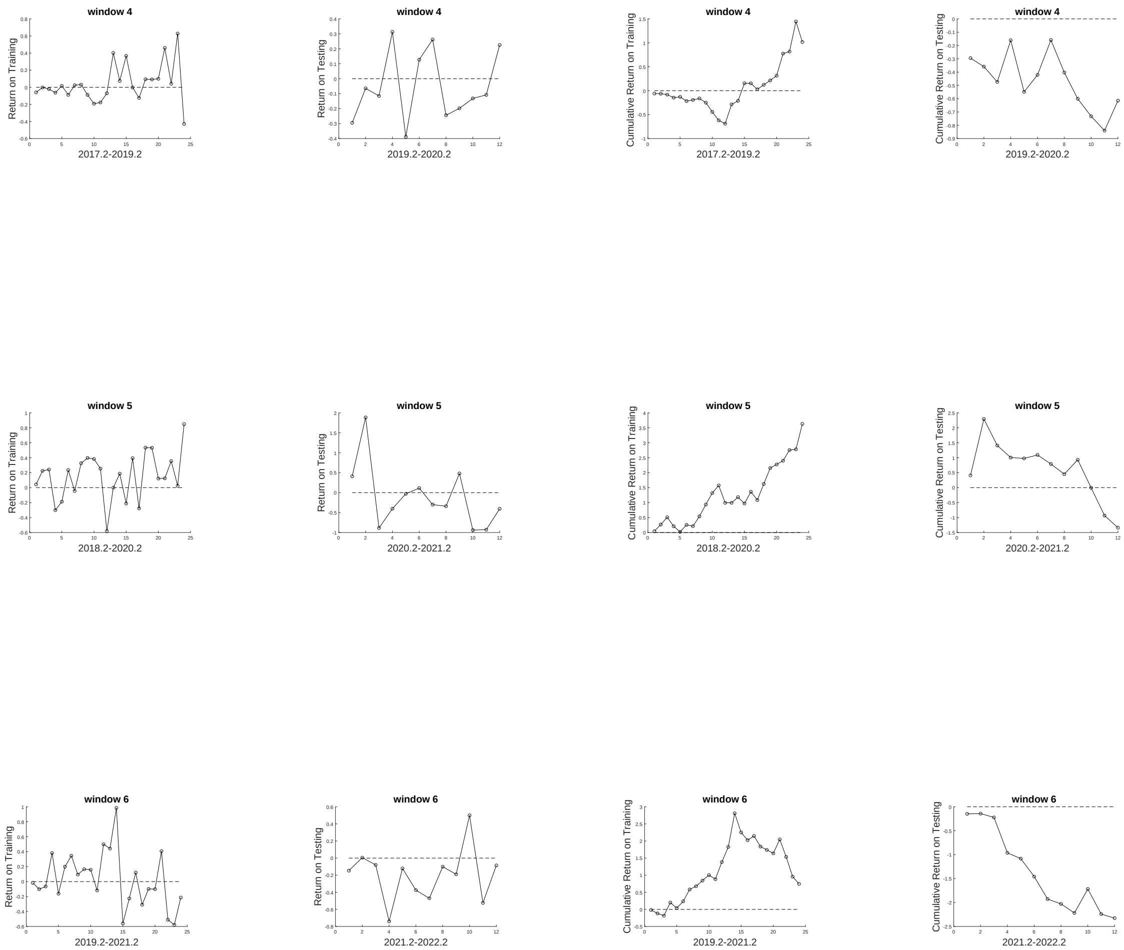


Figure 4. Cumulative return for risk-averse portfolio allocation optimization via incremental learning.