Robust Empirical Risk Minimization from a Single Dependent Trajectory: Non-Asymptotic Theory under Infinite Variance

Jianya Lu *
University of Essex
Colchester, United Kingdom
jylu2018@gmail.com

Bo Pan *
University of Alberta
Edmonton, Canada
pan1@ualberta.ca

Yafei Wang
University of Alberta
Edmonton, Canada
yafei2@ualberta.ca

Lihu Xu
Michigan State University, Michigan, United States
University of Macau, Macau, China
xulihu@msu.edu, lihuxu@um.edu.mo

Bei Jiang
University of Alberta
Edmonton, Canada
bei1@ualberta.ca

Linglong Kong †
University of Alberta
Edmonton, Canada
lkong@ualberta.ca

Abstract

We study robust empirical risk minimization from a single dependent trajectory, a central setting in dynamic risk management, adaptive control, and reinforcement learning. The key difficulty is the simultaneous presence of serial dependence and extreme-risk observations: dependence reduces the effective sample size, while heavy tails can cause standard empirical risk minimization to suffer from poor deviation control and high sensitivity to extreme observations. Existing robust learning theory largely treats heavy tails under independence and often bounded variance conditions, leaving a gap for modern sequential decision-making problems where observations are serially correlated and can exhibit infinite-variance behavior. We close this gap by introducing a logarithmically truncated estimator for heavy-tailed trajectory data. Under mixing and stationary conditions, we establish high-probability excess-risk guarantees that explicitly quantify the joint effects of temporal dependence and tail heaviness. The resulting rate demonstrates that blocking can recover marginal-tail-driven learning behavior up to dependence-induced logarithmic factors. As practical implications, we derive guarantees for downstream applications, including regression under autoregressive dependence and linear dynamical system identification. Numerical experiments show that the proposed estimator remains stable in regimes where standard methods break down.

1 Introduction

Empirical risk minimization (ERM) is a foundational framework for estimating model parameters and decision rules by minimizing an empirical approximation of the population risk. Classical formulations typically assume independent and identically distributed (i.i.d.) observations from an

*These authors contributed equally and in alphabetical order.
†Corresponding author

STAI-X 2026 (Full Paper Track). Do not distribute.

underlying distribution with regular tail conditions, such as boundedness, sub-Gaussianity. However, many modern AI and decision-making systems operate on sequential data, where observations arise from a stochastic process with high variability and are indeed inherently dependent. This paper focuses on a particularly challenging regime: robust ERM estimation from a single dependent trajectory under heavy-tailed uncertainty [Tu et al., 2024, Wang et al., 2022, Duchi et al., 2012]. Such problems arise in healthcare, finance, operational control, and reinforcement learning, where observations are collected along one evolving path and are shaped by both temporal dependence and rare high-impact events [Shi et al., 2023, Wang and Janson, 2021, Gönül et al., 2021]. This sampling structure is intrinsic to sequential data, but it poses substantial methodological and theoretical challenges.

The central difficulty is the coupling between serial dependence and tail risk [Forsyth and Vetzal, 2017, Wang et al., 2020, Roy et al., 2021]. Temporal dependence reduces the effective sample size and weakens the concentration of empirical risks relative to the i.i.d. case, making classical concentration tools for independent samples inapplicable. Rare but large losses further amplify this difficulty, since they can dominate the empirical objective and, under dependence, may appear as clusters or persistent episodes of large fluctuations along the trajectory rather than isolated outliers. Together, these effects lead to poor deviation control for standard ERM. Existing theory for sequential learning often relies on bounded observations, Gaussian noise, or finite-variance assumptions [Wang et al., 2022, Liu and Tao, 2014, Xu et al., 2020]. These assumptions may fail in realistic regimes where observations can have very large, or even infinite, variance [Xu et al., 2023, Brownlees et al., 2015, Davis et al., 1992]. In such settings, expectation-based guarantees are insufficient, and high-probability bounds obtained through crude tools such as Markov's inequality are typically too weak. Although robust optimization methods provide an alternative route [Hsu and Sabato, 2016], they often do not explicitly account for how tail risk interacts with serial dependence in trajectory data and can be computationally demanding, over-parameterized, and sensitive to tuning. This motivates robust ERM methods that provide high-probability control under dependent, infinite-variance regimes while remaining computationally simple.

Main Contribution. In this paper, we initiate a framework for robust empirical risk minimization from a single dependent trajectory under heavy-tailed uncertainty. We propose a logarithmically truncated estimator that provides a regularization of large empirical losses while retaining information from extreme samples. Our theory moves beyond bounded-loss, sub-Gaussian, and finite-variance regimes. The estimator applies broadly to ERM problems under mild Lipschitz-type regularity of the loss. Our main contributions are summarized as follows.

1. We formulate a robust ERM framework for a single temporally dependent stochastic process, allowing serial dependence, unbounded observations.
2. We develop a logarithmically truncated estimator with high-probability deviation control under temporal dependence. The truncation mechanism dampens extreme-risk observations while retaining informative tail behavior. Under suitable mixing and non-degeneracy conditions, we establish excess-risk guarantees whose rates explicitly capture the effects of temporal dependence and tail heaviness.
3. We instantiate the framework in downstream applications, including regression with autoregressive covariates and linear dynamical system identification, and validate the estimator through numerical experiments.

Related work. This paper is motivated by a line of research on learning from a single dependent trajectory, with a particular emphasis on robust estimation under unbounded and heavy-tailed observations. In this section, we review the lines of work most closely related to our framework.

Learning from the trajectory of observations. A central setting in this paper is learning from observations generated along an ordered stochastic process, where dependence arises naturally over time. Such data appear across robotics [Mason et al., 2016, Nguyen-Tuong and Peters, 2011], language modelling [Carta et al., 2023, Zhai et al., 2024], financial risk management [Forsyth and Vetzal, 2017, Hambly et al., 2021], operational control [Wang and Janson, 2021, Pan et al., 2025], and reinforcement learning [Shi et al., 2023]. In supervised learning, this often corresponds to a single dependent trajectory of covariate–response pairs, as in autoregressive modeling and linear dynamical system identification [Wong and Li, 2000, Dean et al., 2020]. In reinforcement learning, trajectories of states, actions, and rewards describe the interaction between a decision-maker and the environment. These settings depart from the i.i.d. paradigm underlying much of classical statistical learning theory, limiting the applicability of standard estimators and guarantees. This gap motivates learning methods

tailored to trajectory data, with explicit control of temporal dependence and sample efficiency. In this work, we aim to fill this gap by focusing on the stationary dependent-trajectory regime, where observations are generated from a fixed stochastic process satisfying suitable mixing conditions. This regime covers a broad class of time-series, autoregressive, and stable dynamical-system problems, and is also relevant to policy-fixed settings in reinforcement learning and control, such as fixed-policy evaluation, off-policy evaluation under a fixed behavior policy, and closed-loop system identification under a prescribed stabilizing controller. In these settings, the policy or controller is fixed during data collection, so the resulting trajectory can often be analyzed as a stationary dependent process under appropriate stability and mixing assumptions.

Trajectory data under heavy-tailed distributions. Existing work on trajectory data often assumes stationarity and stability through boundedness, finite variance, or Gaussian-type conditions. These assumptions ensure convergence to a target distribution, making the learning problem well defined, and provide the concentration guarantee needed along the trajectory. However, they may fail in sequential systems that visit extreme states, generating heavy-tailed losses and destabilizing ERM. Pareto distributions, for example, are unbounded and possess finite moments only below their shape parameter, making them canonical models for extreme events [He et al., 2022]. In such regimes, expectation-based guarantees can be inadequate because the corresponding high-probability bounds may become uninformative when relevant moments are infinite.

Learning under infinite variance has therefore received growing attention. Kanter and Steiger [1974] study regression and time-series estimation with infinite-variance variables, showing that least squares can fail in general heavy-tailed linear models but may remain consistent in certain autoregressive and moving-average settings under stable-law assumptions. More recently, Xu et al. [2023] developed log-truncated robust M-estimators for infinite-variance learning, covering convex regression and high-dimensional nonconvex models such as deep neural networks. de la Peña et al. [2025] propose a transformation-based predictor for random variables with infinite mean or variance and establish its statistical properties and confidence intervals. These works are closely related in their focus on heavy-tailed and infinite-variance regimes, but they do not provide a general robust learning framework for trajectory data where temporal dependence is central. Existing results are typically tailored to i.i.d. samples, specific autoregressive models, or robust reinforcement learning. In contrast, our framework provides high-probability guarantees for broad empirical risk minimization problems under temporal dependence, yielding an estimator suitable for robust learning from dependent trajectories.

2 Problem statement and preliminaries

We consider learning from a single dependent trajectory. Let ξ be a random vector taking values in Ξ with marginal distribution P, and let $\ell : \Theta \times \Xi \to \mathbb{R}$ be a pre-specified loss function. The population risk minimization problem is

$$\theta_\ell^\star := \arg\min_{\theta\in\Theta} \left\{ \mathbb{E}[\ell(\theta;\xi)] = \int_\Xi \ell(\theta;\xi) P(\mathrm{d}\xi) \right\}, \tag{1}$$

where $\Theta \subset \mathbb{R}^p$ is the parameter space and $\Theta^\star$ is the collection of all minimizers $\theta_\ell^\star$. The loss ℓ is assumed to be closed, convex, and proper in θ, but not necessarily differentiable. Let $\widehat{\Xi}_T = \{\widehat{\xi}_t\}_{t=1}^T$ denote the observed training data, viewed as a finite realization of a stationary stochastic process with marginal distribution P. The standard ERM replaces P by the empirical distribution $\widehat{P}_T$ and solves

$$\widehat{\theta}_{\ell,T}^\star := \arg\min_{\theta\in\Theta} \left\{ \widehat{R}_{\ell,T}(\theta) := \int_\Xi \ell(\theta;\xi)\, \widehat{P}_T(\mathrm{d}\xi) = \frac{1}{T}\sum_{t=1}^T \ell\left(\theta;\widehat{\xi}_t\right) \right\}. \tag{2}$$

Throughout, quantities with hats depend on the observed trajectory. To study robust learning under dependence and heavy tails, we impose the following assumptions.

Assumption 1 (Convexity and compactness)**.** *The parameter space $\Theta \subset \mathbb{R}^p$ is convex and compact. In particular, there exists $r > 0$ such that $\|\theta\| \le r$ for all $\theta \in \Theta$.*

Assumption 2 (Local Lipschitz)**.** *For any $\xi \in \mathbb{R}^d$, the loss function $\ell(\theta, \xi)$ is Lipschitz continuous in θ on Θ. That is, for any $\theta_1, \theta_2 \in \Theta$, there exists a constant H_ξ depending on ξ such that $|\ell(\theta_1;\xi) - \ell(\theta_2;\xi)| \le H_\xi \|\theta_1 - \theta_2\|$.*

Assumption 3 (Exponential β-mixing)**.** *The process $\{\xi_t\}_{t\geq 1}$ is strictly stationary with marginal distribution P. For $t \geq 1$, let $\mathcal{F}_1^t := \sigma(\xi_1, \ldots, \xi_t)$, and $\mathcal{F}_{t+k}^{\infty} := \sigma(\xi_{t+k}, \xi_{t+k+1}, \ldots)$. The β-mixing coefficient at lag k is defined as*

$$\beta(k) := \sup_{t\geq 1} \mathbb{E}\left[\sup_{B\in\mathcal{F}_{t+k}^{\infty}} \left|\mathbb{P}\left(B \mid \mathcal{F}_1^t\right) - \mathbb{P}(B)\right|\right],$$

where B ranges over all events determined by the future observations $\{\xi_{t+k}, \xi_{t+k+1}, \ldots\}$. We assume that the process is exponentially β-mixing: there exist constants $C_\beta > 0$ and $c_\beta > 0$ such that $\beta(k) \leq C_\beta e^{-c_\beta k}, k \in \mathbb{N}$.

Remark 1 (β-mixing and effective independence)**.** *The β-mixing coefficient quantifies the dependence between the past $\mathcal{F}_1^t$ and the future $\mathcal{F}_{t+k}^{\infty}$. Exponential decay means that observations far apart in time become approximately independent. This enables blocking arguments, where a dependent trajectory is decomposed into blocks that behave nearly independently up to a controllable error. Thus, Assumption 3 allows temporal dependence while retaining enough weak-independence structure for high-probability learning guarantees.*

Beyond temporal dependence, we allow the observations to be heavy-tailed. When the data are heavy-tailed, the empirical estimator $\widehat{\theta}_{\ell,T}^{\star}$ in (2) can be unstable. Its realized behavior may deviate substantially from its population target, especially when observations are unbounded and only low-order moments exist.

In the following section, we introduce logarithmically truncated estimators for a broad class of learning tasks and establish excess-risk bounds under this infinite-variance, dependent-data regime.

3 Logarithmically truncated loss and estimator

Let $\alpha > 0$ be a robustification parameter and $\rho > 0$ an ℓ_2-regularization parameter. We define the logarithmically truncated M-estimator as

$$\widehat{\theta}_{\psi_\lambda,\ell}^{\star} := \arg\min_{\theta\in\Theta} \left\{\widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta) + \rho\|\theta\|^2\right\}, \text{ with } \widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta) := \frac{1}{T\alpha}\sum_{t=1}^{T} \psi_\lambda\left(\alpha\ell\left(\theta; \widehat{\xi}_t\right)\right). \quad (3)$$

The truncation function ψ_λ satisfies, for a pre-specified nonnegative function λ,

$$-\log[1 - x + \lambda(|x|)] \leq \psi_\lambda(x) \leq \log[1 + x + \lambda(|x|)], \quad \forall x \in \mathbb{R}. \quad (4)$$

The choice of λ controls the influence of extreme losses while retaining information from heavy-tailed observations, making the framework adaptable to different tail behaviours. Let $R_\ell(\theta) := \mathbb{E}[\ell(\theta; \xi)]$. Given $\widehat{\theta}_{\psi_\lambda,\ell}^{\star}$, we measure statistical performance through the excess risk, defined as $R_\ell\left(\widehat{\theta}_{\psi_\lambda,\ell}^{\star}\right) - \min_{\theta\in\Theta} R_\ell(\theta)$.

The logarithmic truncation form in (4) follows the Catoni-type robustification principle and is closely related to existing robust procedures for heavy-tailed learning Catoni [2012], Fan et al. [2017], Chen et al. [2021], Xu et al. [2023]. In particular, Xu et al. [2023] studied Catoni-type and log-truncated M-estimators under infinite variance using a general control function λ. Building on this line of work, the present paper considers log-truncated robust empirical risk minimization for a single temporally dependent trajectory with heavy-tailed, possibly infinite-variance, observations. The dependent single-trajectory setting differs from the i.i.d. setting because serial dependence prevents a direct factorization of exponential-moment bounds, and hence requires blocking and coupling arguments under β-mixing dependence.

Assumption 4 (Truncation-control regularity)**.** *The function $\lambda(x) : \mathbb{R}^+ \to \mathbb{R}^+$ is a continuous non-decreasing function such that $\lim_{x\to\infty} \lambda(x)/x = \infty$. Moreover, there exist positive constant $c_1 > 0$ and a function $f(x) : \mathbb{R}^+ \to \mathbb{R}^+$ such that*

$$\lambda(\nu x) \leq f(\nu)\lambda(x) \text{ for any } \nu, x \in \mathbb{R}^+, \text{ and } \lim_{x\to 0^+} \tfrac{f(x)}{x} = 0,$$
$$\lambda(x + y) \leq c_1(\lambda(x) + \lambda(y)) \text{ for any } x, y \in \mathbb{R}^+.$$

Assumption 5 (Boundness of truncated risk)**.** *We assume that the parameter H_ξ in Assumption 2 satisfies $\mathbb{E}[H_\xi] < \infty$, $\mathbb{E}[\lambda(H_\xi)] < \infty$. Additionally, $R_{\lambda\circ\ell}(\Theta) = \sup_{\theta\in\Theta} R_{\lambda\circ\ell}(\theta) < \infty$, where $R_{\lambda\circ\ell}(\theta) = \mathbb{E}[\lambda(\ell(\theta;\xi))]$.*

Assumption 4 imposes structural conditions on the control function λ used in the logarithmic truncation envelope. The function λ regulates the degree to which large losses are damped. Its growth rate may be chosen to reflect available moment information or anticipated tail behavior of the underlying data, while remaining compatible with the integrability requirements on $\lambda(H_\xi)$ and $\lambda(\ell(\theta;\xi))$ imposed in Assumption 5. The scaling and sub-additivity-type conditions control the effects of rescaling and summation in the concentration arguments.

Assumption 5 requires integrability of the random Lipschitz coefficient and of the λ-transformed loss. The condition $\mathbb{E}[\lambda(H_\xi)] < \infty$ can be interpreted as a moment or tail condition on the data-dependent Lipschitz envelope. For instance, if $\lambda(x) = |x|^\gamma/\gamma$ with $\gamma \in (1,2)$, then $\mathbb{E}[\lambda(H_\xi)] < \infty$ corresponds to a finite γ-th moment condition on H_ξ. Together with the uniform condition $R_{\lambda\circ\ell}(\Theta) < \infty$, these assumptions ensure that the robust empirical criterion is well defined and stable, without requiring the original loss to have finite variance. Potential choices of $\lambda(x)$ are summarized in Table 2.

Remark 2. *The condition $R_{\lambda\circ\ell}(\Theta) < \infty$ requires only the truncated loss $\lambda(\ell(\theta;\xi))$ to have finite expectation uniformly over Θ; it does not require the original loss to have finite variance. If the loss grows too quickly relative to λ, this condition may fail, in which case additional robustification, such as explicit truncation or Huber-type losses, may be needed.*

3.1 Main Result

In this subsection, we state our main results. We establish excess-risk guarantees whose rates explicitly capture the effects of temporal dependence and the tail behavior of the loss distribution.

Theorem 1 (Excess risk bound)**.** *Suppose Assumptions 1–5 hold. Let $\mathcal{N}(\Theta,\kappa)$ denote a κ-net of the parameter space Θ with covering number $N(\Theta,\kappa)$. Let $\|\Theta^\star\| := \inf_{\theta\in\Theta^\star}\|\theta\|$. Then for any confidence level $\delta \in (0,1)$ and any scaling parameter $\alpha > 0$, with probability at least $1-\delta$, we have*

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) \le \underbrace{\frac{(1+c_1)f(\alpha)}{\alpha}R_{\lambda\circ\ell}(\Theta)}_{\textit{Truncation bias}} + \underbrace{\frac{\mathcal{L}(1)+\mathcal{L}(N(\Theta,\kappa))}{T\alpha c_\beta}}_{\textit{Complexity}} + \underbrace{2\kappa\mathbb{E}[H_\xi]}_{\textit{covering-net approximation}}$$
$$+ \underbrace{\frac{c_1 f(\alpha\kappa)}{\alpha}\mathbb{E}[\lambda(H_\xi)]}_{\textit{Lipschitz-envelope tail control}} + \underbrace{\rho\|\Theta^\star\|^2}_{\textit{regularization bias}}.$$

where $\mathcal{L}(\cdot)$ is the logarithmic complexity term defined as below that captures both the confidence level δ and the complexity of the parameter space through its covering number $\mathcal{L}(x) = \log\left(\frac{2C_\beta Tx}{\delta}\right)\log\left(\frac{2x}{\delta c_\beta}\log\left(\frac{2C_\beta Tx}{\delta}\right)\right)$.

Theorem 1 provides a finite-sample excess-risk guarantee for the proposed log-truncated estimator under temporally dependent observations. The bound separates the excess risk into five components: truncation bias, complexity, covering-net approximation error, Lipschitz-envelope tail control, and regularization bias. The sample size T improves the bound through the complexity term, which scales as $1/(T\alpha)$, and through the discretization term when the covering radius is chosen as $\kappa = 1/T$. Compared with the i.i.d. result of Xu et al. [2023], temporal dependence enters our bound through the logarithmic complexity factor $\mathcal{L}(\cdot)$ and the mixing-rate constant c_β. These terms quantify the cost of controlling deviations along a β-mixing trajectory rather than independent samples: faster mixing yields a smaller penalty, while stronger dependence increases the effective complexity of the bound. In particular, let $\kappa = 1/T$ make the covering-net error $O\left(T^{-1}\right)$. Balancing the truncation-bias and complexity terms yields

$$\alpha = f^{-1}\left(\frac{\mathcal{L}(1)+\mathcal{L}\left(N(\Theta,1/T)\right)}{Tc_\beta(1+c_1)R_{\lambda\circ\ell}(\Theta)}\right). \tag{5}$$

Substituting this choice into Theorem 1, we obtain

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) \le \frac{2\left\{\mathcal{L}(1)+\mathcal{L}\left(N(\Theta,1/T)\right)\right\}}{T\alpha c_\beta} + \frac{2}{T}\mathbb{E}[H_\xi] + \frac{c_1}{T}\frac{f(\alpha/T)}{\alpha/T}\mathbb{E}[\lambda(H_\xi)] + \rho\|\Theta^\star\|^2.$$

We emphasize that Theorem 1 is formulated for a general control function λ satisfying Assumptions 4 and 5. Unlike analyses that require monotonicity of the truncation map ψ_λ [Xu et al., 2023], our argument is based on the two-sided logarithmic envelope condition in (4) together with the regularity conditions imposed on λ. This relaxes the monotonicity requirement on ψ_λ and provides additional flexibility in constructing the truncation function. In practical implementations, λ may also be selected in a data-adaptive manner as part of the logarithmic truncation scheme. In this case, the data-adaptive procedure must be accompanied by additional regularity conditions ensuring that the selected λ belongs to an admissible class satisfying Assumptions 4 and 5. Such conditions may involve moment or tail restrictions, boundedness of the admissible tuning class, and concentration guarantees that remain valid under the β-mixing dependence structure in Assumption 3.

Remark 3 (Trade-off induced by α)**.** *The truncation parameter α balances truncation bias and statistical complexity in Theorem 1. The statistical-complexity term decreases with α through the factor $1/\alpha$, whereas the truncation-bias term may increase through $f(\alpha)/\alpha$. Thus, α controls the trade-off between statistical accuracy and robustness to extreme losses. Balancing these two leading contributions gives the optimized choice of α used in the rate analysis.*

We next consider the special case $\lambda(x) = |x|^\gamma/\gamma$, $\gamma \in (1, 2)$, which satisfies Assumptions 4 and 5 and is commonly used in practice.

Corollary 1 (Heavy-tail-dependent excess-risk rate)**.** *Let $\lambda(x) = |x|^\gamma/\gamma$ with $\gamma \in (1, 2)$. Suppose Assumptions 1 - 5 hold, and choose the scaling parameter α as in* (18)*. Then, for any confidence level $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) \le pT^{-\frac{\gamma-1}{\gamma}} \log\left(\frac{2C_\beta T(1+2rT)}{\delta}\right)\left(4C_{\lambda,\ell,\gamma} + 4\mathbb{E}[\lambda(H_\xi)]C^{\gamma-1}_{\lambda,\ell,\gamma}\right) c_\beta^{\frac{1}{\gamma}-1} + \frac{2}{T}\mathbb{E}[H_\xi] + \rho\|\Theta^\star\|^2,$$

where $C_{\lambda,\ell,\gamma} > 0$ is a constant. Consequently,

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) = \mathcal{O}\left(pT^{-\frac{\gamma-1}{\gamma}} \log\left(\frac{2C_\beta T(1+2rT)}{\delta}\right) + \rho\|\Theta^\star\|^2\right).$$

Corollary 1 makes explicit how the tail parameter controls the convergence rate. The leading stochastic term scales as $pT^{-(\gamma-1)/\gamma}\log T$, up to confidence, dependence, and constant factors. Smaller values of γ correspond to heavier tails and slower convergence. As $\gamma \to 2$, the rate approaches the near-parametric order $pT^{-1/2}\log T$, up to logarithmic, dependence-related, and regularization factors. Figure 2 provides a visual illustration of the impact of γ on the strength of logarithmic truncation.

Remark 4. *In this work, we focus on logarithmically truncated loss with L_2 penalty, which is mainly due to the Assumption 1. Assumption 1 is naturally guaranteed by L_2 regularization. We highlight that Theorem 1 and Corollary 1 are more general and also apply to the logarithmically truncated loss in* (3) *with an L_1 penalty. In that case, only the regularization bias changes to $\rho\|\Theta^\star\|_1$, while the remaining terms remain unchanged.*

3.2 Examples

This section demonstrates the proposed framework in two common trajectory-data settings: regression with autoregressive covariates and closed-loop system identification. In both examples, observations arise from a dependent stochastic trajectory, making the i.i.d. framework inappropriate. Further details are provided in Appendix A.

Regression with autoregressive covariates [Duchi et al., 2012]. We first consider regression with covariates generated by an autoregressive process. This setting provides a canonical example of trajectory data: the covariates are not independently sampled, but evolve through a stochastic recursion. Such models arise in time-series analysis, financial modelling, network traffic, and control systems, where current observations depend on past states and may be affected by rare but extreme shocks. Consider the autoregressive model

$$X_t = AX_{t-1} + \Sigma W_t, \qquad Y_t = g(\theta^\star; X_t) + \varepsilon_t, \qquad t = 1, \dots, T, \tag{6}$$

where A is a transition matrix with spectral radius strictly less than one, Σ is a fixed matrix, and $\{W_t\}$ and $\{\varepsilon_t\}$ denote innovation and noise sequences, the variance of $\{\varepsilon_t\}$ does not exist. The parameter

of interest is $\theta^\star \in \Theta$, where $\Theta = \{\theta \in \mathbb{R}^p : \|\theta\| \le R\}$. For $\xi_t = (Y_t, X_t)$, with a pre-specified loss $\ell(\theta; \xi)$, we estimate $\theta^\star$ by the log-truncated estimator as (3). We note that, when the maximum eigenvalue of the transition matrix A is strictly less than 1, i.e., $\rho(A) < 1$, the autoregressive covariate process is stable under standard innovation conditions and admits a stationary distribution. Hence, the resulting observations form a dependent trajectory rather than an i.i.d. sample.

Corollary 2 (Regression with autoregressive covariates)**.** *Assume that the parameter space, loss function, and truncation function satisfy Assumptions 1,2, 4 and 5. Then, for any $\delta \in (0,1)$, $\alpha > 0$, and $\kappa > 0$, with probability at least $1 - \delta$,*

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) \le \frac{(1+c_1)f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\Theta) + \frac{\mathcal{L}(1) + \mathcal{L}(N(\Theta,\kappa))}{T\alpha c_\beta} + 2\kappa\,\mathbb{E}[H_\xi] + \frac{c_1 f(\alpha\kappa)}{\alpha}\mathbb{E}[\lambda(H_\xi)] + \rho\|\Theta^\star\|^2.$$

This corollary illustrates the broad utility of the proposed framework by covering a wide class of Lipschitz loss-based regression problems with autoregressive covariates, including quantile regression, additive models, neural network regression under suitable Lipschitz and moment controls, and certain quasi-likelihood or generalized linear models such as logistic regression. Hence, the guarantee is not tied to a single regression model, but applies broadly to robust learning from dependent covariate trajectories under heavy-tailed noise.

Remark 5. *For the linear model $g(\theta, \xi) = \langle \theta, \xi \rangle$, the assumptions become simpler. The model is automatically Lipschitz in θ, with Lipschitz constant depending only on $\|\xi\|$. Therefore, instead of imposing an abstract Lipschitz condition, it is enough to assume suitable finite moments of the covariates and the noise.*

Learning with linear dynamical trajectories. We next consider system identification from closed-loop trajectory data. Let $A \in \mathbb{R}^{n\times n}$ denote the transition matrix and $B \in \mathbb{R}^{n\times m}$ the control matrix. The system evolves according to

$$x_{t+1} = Ax_t + Bu_t + \varepsilon_t, \tag{7}$$

where $x_t \in \mathbb{R}^n$ is the system state, $u_t \in \mathbb{R}^m$ is the control input, and $\varepsilon_t \in \mathbb{R}^n$ is the additive system noise. Such models arise widely in model-based reinforcement learning, dynamical system identification, and operational control.

We focus on the closed-loop setting, where the control input is generated by a known deterministic feedback law $u_t = \pi(x_t)$. Given trajectory data $\{\xi_t\}_{t=0}^{T-1}$ with $\xi_t = (x_t, u_t, x_{t+1})$, we define the absolute prediction loss $\ell((A,B);\xi_t) = \|x_{t+1} - Ax_t - Bu_t\|_1$. The corresponding penalized estimator is

$$\left(\widehat{A}, \widehat{B}\right) = \arg\min_{(A,B)\in\Theta_{AB}} \left\{\widehat{R}_{\psi_\lambda,\ell,\alpha}([\mathrm{vec}(A)^\top, \mathrm{vec}(B)^\top]^\top) + \rho\left(\|A\|_F^2 + \|B\|_F^2\right)\right\}.$$

Unlike the i.i.d. regression setting, the observations $\{\xi_t\}_{t=0}^{T-1}$ are generated along a single closed-loop trajectory and are therefore serially dependent. This makes system identification a natural application of the proposed framework for robust learning from dependent trajectory data.

Corollary 3 (Learning with linear dynamical trajectories)**.** *Suppose the regression noise $\{\varepsilon_i\}$ is i.i.d. with finite γ-th moment for some $\gamma \in (1,2)$. Assume the closed-loop matrix is stable. Let $\lambda(x) = |x|^\gamma/\gamma$ and Assumption 1 hold. Then, for any $\delta \in (0,1)$, $\alpha > 0$, and $\kappa > 0$, with probability at least $1 - \delta$,*

$$\begin{aligned} &R_\ell\left([vec(\widehat{A})^\top, vec(\widehat{B})^\top]^\top\right) - R_\ell([vec(A)^\top, vec(B)^\top]^\top) \\ \le&(n^2+mn)T^{-\frac{\gamma-1}{\gamma}} \log\left(\frac{2C_\beta T(1+2rT)}{\delta}\right)\left(4C_{\lambda,\ell,\gamma} + 4\mathbb{E}[\lambda(H_\xi)]C_{\lambda,\ell,\gamma}^{\gamma-1}\right)c_\beta^{\frac{1}{\gamma}-1} + \frac{2}{T}\mathbb{E}[H_\xi] + \rho\|\Theta^\star\|^2, \end{aligned}$$

where $C_{\lambda,\ell,\gamma}$ is a positive constant and H_ξ is the local Lipschitz constant bounded by $C_K\|x_t\|$.

Remark 6. *Apart from the above application, a broad class of data-generating processes can be viewed as trajectory data. Examples include time-series, Markov chains, longitudinal measurements, and path-tracking trajectories, where observations are collected along an evolving stochastic path. This viewpoint provides a unified framework for studying learning problems driven by Trajectory observations.*

4 Simulation studies

We evaluate the performance of the proposed estimator. The goal of the simulation study is to examine the qualitative behavior of the proposed estimator under controlled designs that reflect the main theoretical challenges considered in this paper: temporal dependence and heavy-tailedness. We therefore use autoregressive data-generating mechanisms with Pareto-type errors, which allow the dependence strength, tail parameter, and sample size to be varied systematically. These experiments are designed to validate the theoretical predictions and to compare the stability of the proposed estimator with standard baselines under dependent heavy-tailed regimes. Herein, we mainly focus on the dynamic mean estimation and regression tasks across several settings. The detailed simulation settings and the regression and logistic regression experiments are provided in Appendix B.

Dynamic mean estimation. In this experiment, the data $\{\xi_t\}_{t=1}^T$ is generated from the autoregressive process $\xi_t = \rho\xi_{t-1} + \varepsilon_t$, where ε_t denotes the random error. We compare the log-truncated estimator (LTE) in (3) with two baselines: vanilla empirical risk minimization (ERM) and the Cauchy truncated estimator (Cauchy). We set $\psi_\lambda(x) = \left(\log\left(1 + x + |x|^\gamma/\gamma\right) - \log\left(1 - x + |x|^\gamma/\gamma\right)\right)/2$ for the LTE method throughout our analysis. The ERM estimator is given by the sample average, while the loss function for the Cauchy method is defined as:

$$\psi(x) = \begin{cases} \log\left(1 + (x/q)^2\right), & |x| \le \epsilon, \\ \log\left(1 + (\epsilon/q)^2\right), & |x| > \epsilon, \end{cases}$$

where q is the scale parameter and ϵ is the truncation threshold. All tuning parameters were selected through an independent pilot study conducted under the same data-generating design. A grid of candidate values was evaluated using the same performance criterion as in the main simulation study, and the selected values were then fixed for the final comparison. We use mean squared error (MSE) to compare performance across methods. The results are summarized in Table 1.

Table 1: MSE comparison across (T, τ, ρ) settings with $\gamma = 1.5$

		$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
T	τ	LTE	Cauchy	ERM	LTE	Cauchy	ERM	LTE	Cauchy	ERM
50	1.3	3.4827	8.1694	41.4436	8.4531	13.6388	87.6146	46.9136	36.9801	290.2720
	1.6	0.5109	1.2100	2.6790	1.0567	1.7920	5.7454	4.5039	4.6055	19.8040
	2.01	0.1061	0.2712	0.2967	0.2397	0.3751	0.6444	0.9699	0.9819	2.3033
	4.01	0.0064	0.0131	0.0077	0.0139	0.0180	0.0169	0.0528	0.0590	0.0629
100	1.3	2.4138	8.2387	26.6822	5.5460	13.8300	59.8725	22.9623	36.0199	237.4384
	1.6	0.2780	1.1972	1.4302	0.6423	1.7348	3.2053	2.6341	3.8962	12.6642
	2.01	0.0607	0.2592	0.1433	0.1401	0.3386	0.3208	0.5632	0.7213	1.2621
	4.01	0.0031	0.0112	0.0037	0.0071	0.0133	0.0082	0.0290	0.0348	0.0321
500	1.3	1.6972	8.3669	10.1423	3.0693	14.2236	22.8166	11.8340	37.9911	91.1809
	1.6	0.1086	1.2035	0.3665	0.2496	1.7330	0.8243	1.0240	3.6767	3.2914
	2.01	0.0191	0.2567	0.0306	0.0403	0.3206	0.0688	0.1628	0.5684	0.2743
	4.01	0.0009	0.0101	0.0008	0.0017	0.0100	0.0018	0.0068	0.0147	0.0070

Table 1 demonstrates that the proposed LTE generally achieves superior performance, attaining the lowest MSE compared with the two baseline methods. The benefit is especially pronounced when the errors are heavy-tailed. Notably, smaller values of τ correspond to heavier-tailed errors; in these settings, ERM is highly sensitive to extreme observations, while the Cauchy truncated estimator can still exhibit relatively large errors. As the trajectory length T increases, the MSE of LTE decreases steadily, confirming the benefit of larger sample sizes even under temporal dependence. As expected, dependence deteriorates the performance: larger ρ leads to larger MSEs for all methods, reflecting the loss of effective sample size under stronger serial correlation. When $\tau = 4.01$, the error distribution is relatively light-tailed, and all methods perform similarly. These results demonstrate that LTE remains stable under both heavy tails and stronger dependence, while retaining competitive performance in lighter-tailed settings.

We further illustrate the role of the order γ in estimation performance. Note that larger values of γ indicate that more information is retained by the logarithmically truncated M-estimator. As shown in Figure 1, for a fixed sample size, the corresponding MSE decreases as γ increase.

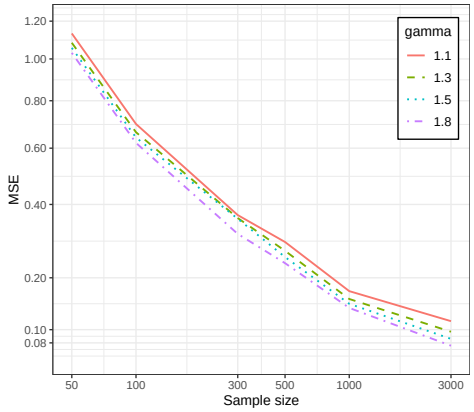

Figure 1: MSE for different γ over various T for Pareto error with $\rho = 0.5$, $\tau = 1.6$

5 Conclusion and discussion

We study robust learning from a single temporally dependent trajectory under an infinite-variance setting and propose a log-truncated estimator for stabilizing empirical risk minimization. The theory quantifies how tail heaviness and temporal dependence jointly affect excess risk, showing that near-classical behavior can still be recovered under suitable mixing and moment conditions. Simulations further support the theory: LTE is most effective under heavier tails, remains competitive in lighter-tailed settings, improves with longer trajectories, and is affected by stronger dependence through reduced effective sample size. For both continuous and binary regression tasks, LTE provides more stable performance than ERM under contaminated covariates.

Several limitations motivate future work. First, the theoretical guarantees depend on unknown problem-specific quantities, such as tail and dependence parameters, which limits their direct use for practical tuning. Developing fully data-adaptive tuning procedures is therefore an important next step. Second, the present theory is developed for a single stationary dependent trajectory satisfying suitable mixing and moment conditions. Consequently, the framework is most directly applicable to stationary dependent time-series settings, including policy-evaluation and fixed-policy learning problems. These assumptions may be restrictive for fully adaptive reinforcement-learning or control environments, where the learned policy may change the data-generating process and stationarity may fail. Extending the analysis to nonstationary, non-mixing, or policy-adaptive data-generating mechanisms remains an important direction for future work. Third, the numerical experiments are primarily theory-guided and are limited to controlled autoregressive and regression settings with Pareto-type heavy-tailed behavior. These designs allow us to systematically study the effects of dependence strength, tail heaviness, and sample size, but they do not fully capture the complexity of real-world trajectory data. In practice, data may exhibit nonstationarity, model misspecification, adaptive sampling, high-dimensional structure, and unknown dependence or tail behavior. Incorporating real-data applications and more complex trajectory-learning benchmarks is an important direction for future work.

Further extensions to modern learning architectures and robust optimization/control applications, such as online optimization and federated learning, would broaden the practical relevance of the framework. In addition, this work focuses on excess-risk guarantees rather than inference. Confidence intervals, hypothesis testing, and uncertainty quantification for robust estimators under heavy-tailed dependence remain open problems. Since classical inference often relies on variance-based quantities and central limit approximations, understanding whether stable-law exists and how to develop valid inferential tools in infinite-variance regimes is an important future direction.

Acknowledgments. Jianya Lu was supported by the London Mathematical Society (LMS) Research in Pairs scheme (Grant No. 42445); Yafei Wang was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC); 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.

References

Henry Berbee. Convergence rates in the strong law for bounded mixing sequences. *Probability Theory and Related Fields*, 74(2):255–270, 1987.

Christian Brownlees, Emilien Joly, and Gábor Lugosi. Empirical risk minimization for heavy-tailed losses. 2015.

Thomas Carta, Clément Romac, Thomas Wolf, Sylvain Lamprier, Olivier Sigaud, and Pierre-Yves Oudeyer. Grounding large language models in interactive environments with online reinforcement learning. In *International Conference on Machine Learning*, pages 3676–3713. PMLR, 2023.

Olivier Catoni. Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et Statistiques*, volume 48, pages 1148–1185, 2012.

Peng Chen, Xinghu Jin, Xiang Li, and Lihu Xu. A generalized catoni's m-estimator under finite α-th moment assumption with $\alpha(1, 2)$. *Electronic Journal of Statistics*, 15(2):5523–5544, 2021.

Richard A Davis, Keith Knight, and Jian Liu. M-estimation for autoregressions with infinite variance. *Stochastic Processes and Their Applications*, 40(1):145–180, 1992.

Victor de la Peña, Henryk Gzyl, Silvia Mayoral, Haolin Zou, and Demissie Alemayehu. Prediction and estimation of random variables with infinite mean or variance. *Communications in Statistics-Theory and Methods*, 54(1):115–129, 2025.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, 20(4):633–679, 2020.

John C Duchi, Alekh Agarwal, Mikael Johansson, and Michael I Jordan. Ergodic mirror descent. *SIAM Journal on Optimization*, 22(4):1549–1578, 2012.

Jianqing Fan, Quefeng Li, and Yuyan Wang. Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(1):247–265, 2017.

Peter A Forsyth and Kenneth R Vetzal. Dynamic mean variance asset allocation: Tests for robustness. *International Journal of Financial Engineering*, 4(02n03):1750021, 2017.

László Gerencsér and Zs Vágó. Adaptive control of multivariable linear stochastic systems. a strong approximation approach. In *1999 European Control Conference*, pages 1643–1647. IEEE, 1999.

Suat Gönül, Tuncay Namlı, Ahmet Coşar, and İsmail Hakkı Toroslu. A reinforcement learning based algorithm for personalization of digital, just-in-time, adaptive interventions. *Artificial Intelligence in Medicine*, 115:102062, 2021.

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

Yi He, Liang Peng, Dabao Zhang, and Zifeng Zhao. Risk analysis via generalized pareto distributions. *Journal of Business & Economic Statistics*, 40(2):852–867, 2022.

Daniel Hsu and Sivan Sabato. Loss minimization and parameter estimation with heavy tails. *Journal of Machine Learning Research*, 17(18):1–40, 2016.

Marek Kanter and WL Steiger. Regression and autoregression with infinite variance. *Advances in Applied Probability*, 6(4):768–783, 1974.

PR Kumar and Pravin Varaiya. Stochastic systems: estimation, identification and adaptive control (prentice-hall information & system sciences series). 1986.

Tongliang Liu and Dacheng Tao. On the robustness and generalization of cauchy regression. In *2014 4th IEEE International Conference on Information Science and Technology*, pages 100–105. IEEE, 2014.

Jianya Lu, Wei Biao Wu, Zhijie Xiao, and Lihu Xu. Almost sure invariance principle of $\beta-$ mixing time series in hilbert space. *arXiv preprint arXiv:2209.12535*, 2022.

Giorgos Mamakoukas, Orest Xherija, and Todd Murphey. Memory-efficient learning of stable linear dynamical systems for prediction and control. *Advances in Neural Information Processing Systems*, 33:13527–13538, 2020.

Sean Mason, Nicholas Rotella, Stefan Schaal, and Ludovic Righetti. Balancing and walking using full dynamics lqr control with contact constraints. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pages 63–68. IEEE, 2016.

Stanislav Minsker. Sub-gaussian estimators of the mean of a random matrix with heavy-tailed entries. *The Annals of Statistics*, 46(6A):2871–2903, 2018.

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

Duy Nguyen-Tuong and Jan Peters. Model learning for robot control: a survey. *Cognitive Processing*, 12(4):319–340, 2011.

Bo Pan, Jianya Lu, Yafei Wang, Hao Li, Bei Jiang, and Linglong Kong. Toward optimal statistical inference in noisy linear quadratic reinforcement learning over a finite horizon. *arXiv preprint arXiv:2508.08436*, 2025.

Abhishek Roy, Krishnakumar Balasubramanian, and Murat A Erdogdu. On empirical risk minimization with dependent and heavy-tailed data. *Advances in Neural Information Processing Systems*, 34:8913–8926, 2021.

Chengchun Shi, Xiaoyu Wang, Shikai Luo, Hongtu Zhu, Jieping Ye, and Rui Song. Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework. *Journal of the American Statistical Association*, 118(543):2059–2071, 2023.

Stephen Tu, Roy Frostig, and Mahdi Soltanolkotabi. Learning from many trajectories. *Journal of Machine Learning Research*, 25(216):1–109, 2024.

Feicheng Wang and Lucas Janson. Exact asymptotics for linear quadratic adaptive control. *Journal of Machine Learning Research*, 22(265):1–112, 2021.

Hao Wang, Zakhary Kaplan, Di Niu, and Baochun Li. Optimizing federated learning on non-iid data with reinforcement learning. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, pages 1698–1707. IEEE, 2020.

Yafei Wang, Bo Pan, Wei Tu, Peng Liu, Bei Jiang, Chao Gao, Wei Lu, Shangling Jui, and Linglong Kong. Sample average approximation for stochastic optimization with dependent data: Performance guarantees and tractability. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3859–3867, 2022.

Chun Shan Wong and Wai Keung Li. On a mixture autoregressive model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 62(1):95–115, 2000.

Lihu Xu, Fang Yao, Qiuran Yao, and Huiming Zhang. Non-asymptotic guarantees for robust statistical learning under infinite variance assumption. *Journal of Machine Learning Research*, 24(92):1–46, 2023.

Yi Xu, Shenghuo Zhu, Sen Yang, Chi Zhang, Rong Jin, and Tianbao Yang. Learning with non-convex truncated losses by sgd. In *Uncertainty in Artificial Intelligence*, pages 701–711. PMLR, 2020.

Yuexiang Zhai, Hao Bai, Zipeng Lin, Jiayi Pan, Shengbang Tong, Yifei Zhou, Alane Suhr, Saining Xie, Yann LeCun, Yi Ma, et al. Fine-tuning large vision-language models as decision-making agents via reinforcement learning. *Advances in Neural Information Processing Systems*, 37: 110935–110971, 2024.

Table 2: Examples of control functions λ for the logarithmic truncation envelope.

Control function $\lambda(x)$	Growth rate	Typical integrability condition
$\frac{x^\gamma}{\gamma}$, $\gamma \in (1,2)$ [Chen et al., 2021]	sub-quadratic polynomial	finite γ-th moment
$\frac{x^2}{2}$, i.e., $\gamma = 2$ [Catoni, 2012]	quadratic	finite second moment
$\left\{\frac{\eta-1}{\eta} \vee \left(\frac{2-\eta}{\eta}\right)^{1/2}\right\} x^\eta$, $\eta \in (1,2)$ [Minsker, 2018]	sub-quadratic polynomial	finite η-th moment
$\sum_{k=2}^m \frac{x^k}{k!}$, $m \geq 2$ [Xu et al., 2020]	polynomial	finite m-th moment
General λ satisfying Assumption 4 &5	user-specified	$R_{\lambda\circ\ell}(\Theta) < \infty$ and $\mathbb{E}[\lambda(H_\xi)] < \infty$

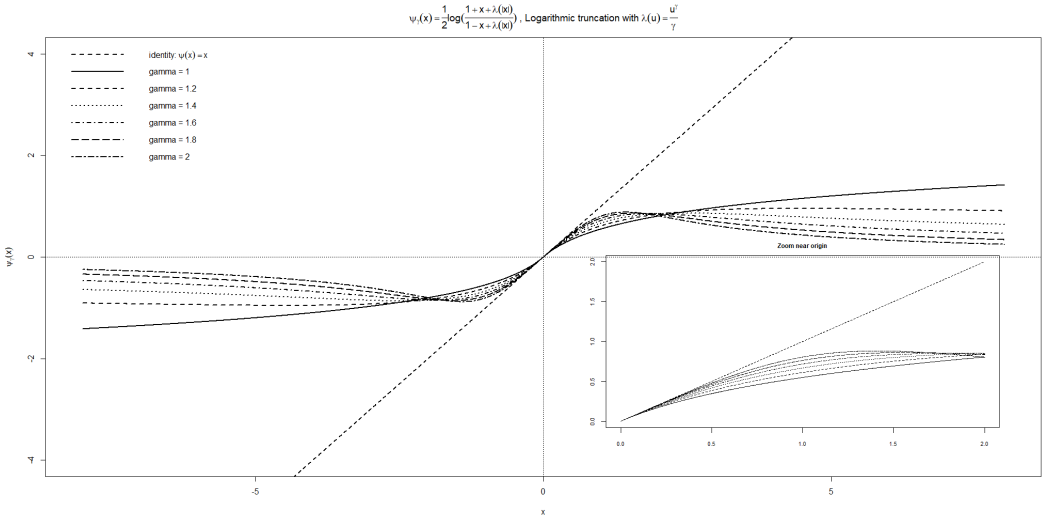

Figure 2: Effect of the control function $\lambda(x) = x^\gamma/\gamma$ on the logarithmic truncation map.

A Trajectory data and related applications

We describe several applications in which the considered data-generating processes naturally arise.

Learning with linear dynamical trajectories. Fix a transition matrix $A \in \mathbb{R}^{n\times n}$ and a control matrix $B \in \mathbb{R}^{n\times m}$. Consider the n-dimensional trajectory $\{x_t\}_{t=1}^T$ generated by the linear system

$$x_{t+1} = Ax_t + Bu_t + \varepsilon_t, \tag{8}$$

where $x_t \in \mathbb{R}^n$ is the system state at time t, initialized either deterministically or randomly at x_0; $u_t \in \mathbb{R}^m$ is the control input; and $\varepsilon_t \in \mathbb{R}^n$ is the additive system noise. Such models arise widely in model-based reinforcement learning, dynamical system identification, and operational control. One canonical example is the *policy estimation for linear-quadratic regulator (LQR)*, where the system evolves according to (8) and the objective is to minimize the cumulative quadratic cost

$$J(K) = \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^T c_t\right], \text{ where } c_t = \left(x_t^\top Q x_t + u_t^\top R u_t\right), u_t = Kx_t,$$

over a feedback policy K, with $Q \succeq 0$ and $R \succ 0$. So the observed data form a single dependent trajectory, making LQR a natural setting in which learning from dependent trajectory data is essential.

This framework also covers *closed-loop system identification under a fixed feedback policy*. Suppose the control input is generated by a known deterministic linear feedback law $u_t = \pi(x_t)$. Given trajectory data $\{\xi_t\}_{t=0}^{T-1}$, where $\xi_t := (x_t, u_t, x_{t+1})$, define the absolute prediction loss

$\ell((A,B);\xi_t) := \|x_{t+1} - Ax_t - Bu_t\|_1$. We then consider the penalized estimator

$$(\widehat{A},\widehat{B}) := \arg\min_{(A,B)\in\Theta_{AB}} \left\{ \widehat{R}_{\psi_\lambda,\ell,\alpha}([\mathrm{vec}(A)^\top, \mathrm{vec}(B)^\top]^\top) + \rho\left(\|A\|_F^2 + \|B\|_F^2\right) \right\}.$$

Since the data are generated by the closed-loop dynamics, the samples $\{\xi_t\}_{t=0}^{T-1}$ are serially dependent rather than independent.

The same formulation also extends naturally to online experimentation and A/B testing in linear dynamical systems. Consider two candidate policies, π_A and π_B. Under policy π_j, $j \in \{A,B\}$, the control input is $u_t^{(j)} = \pi_j(x_t^{(j)})$, and the state evolves as $x_{t+1}^{(j)} = Ax_t^{(j)} + Bu_t^{(j)} + \varepsilon_t^{(j)}$. The resulting trajectory is $\xi_t^{(j)} := \left(x_t^{(j)}, u_t^{(j)}, x_{t+1}^{(j)}\right)$. For a performance metric or outcome $Y(\pi_j)$, the policy contrast can be written as $\Delta := Y(\pi_A) - Y(\pi_B)$, leading to the testing problem

$$H_0 : \Delta = 0 \qquad \text{versus} \qquad H_A : \Delta \neq 0.$$

Because each policy is evaluated along an evolving trajectory, the observations are generally serially dependent, placing online A/B testing within the scope of dependent trajectory data.

Regression tasks from autoregressive process. As a concrete instance of our trajectory-data framework, consider an observed sequence $\{(Y_t, X_t)\}_{t=1}^T$ generated by the autoregressive model

$$X_t = AX_{t-1} + \Sigma W_t, \qquad Y_t = g(\theta^\star; X_t) + \varepsilon_t, \qquad t = 1,\dots,T.$$

Here, $A \in \mathbb{R}^{m\times m}$ is the transition matrix, Σ is a constant matrix, and $\{W_t\}_{t=1}^T$ and $\{\varepsilon_t\}_{t=1}^T$ denote the disturbance and noise sequences, respectively. This model provides a standard representation of dependent trajectory data in time-series analysis, adaptive control, system identification, and dynamic programming [Gerencsér and Vágó, 1999, Kumar and Varaiya, 1986, Mamakoukas et al., 2020]. The observations $\{(Y_t, X_t)\}_{t=1}^T$ encode the evolution of uncertainty along a single stochastic path and may be collected at regular or irregular time points. Under standard assumptions on the innovation sequence, when A is asymptotically stable, meaning that all eigenvalues lie strictly inside the unit circle, the autoregressive process is ergodic and admits a unique stationary distribution [Mokkadem, 1988]. Thus, vector autoregressive dynamics provide a canonical example of trajectory data for which temporal dependence is intrinsic, while stability and tail behavior govern the validity of statistical learning guarantees.

Specifically, we assume that the disturbance sequence $\{W_t\}$ is i.i.d. with a finite covariance matrix, and the regression noise $\{\varepsilon_t\}$ is i.i.d. with a finite γ- th moment for some $\gamma \in (1,2)$. For an observation $\xi_t = (Y_t, X_t)$, we define the loss function as $\ell(\theta,\xi) = |Y - g(\theta,x)|$.

B Additional simulation studies

Dynamic mean estimation. The dynamic mean estimation experiment can be viewed as a stylized prototype for policy-evaluation tasks, where the goal is to estimate an average outcome, reward, or dynamic treatment effect in dynamic A/B testing, reinforcement learning, or sequential decision-making frameworks. Similarly, the regression and logistic regression experiments represent robust supervised learning from dependent covariate-response trajectories.

In this experiment, the data $\{\xi_t\}_{t=1}^T$ is generated from the autoregressive process $\xi_t = \rho\xi_{t-1} + \varepsilon_t$, where ε_t denotes the random error. We vary $\rho \in \{0.25, 0.5, 0.75\}$ throughout this analysis. The error ε_k is generated from a Pareto distribution with tail parameter $\tau \in \{1.3, 1.6, 2.01, 4.01\}$ and is appropriately recentered in order to have zero mean, that is $\varepsilon_i = Z_i - \frac{\tau}{\tau-1}$, $Z_i \sim \mathrm{Pareto}(1,\tau)$. We consider sample sizes $n = 50, 100, 500$, and the replication time is 1000.

Robust logistic regression. We next consider logistic regression with binary outcomes. Let $Y_t \in \{0,1\}$ denote the response and let $\theta^\star \in \mathbb{R}^p$ be the unknown regression coefficient vector, assumed to belong to a compact subset of $\mathbb{R}^p$. The empirical logistic loss is

$$\widehat{R}_{\ell,T}(\theta) = -\frac{1}{T}\sum_{t=1}^T \left[Y_t X_t^\top \theta - \log\left\{1 + \exp(X_t^\top\theta)\right\}\right].$$

In our simulation, the covariates are generated from the autoregressive process $X_t = AX_{t-1} + W_t$, where W_t are i.i.d. $N(0, I_p)$. The entries of A are first generated independently from the standard

normal distribution and then rescaled so that $\|A\|_F = 0.8$. Conditional on X_t, the binary response is generated as

$$Y_t \mid X_t \sim \text{Bernoulli}\left(\frac{\exp(X_t^\top \theta^\star)}{1+\exp(X_t^\top \theta^\star)}\right).$$

The true coefficient vector $\theta^\star$ is generated coordinatewise from $U(-1,1)$. To assess robustness, we contaminate the covariates by heavy-tailed Pareto noise. Specifically, the contaminated covariate is given by $X_t' = X_t + \epsilon_t$, where $\epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{tp})^\top$ and the nonzero contaminated entries are generated independently from a Pareto distribution with scale parameter 1 and shape parameter $\tau \in \{1.3, 1.6, 2.01, 4.01\}$. We vary the contamination proportion from $\{0.1, 0.15, 0.2\}$. We set $p = 20$, $T \in \{100, 500, 1000\}$, and use 100 Monte Carlo replications. As a baseline, we also compute the L_2 penalized non-truncated estimator by adding penalization to (2). Table 3 summarizes the results.

Table 3: Comparison of L_2-estimation error for logistic regression under different (n, τ) and contamination percentage (cp) settings with $\gamma = 1.5$

		cp=0.1		cp=0.15		cp=0.2	
n	τ	LTE	ERM	LTE	ERM	LTE	ERM
100	1.3	1.6456	1.6518	1.6610	1.6819	1.8577	1.8836
	1.6	1.5493	1.5676	1.5553	1.5747	1.7419	1.7620
	2.01	1.4646	1.4763	1.4355	1.4562	1.6071	1.6232
	4.01	1.3394	1.3429	1.2728	1.2767	1.3758	1.3799
500	1.3	1.0849	1.6705	1.3177	2.0054	1.4360	2.1583
	1.6	0.8996	1.3315	1.0612	1.6321	1.1531	1.8153
	2.01	0.8156	1.0175	0.8825	1.2272	0.9430	1.3858
	4.01	0.7156	0.7146	0.7437	0.7467	0.7590	0.7697
1000	1.3	0.9855	1.7962	1.1254	2.1488	1.2311	2.2927
	1.6	0.7348	1.4021	0.8358	1.7407	0.9033	1.9021
	2.01	0.6161	1.0223	0.6834	1.2755	0.7304	1.4064
	4.01	0.5779	0.6010	0.6381	0.6717	0.6689	0.7030

Under the dependent data setting, Table 3 shows that the proposed log-truncated estimator achieves a smaller L_2 estimation error than the non-truncated estimator, particularly when the data are highly heavy-tailed, such as when $\tau = 1.3$ or 1.6. As τ increases, extreme outliers become less frequent and the data become less heavy-tailed; consequently, both methods improve. Nevertheless, the log-truncated estimator continues to outperform the non-truncated estimator across the considered settings.

C Proof of Lemma 1

Lemma 1. *Suppose Assumptions 1 - 5 hold. Let $\theta_\ell^\star$ denote the minimizer of the population risk $R_\ell(\theta)$. Then for any $\delta \in (0,1)$, with probability at least $1-\delta$, we have*

$$\widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta_\ell^\star) - R_\ell(\theta_\ell^\star) \le \frac{f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\theta_\ell^\star) + \frac{1}{T\alpha c_\beta}\log\left(\frac{2C_\beta T}{\delta}\right)\log\left(\frac{2}{\delta c_\beta}\log\left(\frac{2C_\beta T}{\delta}\right)\right).$$

Proof. Substituting the definitions of $\widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta_\ell^\star)$ and $R_\ell(\theta_\ell^\star)$, we obtain

$$\begin{aligned}\widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta_\ell^\star) - R_\ell(\theta_\ell^\star) &= \frac{1}{T\alpha}\sum_{t=1}^{T}\psi_\lambda\left(\alpha\ell\left(\theta_\ell^\star;\widehat{\xi}_t\right)\right) - \mathbb{E}\left[\ell(\theta_\ell^\star;\xi)\right] \\ &= \frac{1}{T\alpha}\sum_{t=1}^{T}\left(\psi_\lambda(\alpha\ell\left(\theta_\ell^\star;\widehat{\xi}_t\right)) - \alpha R_\ell(\theta_\ell^\star)\right).\end{aligned}$$

To address the dependence in the sequence $\left\{\widehat{\xi}_t\right\}_{t=1}^T$, we employ a blocking technique. Without loss of generality, assume that $T = km$ for positive integers k and m. Partition the index set $\{1, \ldots, T\}$

into m interleaving subsequences of length k. For $j = 1, \ldots, m$, define

$$\mathcal{S}_j = \{j, j+m, j+2m, \ldots, j+(k-1)m\}.$$

Define

$$Z_{\mathcal{S}_j} = \sum_{t \in \mathcal{S}_j} \left(\psi_\lambda \left(\alpha \ell \left(\theta_\ell^\star; \widehat{\xi}_t \right) \right) - \alpha R_\ell(\theta_\ell^\star) \right).$$

Then

$$\mathbb{P}\left(\widehat{R}_{\psi_\lambda, \ell, \alpha}(\theta_\ell^\star) - R_\ell(\theta_\ell^\star) \geq x \right) = \mathbb{P}\left(\frac{1}{T\alpha} \sum_{j=1}^{m} Z_{\mathcal{S}_j} \geq x \right) \leq \sum_{j=1}^{m} \mathbb{P}\left(Z_{\mathcal{S}_j} \geq k\alpha x \right).$$

By Berbee [1987, Lemma 2.1], since the spacing between consecutive elements in each subsequence $\mathcal{S}_j$ is m, we can construct independent random variables $\tilde{\xi}_t \sim \widehat{\xi}_t$ for $t \in \mathcal{S}_j$ such that the distance between the distributions is controlled by the mixing coefficient. Define

$$\tilde{Z}_{\mathcal{S}_j} = \sum_{t \in \mathcal{S}_j} \left(\psi_\lambda \left(\alpha \ell \left(\theta^\star; \tilde{\xi}_t \right) \right) - \alpha R_\ell(\theta^\star) \right).$$

Then

$$\begin{aligned} \mathbb{P}\left(Z_{\mathcal{S}_j} \geq k\alpha x \right) &\leq \mathbb{P}\left(\tilde{Z}_{\mathcal{S}_j} \geq k\alpha x \right) + \mathbb{P}\left(\tilde{\xi}_t \neq \xi_t \text{ for some } t \in \mathcal{S}_j \right) \\ &\leq \mathbb{P}\left(\tilde{Z}_{\mathcal{S}_j} \geq k\alpha x \right) + k\,\beta(m). \end{aligned}$$

Combining over all m subsequences gives

$$\mathbb{P}\left(\widehat{R}_{\psi_\lambda, \ell, \alpha}(\theta_\ell^\star) - R_\ell(\theta_\ell^\star) \geq x \right) \leq \sum_{j=1}^{m} \mathbb{P}\left(\tilde{Z}_{\mathcal{S}_j} \geq k\alpha x \right) + mk\,\beta(m). \tag{9}$$

Next, we bound the first term. Using (4) and Assumption 4,

$$\exp\left\{ \psi_\lambda \left(\alpha \ell \left(\theta_\ell^\star; \tilde{\xi}_t \right) \right) \right\} \leq 1 + \alpha \ell \left(\theta_\ell^\star; \tilde{\xi}_t \right) + f(\alpha) \lambda \left(\ell \left(\theta_\ell^\star; \tilde{\xi}_t \right) \right). \tag{10}$$

We now bound the exponential moment

$$\psi_\lambda \left(\alpha \ell \left(\theta^\star; \tilde{\xi}_t \right) \right) - \alpha R_\ell(\theta^\star).$$

Applying (10),

$$\begin{aligned} \mathbb{E}\left[\exp\left(\psi_\lambda \left(\alpha \ell \left(\theta_\ell^\star; \tilde{\xi}_t \right) \right) - \alpha R_\ell(\theta_\ell^\star) \right) \right] &= e^{-\alpha R_\ell(\theta_\ell^\star)} \mathbb{E}\left[\exp\left\{ \psi_\lambda \left(\alpha \ell \left(\theta^\star; \tilde{\xi}_t \right) \right) \right\} \right] \\ &\leq e^{-\alpha R_\ell(\theta_\ell^\star)} \mathbb{E}\left[1 + \alpha \ell \left(\theta_\ell^\star; \tilde{\xi}_t \right) + f(\alpha) \lambda \left(\ell \left(\theta_\ell^\star; \tilde{\xi}_t \right) \right) \right] \\ &= e^{-\alpha R_\ell(\theta_\ell^\star)} \left(1 + \alpha R_\ell(\theta_\ell^\star) + f(\alpha) R_{\lambda \circ \ell}(\theta_\ell^\star) \right) \\ &\leq \exp\left\{ f(\alpha) R_{\lambda \circ \ell}(\theta_\ell^\star) \right\}, \end{aligned}$$

where the last step uses $1 + y \leq e^y$.

Since the variables in each subsequence are independent, Markov's inequality implies

$$\begin{aligned} \mathbb{P}\left(\tilde{Z}_{\mathcal{S}_j} \geq k\alpha x \right) &\leq e^{-k\alpha x} \mathbb{E}\left[\exp\left\{ \tilde{Z}_{\mathcal{S}_j} \right\} \right] \\ &= e^{-k\alpha x} \prod_{t \in \mathcal{S}_j} \mathbb{E}\left[\exp\left\{ \psi_\lambda \left(\alpha \ell \left(\theta_\ell^\star; \tilde{\xi}_t \right) \right) - \alpha R_\ell(\theta_\ell^\star) \right\} \right] \\ &\leq \exp\left\{ -k\alpha x + k f(\alpha) R_{\lambda \circ \ell}(\theta_\ell^\star) \right\}. \end{aligned} \tag{11}$$

Substituting (11) into (9) and noting that $T = mk$, we obtain

$$\mathbb{P}\left(\widehat{R}_{\psi_\lambda, \ell, \alpha}(\theta_\ell^\star) - R_\ell(\theta_\ell^\star) \geq x \right) \leq m \exp\left\{ -k\alpha x + k f(\alpha) R_{\lambda \circ \ell}(\theta_\ell^\star) \right\} + T C_\beta e^{-c_\beta m},$$

where Assumption 3 gives $\beta(m) \leq C\beta e^{-c_\beta m}$.

Choose

$$m = \frac{1}{c_\beta} \log\left(\frac{2C_\beta T}{\delta}\right),$$

and set

$$\begin{aligned} x &= \frac{f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\theta_\ell^\star) + \frac{m}{T\alpha} \log\left(\frac{2m}{\delta}\right) \\ &= \frac{f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\theta_\ell^\star) + \frac{1}{T\alpha c_\beta} \log\left(\frac{2C_\beta T}{\delta}\right) \log\left(\frac{2}{\delta c_\beta} \log\left(\frac{2C_\beta T}{\delta}\right)\right). \end{aligned}$$

This yields

$$\mathbb{P}\left(\widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta_\ell^\star) - R_\ell(\theta_\ell^\star) \geq x\right) \leq \delta,$$

which completes the proof. □

D Proof of Lemma 2

Lemma 2. *Suppose Assumptions Assumptions 1 - 5 hold. Let $N(\Theta, \kappa)$ denote the covering number of the parameter space Θ with radius κ. Then for any $\delta \in (0, 1)$, with probability at least $1 - \delta$, we have*

$$\begin{aligned} R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - \widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) \leq 2\kappa\mathbb{E}[H_\xi] + \frac{c_1 f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\Theta) + \frac{c_1 f(\alpha\kappa)}{\alpha}\mathbb{E}[\lambda(H_\xi)] \\ + \frac{1}{T\alpha c_\beta} \log\left(\frac{2C_\beta T\, N(\Theta,\kappa)}{\delta}\right) \log\left(\frac{2N(\Theta,\kappa)}{\delta c_\beta} \log\left(\frac{2C_\beta T\, N(\Theta,\kappa)}{\delta}\right)\right). \end{aligned}$$

Proof. Let $\mathcal{N}(\Theta, \kappa)$ be a κ-net of the parameter space Θ and denote its covering number as $N(\Theta, \kappa)$. By definition, for any $\widehat{\theta}^\star_{\psi_\lambda,\ell} \in \Theta$, there exists a $\tilde{\theta} \in \mathcal{N}(\Theta, \kappa)$ such that $\left\|\widehat{\theta}^\star_{\psi_\lambda,\ell} - \tilde{\theta}\right\| \leq \kappa$. By Assumption 2, the loss function satisfies the Lipschitz condition

$$\left|\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}; \xi\right) - \ell\left(\tilde{\theta}; \xi\right)\right| \leq H_\xi \left\|\widehat{\theta}^\star_{\psi_\lambda,\ell} - \tilde{\theta}\right\| \leq \kappa H_\xi.$$

This implies

$$\ell\left(\tilde{\theta}; \xi\right) - \kappa H_\xi \leq \ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}; \xi\right) \leq \ell\left(\tilde{\theta}; \xi\right) + \kappa H_\xi. \tag{12}$$

Taking the expectation with respect to the stationary measure P, we obtain,

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) = \mathbb{E}\left[\ell(\widehat{\theta}^\star_{\psi_\lambda,\ell})\right] \leq \mathbb{E}\left[\ell(\tilde{\theta}; \xi) + \kappa H_\xi\right] = R_\ell(\tilde{\theta}) + \kappa\mathbb{E}[H_\xi], \tag{13}$$

For the empirical risk $\widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right)$, (4) implies

$$\begin{aligned} -\widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) &= -\frac{1}{T\alpha}\sum_{t=1}^{T} \psi_\lambda(\alpha l(\widehat{\theta}^\star_{\psi_\lambda,\ell}; \widehat{\xi}_t)) \qquad (14) \\ &\leq \frac{1}{T\alpha}\sum_{t=1}^{T} \log\left[1 - \alpha l(\widehat{\theta}^\star_{\psi_\lambda,\ell}; \widehat{\xi}_t) + \lambda\left(\alpha|l(\widehat{\theta}^\star_{\psi_\lambda,\ell}; \widehat{\xi}_t)|\right)\right] \\ &\leq \frac{1}{T\alpha}\sum_{t=1}^{T} \log\left[1 - \alpha l(\tilde{\theta}; \widehat{\xi}_t) + \alpha\kappa H_{\widehat{\xi}_t} + c_1 f(\alpha)\lambda\left(|l(\tilde{\theta}; \widehat{\xi}_t)|\right) + c_1 f(\alpha\kappa)\lambda(H_{\widehat{\xi}_t})\right], \end{aligned}$$

where the last inequality follows (12) and Assumption 4. Denote

$$\mathcal{A}_{\tilde{\theta},\widehat{\xi}_t} = 1 - \alpha l(\tilde{\theta}; \widehat{\xi}_t) + \alpha\kappa H_{\widehat{\xi}_t} + c_1 f(\alpha)\lambda\left(|l(\tilde{\theta}; \widehat{\xi}_t)|\right) + c_1 f(\alpha\kappa)\lambda(H_{\widehat{\xi}_t}).$$

Combining (13) and (14), the event $\left\{R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - \widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) \geq x\right\}$ implies that for some $\tilde{\theta} \in \mathcal{N}(\Theta,\kappa)$,

$$R_\ell(\tilde{\theta}) + \kappa\mathbb{E}[H_\xi] + \frac{1}{T\alpha}\sum_{t=1}^{T}\log\left(\mathcal{A}_{\tilde{\theta},\widehat{\xi}_t}\right) \geq x.$$

Applying the union bound over $\mathcal{N}(\Theta,\kappa)$, we obtain

$$\begin{aligned}
&\mathbb{P}\left(R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - \widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) \geq x\right) \\
&\quad\leq N(\Theta,\kappa)\mathbb{P}\left(\sum_{t=1}^{T}\log\left(\mathcal{A}_{\tilde{\theta},\widehat{\xi}_t}\right) \geq T\alpha\left(x - R_\ell\left(\tilde{\theta}\right) - \kappa\mathbb{E}[H_\xi]\right)\right).
\end{aligned}$$

We now proceed to bound the probability for a fixed $\tilde{\theta}$ using the block technique similar to the proof of Lemma 1. We assume that $T = km$ where m is the block gap and k is the length. We partition the index set into m interleaving sub-sequences $\mathcal{S}_j = \{j, j+m, \ldots, j+(k-1)m\}$. Applying Berbee [1987, Lemma 2.1], for each sub-sequence $\mathcal{S}_j$, there exists an independent sequence $\tilde{\xi}_t \sim \widehat{\xi}_t$. Then we have,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{t=1}^{T}\log\left(\mathcal{A}_{\tilde{\theta},\widehat{\xi}_t}\right) \geq T\alpha\left(x - R_\ell\left(\tilde{\theta}\right) - \kappa\mathbb{E}[H_\xi]\right)\right) \\
&\leq \sum_{j=1}^{m}\mathbb{P}\left(\sum_{t\in\mathcal{S}_j}\log\left(\mathcal{A}_{\tilde{\theta},\widehat{\xi}_t}\right) \geq k\alpha\left(x - R_\ell\left(\tilde{\theta}\right) - \kappa\mathbb{E}[H_\xi]\right)\right) \\
&\leq \sum_{j=1}^{m}\mathbb{P}\left(\sum_{t\in\mathcal{S}_j}\log\left(\mathcal{A}_{\tilde{\theta},\tilde{\xi}_t}\right) \geq k\alpha\left(x - R_\ell\left(\tilde{\theta}\right) - \kappa\mathbb{E}[H_\xi]\right)\right) + mk\beta(m).
\end{aligned}$$

Similar with the estimation of (11) in the proof of Lemma 1, equation (4), Assumption 4 and the independence of $\left\{\tilde{\xi}_t\right\}_{t\in\mathcal{S}_j}$ imply,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{t\in\mathcal{S}_j}\log\left(\mathcal{A}_{\tilde{\theta},\tilde{\xi}_t}\right) \geq k\alpha\left(x - R_\ell\left(\tilde{\theta}\right) - \kappa\mathbb{E}[H_\xi]\right)\right) \\
&\leq \exp\left\{-k\alpha\left(x - R_\ell\left(\tilde{\theta}\right) - \kappa\mathbb{E}[H_\xi]\right)\right\}\mathbb{E}\left[\exp\left\{\sum_{t\in\mathcal{S}_j}\log\left(\mathcal{A}_{\tilde{\theta},\tilde{\xi}_t}\right)\right\}\right] \\
&= \exp\left\{-k\alpha\left(x - R_\ell\left(\tilde{\theta}\right) - \kappa\mathbb{E}[H_\xi]\right)\right\}\prod_{t\in\mathcal{S}_j}\mathbb{E}\left[\exp\left\{\log\left(\mathcal{A}_{\tilde{\theta},\tilde{\xi}_t}\right)\right\}\right] \\
&\leq \exp\left\{-k\alpha\left(x - R_\ell\left(\tilde{\theta}\right) - \kappa\mathbb{E}[H_\xi]\right)\right\} \\
&\quad\prod_{t\in\mathcal{S}_j}\exp\left\{-\alpha R_\ell\left(\tilde{\theta}\right) + \alpha\kappa\mathbb{E}[H_\xi] + c_1 f(\alpha)R_{\lambda\circ\ell}\left(\tilde{\theta}\right) + c_1 f(\alpha\kappa)\mathbb{E}[\lambda(H_\xi)]\right\} \\
&\leq \exp\left\{-k\alpha x + 2k\alpha\kappa\mathbb{E}[H_\xi] + c_1 k f(\alpha)R_{\lambda\circ\ell}(\Theta) + c_1 k f(\alpha\kappa)\mathbb{E}[\lambda(H_\xi)]\right\},
\end{aligned}$$

where $R_{\lambda\circ\ell}(\Theta) = \sup_{\theta\in\Theta} R_{\lambda\circ\ell}(\theta)$. Combining the estimates above, we obtain

$$\begin{aligned}
&\mathbb{P}\left(R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - \widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) \geq x\right) \\
&\quad\leq N(\Theta,\kappa)m\exp\left\{-k\alpha x + 2k\alpha\kappa\mathbb{E}[H_\xi] + c_1 k f(\alpha)R_{\lambda\circ\ell}(\Theta) + c_1 k f(\alpha\kappa)\mathbb{E}[\lambda(H_\xi)]\right\} \\
&\quad + N(\Theta,\kappa)TC_\beta\exp\{-c_\beta m\}.
\end{aligned}$$

Set

$$x = 2\kappa\mathbb{E}[H_\xi] + \frac{c_1 f(\alpha)}{\alpha}R_{\lambda\circ\ell}(\Theta) + \frac{c_1 f(\alpha\kappa)}{\alpha}\mathbb{E}[\lambda(H_\xi)] + \frac{m}{T\alpha}\log\left(\frac{2mN(\Theta,\kappa)}{\delta}\right),$$

and $m = \frac{1}{c_\beta} \log\left(\frac{2C_\beta TN(\Theta,\kappa)}{\delta}\right)$ which yields $N(\Theta,\kappa)mkC_\beta e^{-c_\beta m} \leq \delta/2$. We conclude that

$$\mathbb{P}\Bigg(R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - \widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) \geq 2\kappa\mathbb{E}[H_\xi] + \frac{c_1 f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\Theta) + \frac{c_1 f(\alpha\kappa)}{\alpha}\mathbb{E}[\lambda(H_\xi)] + \frac{1}{T\alpha c_\beta}\log\left(\frac{2C_\beta TN(\Theta,\kappa)}{\delta}\right)\log\left(\frac{2N(\Theta,\kappa)}{\delta c_\beta}\log\left(\frac{2C_\beta TN(\Theta,\kappa)}{\delta}\right)\right)\Bigg) \leq \delta.$$

□

E Proof of Theorem 1

We first establish the error decomposition for the excess risk $R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell\left(\theta^\star_\ell\right)$. Decompose the excess risk, one has,

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) = \left(R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - \widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right)\right) \tag{15}$$

$$+ \left(\widehat{R}_{\psi_\lambda,\ell,\alpha}(\widehat{\theta}^\star_{\psi_\lambda,\ell}) - \widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta^\star_\ell)\right) + \left(\widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta^\star_\ell) - R_\ell(\theta^\star_\ell)\right). \tag{16}$$

According to the definition of $\widehat{\theta}^\star_{\psi_\lambda,\ell}$, we have

$$\widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) + \rho\left\|\widehat{\theta}^\star_{\psi_\lambda,\ell}\right\|^2 \leq \widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta^\star_\ell) + \rho\left\|\theta^\star\right\|^2.$$

Rearranging this inequality, we obtain an upper bound for the second term of (15)

$$\widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - \widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta^\star_\ell) \leq \rho|\theta^\star|^2 - \rho\left\|\widehat{\theta}^\star_{\psi_\lambda,\ell}\right\|^2 \leq \rho\left\|\theta^\star\right\|^2.$$

Substituting it back into (15), we obtain the decomposition,

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) \leq R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - \widehat{R}_{\psi_\lambda,\ell,\alpha}\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) + \widehat{R}_{\psi_\lambda,\ell,\alpha}(\theta^\star_\ell) - R_\ell(\theta^\star_\ell) + \rho\|\theta^\star_\ell\|^2. \tag{17}$$

Applying Lemma 1 and Lemma 2 with probability $\delta/2$ for each, respectively, we obtain that with probability at most δ,

$$\begin{aligned}
R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell\left(\theta^\star_\ell\right) \leq & \frac{f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\theta^\star_\ell) + \frac{1}{T\alpha c_\beta}\log\left(\frac{2C_\beta T}{\delta}\right)\log\left(\frac{2}{\delta c_\beta}\log\left(\frac{2C_\beta T}{\delta}\right)\right) + \rho\left\|\theta^\star\right\|^2 \\
& + 2\kappa\mathbb{E}[H_\xi] + \frac{c_1 f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\Theta) + \frac{c_1 f(\alpha\kappa)}{\alpha}\mathbb{E}[\lambda(H_\xi)] \\
& + \frac{1}{T\alpha c_\beta}\log\left(\frac{2C_\beta TN(\Theta,\kappa)}{\delta}\right)\log\left(\frac{2N(\Theta,\kappa)}{\delta c_\beta}\log\left(\frac{2C_\beta TN(\Theta,\kappa)}{\delta}\right)\right) \\
\leq & \frac{(1+c_1)f(\alpha)}{\alpha} R_{\lambda\circ\ell}(\Theta) + \frac{1}{T\alpha c_\beta}\left(\mathcal{L}(1) + \mathcal{L}(N(\Theta,\kappa))\right) + \rho\left\|\Theta^\star\right\|^2 \\
& + 2\kappa\mathbb{E}[H_\xi] + \frac{c_1 f(\alpha\kappa)}{\alpha}\mathbb{E}[\lambda(H_\xi)],
\end{aligned}$$

where

$$\mathcal{L}(x) = \log\left(\frac{2C_\beta Tx}{\delta}\right)\log\left(\frac{2x}{\delta c_\beta}\log\left(\frac{2C_\beta Tx}{\delta}\right)\right),$$

and $\|\Theta^\star\| = \inf_{\theta^\star\in\Theta^\star}\|\theta\|$. The final inequality follows by bounding $R_{\lambda\circ\ell}(\theta^\star_\ell) \leq \sup_{\theta\in\Theta} R_{\lambda\circ\ell}(\theta) := R_{\lambda\circ\ell}(\Theta)$ and taking the infimum over all $\theta^\star_\ell \in \Theta^\star$.

Let the covering net radius be $\kappa = 1/T$. By equating the bias and variance trade-off terms, we choose the optimal scaling parameter α as

$$\alpha = f^{-1}\left(\frac{\mathcal{L}(1) + \mathcal{L}(N(\Theta,1/T))}{Tc_\beta(1+c_1)R_{\lambda\circ\ell}(\Theta)}\right).$$

Substituting this α into the main inequality, and noting that $\frac{f(\alpha/T)}{\alpha} = \frac{1}{T}\frac{f(\alpha/T)}{\alpha/T} \to 0$ as $T\to\infty$, the excess risk of $\widehat{\theta}^\star_{\psi_\lambda,\ell}$ is bounded by

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) \leq \frac{1}{T}\frac{2\left(\mathcal{L}(1) + \mathcal{L}\left(N\left(\Theta,1/T\right)\right)\right)}{\alpha c_\beta} + \frac{2}{T}\mathbb{E}[H_\xi] + \frac{c_1}{T}\frac{f(\alpha/T)}{\alpha/T}\mathbb{E}[\lambda(H_\xi)] + \rho|\Theta^\star|^2.$$

F Proof of Corollary 1

Given the definition of $\lambda(x) = |x|^\gamma/\gamma$ with $\gamma \in (1,2)$, we can explicitly determine the function $f(t)$ and constant c_1 in Assumption 4. Since $\lambda(\nu x) = |\nu x|^\gamma/\gamma = \nu^\gamma \lambda(x)$, we have $f(\nu) = \nu^\gamma$. Furthermore, utilizing the standard inequality $|x+y|^\gamma \le 2^{\gamma-1}(|x|^\gamma + |y|^\gamma)$ for $\gamma > 1$, we obtain $c_1 = 2^{\gamma-1}$.

Applying the general bound from Theorem 1 and equating the bias and variance terms to tune the optimal scaling parameter α, we have

$$\alpha = f^{-1}\left(\frac{\mathcal{L}(1)+\mathcal{L}(N(\Theta,1/T))}{Tc_\beta(1+c_1)R_{\lambda\circ\ell}(\Theta)}\right) = \left(\frac{\mathcal{L}(1)+\mathcal{L}(N(\Theta,1/T))}{Tc_\beta(1+2^{\gamma-1})R_{\lambda\circ\ell}(\Theta)}\right)^{\frac{1}{\gamma}}. \tag{18}$$

For the covering number $N(\Theta, \frac{1}{T})$, [Xu et al., 2023, Lemma 15] implies that $N(\Theta, \frac{1}{T}) \le (1+2rT)^p$. Substituting α into the excess risk bound of Theorem 1 yields

$$R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) \le \frac{2\Big(\mathcal{L}(1)+\mathcal{L}(N(\Theta,1/T))\Big)}{T\alpha c_\beta} + \frac{c_1 f(\alpha/T)}{\alpha}\mathbb{E}[\lambda(H_\xi)] + \frac{2}{T}\mathbb{E}[H_\xi] + \rho|\Theta^\star|^2. \tag{19}$$

To simplify the logarithmic complexity, note that

$$\mathcal{L}(x) = \log\left(\frac{2C_\beta Tx}{\delta}\right)\log\left(\frac{2x}{\delta c_\beta}\log\Big(\frac{2C_\beta Tx}{\delta}\Big)\right) \le 2\log^2\left(\frac{2C_\beta Tx}{\delta}\right).$$

Since $\gamma \in (1,2)$, the exponent satisfies $\frac{\gamma-1}{\gamma} < \frac{1}{2}$. Combined with the bound for $N(\Theta, \frac{1}{T})$, we can get

$$\begin{aligned}\big(\mathcal{L}(1)+\mathcal{L}\big(N(\Theta,1/T)\big)\big)^{\frac{\gamma-1}{\gamma}} &\le \left(2\log^2\left(\frac{2C_\beta T}{\delta}\right) + 2\log^2\left(\frac{2C_\beta T(1+2rT)^p}{\delta}\right)\right)^{\frac{1}{2}} \\ &\le \left(4p^2\log^2\left(\frac{2C_\beta T(1+2rT)}{\delta}\right)\right)^{\frac{1}{2}} = 2p\log\left(\frac{2C_\beta T(1+2rT)}{\delta}\right).\end{aligned}$$

Substituting this upper bound and α into the first term of (19), we have

$$\begin{aligned}\frac{2\Big(\mathcal{L}(1)+\mathcal{L}(N(\Theta,1/T))\Big)}{T\alpha c_\beta} &= 2\left(\frac{\mathcal{L}(1)+\mathcal{L}(N(\Theta,1/T))}{Tc_\beta}\right)^{\frac{\gamma-1}{\gamma}}\Big((1+2^{\gamma-1})R_{\lambda\circ\ell}(\Theta)\Big)^{\frac{1}{\gamma}} \\ &\le 4pT^{-\frac{\gamma-1}{\gamma}}\Big((1+2^{\gamma-1})R_{\lambda\circ\ell}(\Theta)\Big)^{\frac{1}{\gamma}} c_\beta^{\frac{1}{\gamma}-1}\log\left(\frac{2C_\beta T(1+2rT)}{\delta}\right).\end{aligned}$$

For the second term of (19), we have

$$\begin{aligned}\frac{c_1 f(\alpha/T)}{\alpha}\mathbb{E}[\lambda(H_\xi)] &= \frac{2^{\gamma-1}\alpha^{\gamma-1}}{T^\gamma}\mathbb{E}[\lambda(H_\xi)] \\ &= \frac{2^{\gamma-1}}{T^\gamma}\left(\frac{\mathcal{L}(1)+\mathcal{L}(N(\Theta,1/T))}{Tc_\beta(1+2^{\gamma-1})R_{\lambda\circ\ell}(\Theta)}\right)^{\frac{\gamma-1}{\gamma}}\mathbb{E}[\lambda(H_\xi)] \\ &\le 4pT^{-\gamma-\frac{\gamma-1}{\gamma}}\big((1+2^{\gamma-1})R_{\lambda\circ\ell}(\Theta)\big)^{\frac{1}{\gamma}-1}c_\beta^{\frac{1}{\gamma}-1}\mathbb{E}[\lambda(H_\xi)]\log\left(\frac{2C_\beta T(1+2rT)}{\delta}\right).\end{aligned}$$

To ease the notation, let us denote $C_{\lambda,\ell,\gamma} = \big((1+2^{\gamma-1})R_{\lambda\circ\ell}(\Theta)\big)^{\frac{1}{\gamma}}$. Summing up all the terms, we arrive at the excess risk bound

$$\begin{aligned}R_\ell\left(\widehat{\theta}^\star_{\psi_\lambda,\ell}\right) - R_\ell(\theta^\star_\ell) &\le pT^{-\frac{\gamma-1}{\gamma}}\log\left(\frac{2C_\beta T(1+2rT)}{\delta}\right)\Big(4C_{\lambda,\ell,\gamma} + 4\mathbb{E}[\lambda(H_\xi)]C_{\lambda,\ell,\gamma}^{\gamma-1}\Big)c_\beta^{\frac{1}{\gamma}-1} \\ &+ \frac{2}{T}\mathbb{E}[H_\xi] + \rho|\Theta^\star|^2 \\ &= O\big(pT^{-\frac{\gamma-1}{\gamma}}\log\left(2C_\beta T(1+2rT)/\delta\right) + \rho|\Theta^\star|\big).\end{aligned}$$

G Proof of Corollary 2

Following [Duchi et al., 2012], the sequence $\{X_t\}$ is exponentially ergodic with a mixing rate bounded by $M\eta^T$, where $M > 0$ and $\eta \in (0, 1)$ are constants depending on the matrices A and Σ. As shown by [Lu et al., 2022], this ergodicity implies that the sequence is exponentially β-mixing. Furthermore, the joint process $\xi_t^\top = (Y_t^\top, X_t)$ is also exponentially β-mixing. This formulation maps directly to our condition with the mixing constants given by $C_\beta = M$ and $c_\beta = -\ln \eta$. Thus, Assumption 3 is satisfied, which completes the proof of Corollary 2.

In the special case where $g(\theta, \xi) = \langle \theta, X \rangle$, it is straightforward to verify Assumptions 2 and 5. For any $\theta_1, \theta_2 \in \Theta$, we have

$$\begin{aligned} |\ell(\theta_1; \xi) - \ell(\theta_2; \xi)| &= \Big| |Y - \langle \theta_1, X \rangle| - |Y - \langle \theta_2, X \rangle| \Big| \\ &\le \Big| (Y - \langle \theta_1, X \rangle) - (Y - \langle \theta_2, X \rangle) \Big| \\ &\le \|X\| \|\theta_1 - \theta_2\|. \end{aligned}$$

This demonstrates that the loss function is Lipschitz continuous with respect to θ. Consequently, we can define $H_\xi = \|X\|$. Since $\{W_t\}$ is an i.i.d. sequence with finite variance, and the γ-th moment of ϵ_t is finite, it follows that the γ-th moment of $\{\xi_t\}$ also exists. Given that the sequence $\{\xi_t\}_{t=1}^T$ is ergodic and has finite expectation, both Assumption 2 and Assumption 5 are satisfied. Consequently, under Assumption 1 and the specification $g(\theta, X) = \langle \theta, X \rangle$, Corollary 2 holds.

H Proof of Corollary 3

The proof of Corollary 3 is similar with the proof of Corollary 2. The data-generating process is a closed-loop system evolving according to $x_{t+1} = Ax_t + Bu_t + \varepsilon_t$. Under the feedback law $u_t = Kx_t$, the system can be rewritten as a standard vector autoregressive process:

$$x_{t+1} = Fx_t + \varepsilon_t, \quad \text{where } F = A + BK.$$

Let $\theta = [\mathrm{vec}(A)^\top, \mathrm{vec}(B)^\top]^\top \in \mathbb{R}^p$ where $p = n^2 + nm$. Notice that $\|\theta\|^2 = \|A\|_F^2 + \|B\|_F^2$.

Because $\rho(F) < 1$, the closed-loop transition matrix F is asymptotically stable, and the noise $\{\varepsilon_t\}$ is i.i.d. with a finite γ-th moment, the sequence $\{x_t\}$ is exponentially ergodic and admits a stationary distribution with a finite γ-th moment. This directly implies the sequence is exponentially β-mixing, satisfying Assumption 3.

For any two parameters θ_1 and θ_2 corresponding to matrix pairs (A_1, B_1) and (A_2, B_2) in the compact set Θ_{AB}, the L_1 loss satisfies

$$\begin{aligned} |l(\theta_1; \xi_t) - l(\theta_2; \xi_t)| &= |\|x_{t+1} - A_1 x_t - B_1 u_t\|_1 - \|x_{t+1} - A_2 x_t - B_2 u_t\|_1| \\ &\le \|(A_1 - A_2)x_t + (B_1 - B_2)Kx_t\|_1 \\ &\le \sqrt{n} \|(A_1 - A_2)x_t + (B_1 - B_2)Kx_t\| \\ &\le \sqrt{n} \left(\|A_1 - A_2\|_F \|x_t\| + \|B_1 - B_2\|_F \|K\|_F \|x_t\| \right) \\ &\le \sqrt{n} \max(1, \|K\|_F) \left(\|A_1 - A_2\|_F + \|B_1 - B_2\|_F \right) \|x_t\| \\ &\le \sqrt{2n} \max(1, \|K\|_F) \sqrt{\|A_1 - A_2\|_F^2 + \|B_1 - B_2\|_F^2} \|x_t\| \\ &= C_K \|x_t\| \|\theta_1 - \theta_2\| \end{aligned}$$

where $C_K = \sqrt{2n} \max(1, \|K\|_F)$. Thus, the loss is Lipschitz continuous with respect to the vectorized parameter θ, with $H_{\xi_t} = C_K \|x_t\|$, satisfying Assumption 2.

Since $\gamma \in (1, 2)$, the existence of the γ-th moment guarantees the existence of the first moment. Therefore, $\mathbb{E}[H_{\xi_t}] = C_K \mathbb{E}[\|x_t\|] < \infty$. Additionally, using the truncation function $\lambda(x) = |x|^\gamma / \gamma$, we have $\mathbb{E}[\lambda(H_{\xi_t})] = \frac{(C_K)^\gamma}{\gamma} \mathbb{E}[\|x_t\|^\gamma] < \infty$. Thus, Assumption 5 is fully satisfied.