VC Theory for Inventory Policies

Yaqi Xie (Chicago Booth)

joint work with

Will Ma (Columbia GSB),

Linwei Xin (Cornell ORIE)

CHICAGO BOOTH

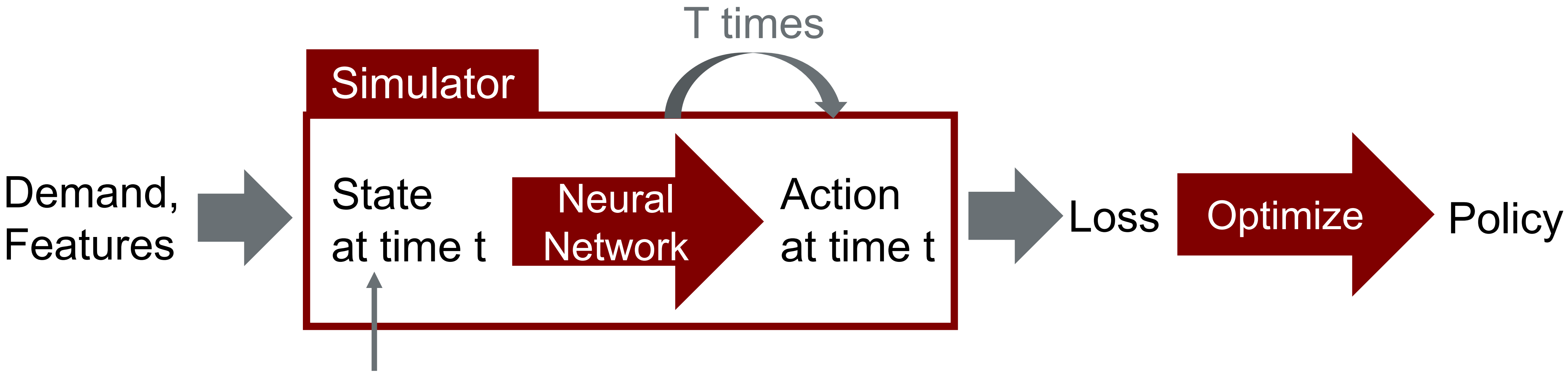
Background: Data-driven Inventory Control

Approaches	
Classical Approach	Predict and then optimize
Deep Reinforcement Learning (DRL)	End-to-end e.g., Gijsbrechts Boute Van Mieghem Zhang 2022
Supervised Learning (SL)	Sinclair et al. 2023: Hindsight evaluation
	Madeka Torkkola Eisenach Luo Foster Kakade 2022: SL > DRL by Amazon A/B testing
	Alvo Russo Kanoria 2023: SL > heuristics by synthetic data

Our paper: theoretical analysis for classical inventory policies

Supervised Learning for Inventory

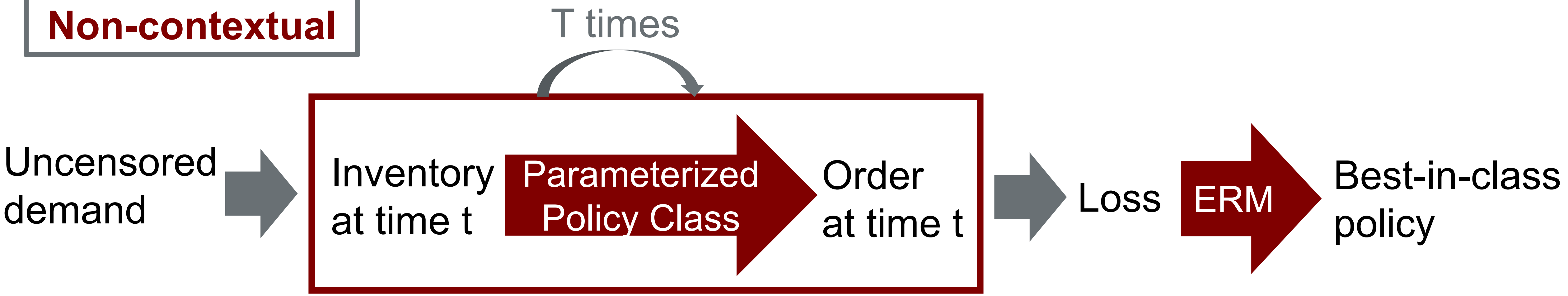
- Assume uncensored demand ⟵ exogenous
- Can evaluate counterfactual performance of any policy on past demand trajectories



Contextual

State Vector = (Inventory State, Features)

Non-contextual



Question:

- How badly could we be overfitting? **Estimation/Generalization error**

Summary of “VC Theory for Inventory Policies”: in **SL framework**, we use **VC theory** to analyze the **generalizability** of inventory policies, and how many demand samples are needed to learn **near-optimal** ones.

Sample complexity

Inventory Basics: Single Durable Good, Periodic Review

- $t = 1, \ldots, T$: finite time horizon
- x^t: inventory at start of time t, with $x^1 = 0$
- y^t: inventory after replenishing at time t

Order quantity

- d^t: realized demand at time t
- c^t: cost paid at time t

$$c^t = b\max\{d^t - y^t, 0\} + h\max\{y^t - d^t, 0\} + K\mathbb{I}(y^t > x^t)$$

  - b, h: unit costs for understocking, overstocking
  - K: fixed cost for each replenishment

For the talk, we assume lead time is 0 and lost-sales demand:

- y^t can be any number $\geq x^t$
- $x^{t+1} = \max\{y^t - d^t, 0\}$

Can handle positive lead times for backlogged demand ($x^{t+1} = y^t - d^t$)

Classes of Inventory Policies considered

1) ***S* Policies**: defined by **base stock** $S \in \mathbb{R}_{\geq 0}$

$$y^t = S$$

- optimal if d^t drawn IID across t and $K = 0$

2) $(\boldsymbol{s}, \boldsymbol{S})$ **Policies**: defined by $S \in \mathbb{R}_{\geq 0}$ and **re-order point** $s \in [0, S]$

$$y^t = x^t + (S - x^t)\mathbb{I}(x^t \leq s)$$

- asymptotically optimal as $T \to \infty$ if d^t drawn IID across t and $K > 0$

3) $(\boldsymbol{S^t})$ **Policies**: defined by a vector of base stocks $(S^t)_{t=1}^T \in \mathbb{R}_{\geq 0}^T$

$$y^t = \max\{S^t, x^t\}$$

- optimal if d^t is independent (non-identical) across t and $K = 0$

Learning **near-optimal** policy for **independent** demands

We learn **best-in-class** policies for 1) – 3) **for arbitrary demands**

Learning Theory Basics

- $\pi \in \Pi$ ($\Pi = \Pi_S, \Pi_{(s,S)}, \text{or } \Pi_{(S^t)}$): inventory policies
- $\boldsymbol{d} = (d^1, \ldots, d^T)$: demand sequence/trajectory
- $\ell(\pi, \boldsymbol{d}) = \frac{1}{T}\sum_{t=1}^{T} c^t(\pi, \boldsymbol{d})$: loss of policy π on sequence $\boldsymbol{d}$ ⟵ normalized
- $\boldsymbol{D}$: (unknown) distribution from which $\boldsymbol{d}$ is drawn
  - $L(\pi) = \mathbb{E}_{\boldsymbol{d} \sim \boldsymbol{D}}[\ell(\pi, \boldsymbol{d})]$: true loss of π
  - $\pi^* = \arg\inf_{\pi \in \Pi} L(\pi)$: policy in Π minimizing true loss ⟵ best-in-class policy
- $(\boldsymbol{d}_i)_{i=1}^{N}$: N training samples drawn IID from $\boldsymbol{D}$
  - $\hat{L}(\pi) = \frac{1}{N}\sum_{i=1}^{N} \ell(\pi, \boldsymbol{d}_i)$: empirical loss of π
  - $\hat{\pi} = \arg\inf_{\pi \in \Pi} \hat{L}(\pi)$: policy in Π minimizing in-sample loss ⟵ empirical risk minimizer (ERM)

Estimation and Generalization Errors

Expectation w.r.t. samples

$$\underbrace{\mathbb{E}[L(\hat{\pi})]}_{\text{Out-of-sample Loss}} = \underbrace{\mathbb{E}[L(\overset{\text{ERM}}{\hat{\pi}}) - L(\overset{\text{Best-in-class}}{\pi^*})]}_{\text{Estimation Error}} + L(\pi^*)$$

Not depend on samples

$$\underbrace{\mathbb{E}[L(\hat{\pi}) - L(\pi^*)]}_{\text{Estimation Error}} \le \underbrace{\mathbb{E}\left[\sup_{\pi\in\Pi}\left(L(\pi) - \hat{L}(\pi)\right)\right]}_{\text{Generalization Error}} = O\left(\sqrt{\mathrm{PDim}(\Pi)/N}\right) \quad \forall \text{ distributions } \boldsymbol{D}$$

<u>Goal</u>: bound the **generalization errors** of S, (s, S) and (S^t) policy classes

- w.r.t. sample size $\boldsymbol{N}$ and horizon length $\boldsymbol{T}$
- **without any assumption** on the demand distribution $\boldsymbol{D}$

Generalization Error, VC/Pseudo-dimension

VC Theory: $\underbrace{\mathbb{E}[L(\hat{\pi}) - L(\pi^*)]}_{\text{Estimation Error}} \leq \underbrace{\mathbb{E}\left[\sup_{\pi\in\Pi}\left(L(\pi) - \hat{L}(\pi)\right)\right]}_{\text{Generalization Error (GE)}} = O\left(\boxed{\sqrt{\mathrm{PDim}(\Pi)/N}}\right)$ $\forall$ distributions $\boldsymbol{D}$

<u>Our results</u>:

1) $\mathrm{PDim}(\Pi_S) \leq 2$

2) $\mathrm{PDim}\left(\Pi_{(s,S)}\right) = O(\log T)$, $\mathrm{GE}\left(\Pi_{(s,S)}\right) = \Omega\left(\sqrt{\log T \,/\, \log\log T \,/ N}\right)$

3) $\mathrm{PDim}\left(\Pi_{(S^t)}\right) = \Omega(T)$, $\mathrm{GE}\left(\Pi_{(S^t)}\right) = O\left(\sqrt{1/N}\right)$

Policy Class	Lower Bound	Upper Bound
S Policies	$\Omega\left(\sqrt{1/N}\right)$	$\boldsymbol{O\left(\sqrt{1/N}\right)}$
(s,S) Policies	$\boldsymbol{\Omega\left(\sqrt{\log T \,/ \log\log T \,/ N}\right)}$	$\boldsymbol{O\left(\sqrt{\log T \,/ N}\right)}$
(S^t) Policies	$\Omega\left(\sqrt{1/N}\right)$	$\boldsymbol{O\left(\sqrt{1/N}\right)}$

New results in **Red**

Surprising improvement from literature

Arbitrary demand

Independent demand

S Policies, (s, S) Policies

Novel way to prove
Newsvendor (uncensored)

Policy Class	Lower Bound	Upper Bound
S Policies	$\Omega\left(\sqrt{1/N}\right)$	$\boldsymbol{O}\left(\sqrt{\mathbf{1}/\boldsymbol{N}}\right)$
(s, S) Policies	$\boldsymbol{\Omega}\left(\sqrt{\log \boldsymbol{T} \,/\, \log\log \boldsymbol{T} \,/\boldsymbol{N}}\right)$	$\boldsymbol{O}\left(\sqrt{\log \boldsymbol{T} \,/\boldsymbol{N}}\right)$

$\leftarrow \mathrm{PDim}(\Pi_S) \le 2$

$\leftarrow \mathrm{PDim}\left(\Pi_{(s,S)}\right) = O(\log T)$

∃instance, s.t.
$\mathrm{PDim}_\gamma\left(\Pi_{(s,S)}\right) = \Omega(\log T / \log\log T)$

Fan Zhou 2024 derive $O\left(\sqrt{T/N}\right)$
for IID integer demands

- Policy class Π **shatters** samples $\boldsymbol{d}_1, \dots \boldsymbol{d}_m$ with **witnesses** $\tau_1, \dots, \tau_m$, if for all $G \subseteq \{1, \dots, m\}$ ("Good"), there exists $\pi \in \Pi$ such that

$$\ell(\pi, \boldsymbol{d}_i) \le \tau_i \ \forall i \in G;$$
$$\ell(\pi, \boldsymbol{d}_i) > \tau_i \ \forall i \notin G$$

- **Pseudo-dimension** of Π, or $\mathrm{PDim}(\Pi)$, is the maximum of samples that it can shatter with witnesses

(S^t) Policies: $\mathrm{GE}\left(\boldsymbol{\Pi}_{(\mathbf{S}^t)}\right) = \boldsymbol{O}\left(\sqrt{\mathbf{1}/\boldsymbol{N}}\right)$

$\underbrace{\mathbb{E}\left[\sup_{\pi \in \Pi}\left(L(\pi) - \hat{L}(\pi)\right)\right]}_{\text{Generalization Error (GE)}} \leq$ **Rademacher Complexity** $\leq$ **Rademacher for Inventory Level** $\leq$ **γ-shattering dimension $\mathbf{PDim}_\gamma(\boldsymbol{Y}(\boldsymbol{\Pi}))$**

Lipschitz, Talagrand

Ending inventory by Π

γ-shattering dimension:

- $\boldsymbol{Y}(\boldsymbol{\Pi})$ **shatters** samples $\boldsymbol{d}_1, \ldots \boldsymbol{d}_m$ with **witnesses** $\tau_1, \ldots, \tau_m$ at scale $\gamma > 0$, if for all $G \subseteq \{1, \ldots, m\}$ ("Good"), there exists $\pi \in \Pi$ such that

  Ending inventory
  $$\begin{cases} \mathrm{y}(\pi, \boldsymbol{d}_i) \leq \tau_i - \gamma \ \forall i \in G; \\ \mathrm{y}(\pi, \boldsymbol{d}_i) > \tau_i + \gamma \ \forall i \notin G \end{cases}$$

- $\mathbf{PDim}_\gamma$ is the maximum m of samples that it can shatter with witnesses at scale γ

$\mathbf{PDim}_\gamma\left(\Pi_{(s^t)}\right) = O(1/\gamma)$: Proof Sketch I

We prove: max # of trajectories does not grow with T

Necessary

Demand trajectories

policies

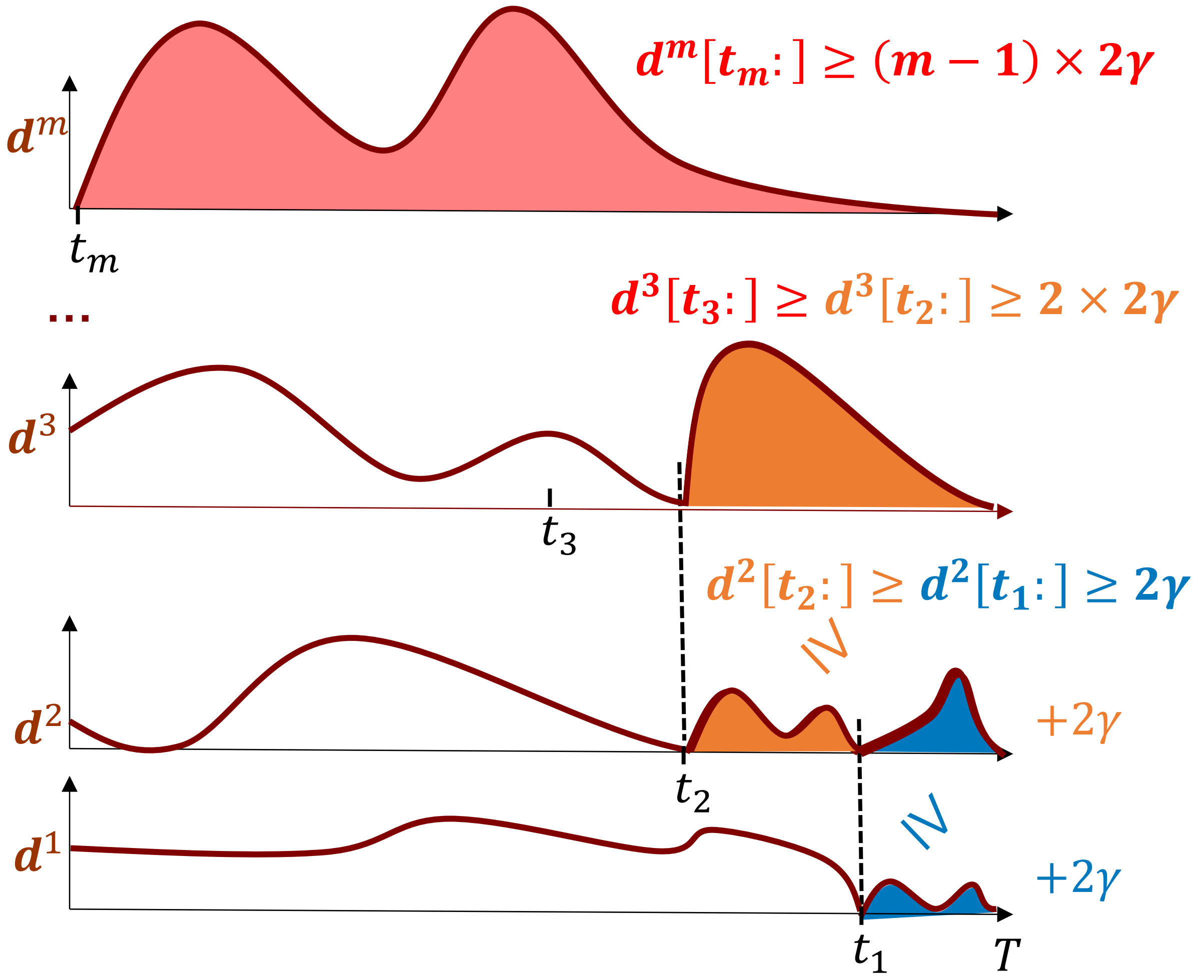
Ending Inventory	$\boldsymbol{\pi}^1$	$\boldsymbol{\pi}^2$	$\boldsymbol{\pi}^3$	…	$\boldsymbol{\pi}^m$
$\boldsymbol{d}^1$	$> \tau_1 + \gamma$	$\leq \tau_1 - \gamma$	$\leq \tau_1 - \gamma$		$\leq \tau_1 - \gamma$
$\boldsymbol{d}^2$	$\leq \tau_2 - \gamma$	$> \tau_2 + \gamma$	$\leq \tau_2 - \gamma$		$\leq \tau_2 - \gamma$
$\boldsymbol{d}^3$	$\leq \tau_3 - \gamma$	$\leq \tau_3 - \gamma$	$> \tau_3 + \gamma$		$\leq \tau_3 - \gamma$
…					
$\boldsymbol{d}^m$	$\leq \tau_m - \gamma$	$\leq \tau_m - \gamma$	$\leq \tau_m - \gamma$		$> \tau_m + \gamma$

- Assume $\tau_1 = \cdots = \tau_m =: \tau$ for proof sketch
- Let t_i be the last replenishment point for π^i on $\boldsymbol{d}^i$, and $d^i[t_i:] \coloneqq \sum_{t'=t}^{T} d_{t'}^i$ be demand from that point onward

$$\pi_{t_i}^i - \boldsymbol{d}^i[t_i:] = y_T(\pi^i, \boldsymbol{d}^i) > \tau + \gamma$$
$$\pi_{t_i}^i - \boldsymbol{d}^j[t_i:] \leq y_T(\pi^i, \boldsymbol{d}^j) \leq \tau - \gamma \ \ \forall j \neq i$$

$$\Longrightarrow \boldsymbol{d}^j[t_i:] - \boldsymbol{d}^i[t_i:] > 2\gamma \ \ \forall j \neq i$$

$\mathbf{PDim}_\gamma\left(\Pi_{(S^t)}\right) = \mathbf{O}(\mathbf{1}/\boldsymbol{\gamma})$: Proof Sketch II



Re-index so that $t_m < \cdots < t_1$,

$d^j[t_i:] > d^i[t_i:] + 2\gamma \quad \forall i \neq j$ ★

But $1 \geq \pi^m_{t_m} \geq d^m[t_m:]$ because t_m is a last replenishment point on d^m!

Hence can shatter at most $m = O(1/\gamma)$ trajectories, for any $\gamma > 0$.

(S^t) Policies: "Horizon-free" Learning Guarantees

Data-driven inventory: Empirical dynamic program or Genetic RL

Episodic RL: Finite state and action spaces

Stochastic Optimization - SAA for T-dim decision space: Interpolation or Covering number

Independent Demands	Arbitrary Demands
Levi Roundy Shmoys 07: $O\left(\sqrt{T^5 \log T / N}\right)$ Cheung Simchi-Levi 19: $O\left(\sqrt{T^6 \log T / N}\right)$ Qin Simchi-Levi Zhu 23: $O\left(\sqrt{T/N}\right)$	Shapiro Dentcheva Ruszczynski 09: $O\left(\sqrt{T/N}\right)$ Shalev-Shwartz Shamir Srebro Sridharan 10: $O\left(\sqrt{T/N}\right)$
Yin Bai Wang 21: $O\left(\sqrt{T \log T / N}\right)$ Zhang Ji Du 22: $O\left(\sqrt{T \log N / N}\right)$	
Our Paper: $O\left(\sqrt{1/N}\right)$	

VC Theory

Conclusion

Our paper provides the theoretical analysis for the **supervised learning** framework:

we use **VC theory** to analyze the **generalization error/sample complexity** of **inventory** policies.

Policy Class	Lower Bound	Upper Bound
S Policies	$\Omega\left(\sqrt{1/N}\right)$	$\boldsymbol{O}\left(\sqrt{\mathbf{1}/\boldsymbol{N}}\right)$
(s,S) Policies	$\boldsymbol{\Omega}\left(\sqrt{\mathbf{log}\,\boldsymbol{T}\,/\,\mathbf{log}\,\mathbf{log}\,\boldsymbol{T}\,/\boldsymbol{N}}\right)$	$\boldsymbol{O}\left(\sqrt{\mathbf{log}\,\boldsymbol{T}\,/\boldsymbol{N}}\right)$
(S^t) Policies	$\Omega\left(\sqrt{1/N}\right)$	$\boldsymbol{O}\left(\sqrt{\mathbf{1}/\boldsymbol{N}}\right)$

Arbitrary demand

Surprising improvement from literature

Independent demand

- VC theory is a powerful tool.
- The number of policy parameters may not be the measure of overfitting.

Thanks for Your Attention!

VC Theory for Inventory Policies
Yaqi Xie, Will Ma, Linwei Xin
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4794903