Machine Learning with Kernels for Portfolio Valuation and Risk Management

Damir Filipović
joint with Lotfi Boudabsa

École Polytechnique Fédérale de Lausanne
Swiss Finance Institute

Bachelier Finance Society One World Seminar
October 8, 2020

Preprint available online

- https://ssrn.com/abstract=3401539
- https://arxiv.org/abs/1906.03726

What this talk is about

- Develop a computational framework for quantitative portfolio risk management, based on dynamic value process
- Approximate and learn value process from a finite sample (learning from examples = machine learning) using kernel methods
- Rigorous implementation of kernel methods for simulation based approximation of functions

Main results

- Representer theorem yields closed form conditional expectations
- Central limit theorem in function space
- Finite sample guarantees thanks to weighted sampling
- Good quantitative results with relatively small training sample size
- "Regress-later" outperforms "regress-now"

Outline

Outline

Most economic scenario generators are of this form

- Economy with finite time horizon T
- $\mathbb{R}^d$-valued random driver $X = (X_1, \dots, X_T)$ with independent X_t's
- All financial values and cash flows discounted by some numeraire
- $\mathbb{Q}$ corresponding risk-neutral measure
- Physical measure $\mathbb{P} \sim \mathbb{Q}$

Dynamic value process

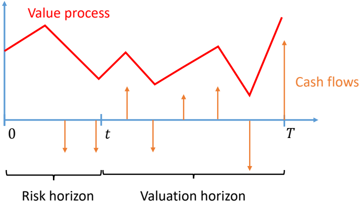
- **Objective:** a portfolio whose present value is not directly observable but has to be derived from its **cumulative cash flow** at T

$$f(X) = \sum_{t=1}^{T} \zeta_t$$

with time-t cash flows $\zeta_t = \zeta_t(X_1, \ldots, X_t) \in L^2_{\mathbb{Q}}$

- **Examples:** options, insurance liabilities, mortgage-backed instruments

- **Goal:** find **value process** of the portfolio, i.e., the $L^2_{\mathbb{Q}}$-martingale



$$V_t = \mathbb{E}_t^{\mathbb{Q}}[f(X)] := \mathbb{E}^{\mathbb{Q}}[f(X) \mid X_1, \ldots, X_t] = \underbrace{\sum_{s=1}^{t} \zeta_s}_{\text{cumulative CF at } t} + \underbrace{\mathbb{E}_t^{\mathbb{Q}}[\sum_{s=t+1}^{T} \zeta_s]}_{\text{time-}t\text{ value}}$$

Application: portfolio risk management

- **Risk measurement:** calculate $\mathbb{P}$-risk measure of 1-year P&L

$$\Delta V_{t+1} = V_{t+1} - V_t$$

E.g., (conditional) VaR or ES

- **Hedging:** given tradable financial instruments with gains processes G, i.e., $L^2_{\mathbb{Q}}$-martingales, find predictable hedging strategy ψ such that

$$\psi_t^\top \Delta G_t \approx \Delta V_t$$

E.g., variance minimizing

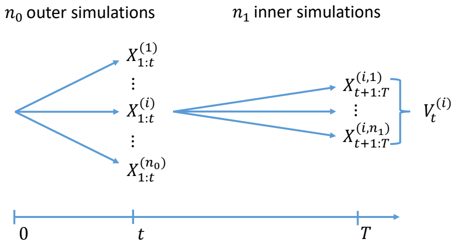
$$\psi_t = \mathbb{E}^{\mathbb{Q}}_{t-1}[\Delta G_t \Delta G_t^\top]^{-1} \mathbb{E}^{\mathbb{Q}}_{t-1}[\Delta G_t \, \Delta V_t]$$

- This is a **universal framework** for financial models (incl. GARCH)

Problem

How to compute value process $V_t = \mathbb{E}_t^{\mathbb{Q}}[f(X)]$, subject to

- lack of analytic solution: requires simulations



- limited computing budget: in realistic contexts train on small datasets, i.e., 1,000 to 32,000 paths
- Training with small datasets is vulnerable to overfitting

Machine learning approach—"Regress later"

1. **Approximate** f by some f_λ in $L^2_{\mathbb{Q}}$, for regularization parameter $\lambda > 0$
2. **Learn** f_λ from a finite sample $\boldsymbol{x} = (x(1), \dots, x(n))$, drawn from some sampling measure $\widetilde{\mathbb{Q}} \sim \mathbb{Q}$, along with function values $f(x(1)), \dots, f(x(n))$, gives sample estimator of the form

$$f_{\boldsymbol{x}}(x) = \sum_{j=1}^n \phi_j(x)\beta_j$$

such that $\mathbb{E}_t^{\mathbb{Q}}[\phi_j(X)]$ are given **in closed form**

Result: estimated value process **in closed form**:

$$V_{\boldsymbol{x},t} = \mathbb{E}_t^{\mathbb{Q}}[f_{\boldsymbol{x}}(X)], \quad t = 0, \dots, T$$

Basic estimation error bound

How good is this estimator? According to Doob's inequality, we have

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\max_{t\le T}|V_t - V_{\boldsymbol{x},t}|\right] &\le \|\tfrac{d\mathbb{P}}{d\mathbb{Q}}\|_{2,\mathbb{Q}}\,\mathbb{E}^{\mathbb{Q}}\left[\max_{t\le T}|V_t - V_{\boldsymbol{x},t}|^2\right]^{1/2} \\
&\le 2\|\tfrac{d\mathbb{P}}{d\mathbb{Q}}\|_{2,\mathbb{Q}}\,\|f - f_{\boldsymbol{x}}\|_{2,\mathbb{Q}} \\
&\le 2\|\tfrac{d\mathbb{P}}{d\mathbb{Q}}\|_{2,\mathbb{Q}}\Big(\underbrace{\|f - f_\lambda\|_{2,\mathbb{Q}}}_{\text{approximation error}} + \underbrace{\|f_\lambda - f_{\boldsymbol{x}}\|_{2,\mathbb{Q}}}_{\text{sample error}}\Big)
\end{aligned}$$

Outline

Hypothesis space

- $(E, \mathcal{E}, \mathbb{Q})$ probability space, e.g., path space $E = \mathbb{R}^{d \times T}$
- Payoff function $f : E \to \mathbb{R}$, $f \in L^2_{\mathbb{Q}}$
- **Goal:** approximate f by some f_λ in a **hypothesis space** $\mathcal{H}$
- Convenient choice of $\mathcal{H}$: a reproducing kernel Hilbert space ...

Reproducing kernel Hilbert space

Definition 2.1.

A function $k : E \times E \to \mathbb{R}$ is a **kernel** if for any $n \in \mathbb{N}$ and $x_1, \ldots, x_n \in E$ the $n \times n$-matrix $k(x_i, x_j)$ is symmetric and positive semidefinite.

Definition 2.2.

A Hilbert space $\mathcal{H}$ of functions $h : E \to \mathbb{R}$ is a **reproducing kernel Hilbert space (RKHS)** if, for any $x \in E$ there exists $k_x \in \mathcal{H}$ such that

$$\langle h, k_x \rangle_{\mathcal{H}} = h(x), \quad h \in \mathcal{H}.$$

The kernel $k(x, y) = \langle k_x, k_y \rangle_{\mathcal{H}}$ is the **reproducing kernel** of $\mathcal{H}$.

Theorem 2.3 (Moore).

For any kernel $k : E \times E \to \mathbb{R}$ there exists a unique RKHS $\mathcal{H}$ such that $k(\cdot, x) \in \mathcal{H}$ and $\langle h, k(\cdot, x) \rangle_{\mathcal{H}} = h(x)$ for all $h \in \mathcal{H}$ and $x \in E$.

Basic assumptions

- Let $k : E \times E \to \mathbb{R}$ be a measurable kernel with a separable RKHS $\mathcal{H}$
- Then $\mathcal{H}$ consists of measurable functions
- Define $\kappa(x) = \sqrt{k(x,x)} = \|k(\cdot,x)\|_{\mathcal{H}}$, so that

$$|h(x)| \leq \kappa(x)\|h\|_{\mathcal{H}}, \quad h \in \mathcal{H}$$

- **Assume:** $\|\kappa\|_{2,\mathbb{Q}} < \infty$
- Define embedding $J : \mathcal{H} \to L^2_{\mathbb{Q}}$ by $Jh = \mathbb{Q}$-equivalence class of h

Approximation

Approximate f by f_λ in $L^2_{\mathbb{Q}}$, where $h = f_\lambda \in \mathcal{H}$ solves

$$\min_{h\in\mathcal{H}}(\|f-h\|^2_{2,\mathbb{Q}} + \lambda\|h\|^2_{\mathcal{H}}), \tag{2.1}$$

for some regularization parameter $\lambda > 0$:

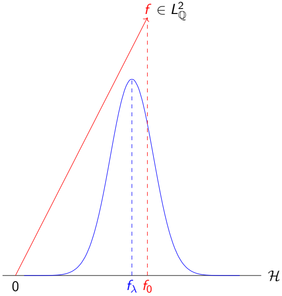


Figure: f_λ with sample error distribution.

Existence, uniqueness, and representer theorem

1. There exists a unique solution $h = f_\lambda \in \mathcal{H}$ to (2.1), and it is given by
$$f_\lambda = (J^*J + \lambda)^{-1} J^* f.$$
2. **[Representer theorem]** The solution f_λ can be represented as
$$f_\lambda = J^* g_\lambda = \int_E k(\cdot, x) g_\lambda(x) \mathbb{Q}(dx)$$
where
$$g_\lambda = (JJ^* + \lambda)^{-1} f.$$

Corollary of representer theorem

Definition 2.4.

We call the kernel k **tractable** if the conditional kernel embeddings

$$M_t(y) = \mathbb{E}_t^{\mathbb{Q}}[k(X, y)]$$

are given in closed form, for all y, t.

Lemma 2.5.

Assume that k is tractable. Then

$$\mathbb{E}_t^{\mathbb{Q}}[f_\lambda(X)] = \int_E M_t(y) g_\lambda(y) \mathbb{Q}(dy)$$

is given in closed form, subject to $\mathbb{Q}$-integration, for all t.

Projection and universal kernels

- Let $f_0 = \mathrm{Proj}_{\overline{\mathrm{Im}\, J}} f$ be the orthogonal projection of f onto $\overline{\mathrm{Im}\, J}$ in $L^2_{\mathbb{Q}}$
- Orthogonality: decompose the squared **approximation error**

$$\|f - f_\lambda\|^2_{2,\mathbb{Q}} = \underbrace{\|f - f_0\|^2_{2,\mathbb{Q}}}_{\text{"best approximation error"}} + \underbrace{\|f_0 - f_\lambda\|^2_{2,\mathbb{Q}}}_{\text{"regularization error"}}.$$

Lemma 2.6.

$\|f_0 - f_\lambda\|_{2,\mathbb{Q}} \to 0$ *as* $\lambda \to 0$ *(but can be slow).*

- Desirable kernel property is $\|f - f_0\|^2_{2,\mathbb{Q}} = 0$:

Definition 2.7.

The kernel k is called **$L^2_{\mathbb{Q}}$-universal** if $\overline{\mathrm{Im}\, J} = L^2_{\mathbb{Q}}$.

Outline

Random sample

- Let $\boldsymbol{x} = (x(1), \ldots, x(n))$ be i.i.d. E-valued rv's with $x(i) \sim \mathbb{Q}$
- Define the empirical measure $\mathbb{Q}_{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^n \delta_{x(i)}$
- All results above apply for $\mathbb{Q}_{\boldsymbol{x}}$ in lieu of $\mathbb{Q}$
- Denote by $J_{\boldsymbol{x}} : \mathcal{H} \to L^2_{\mathbb{Q}_{\boldsymbol{x}}}$ the sample version of $J : \mathcal{H} \to L^2_{\mathbb{Q}}$
- Obtain $\mathcal{H}$-valued sample estimator

$$f_{\boldsymbol{x}} = (J_{\boldsymbol{x}}^* J_{\boldsymbol{x}} + \lambda)^{-1} J_{\boldsymbol{x}}^* f = J_{\boldsymbol{x}}^* \underbrace{(J_{\boldsymbol{x}} J_{\boldsymbol{x}}^* + \lambda)^{-1} f}_{= g_{\boldsymbol{x}} \in L^2_{\mathbb{Q}_{\boldsymbol{x}}}}$$

How to compute $f_{\boldsymbol{x}}$

- Denote $\boldsymbol{f} = (f(x(1)), \dots, f(x(n)))^\top$
- Define $n \times n$-kernel matrix $\boldsymbol{K}_{ij} = k(x(i), x(j))$
- $\frac{1}{n}\boldsymbol{K}$ is matrix representation of $J_{\boldsymbol{x}} J_{\boldsymbol{x}}^* : L^2_{\mathbb{Q}_x} \to L^2_{\mathbb{Q}_x}$

Corollary 3.1 (of representer theorem).

The unique solution $\boldsymbol{g} \in \mathbb{R}^n$ *to*

$$(\tfrac{1}{n}\boldsymbol{K} + \lambda)\boldsymbol{g} = \boldsymbol{f}$$

gives $f_{\boldsymbol{x}} = \frac{1}{n}\sum_{j=1}^n k(\cdot, x(j))\boldsymbol{g}_j$*. If, moreover, the kernel* k *is tractable then*

$$V_{\boldsymbol{x},t} = \frac{1}{n}\sum_{j=1}^n M_t(x(j))\boldsymbol{g}_j$$

is given in closed form.

Law of large numbers and central limit theorem

Theorem 3.2.

Assume that $\|\kappa\|_{4,\mathbb{Q}} < \infty$ *and* $\|f\kappa\|_{2,\mathbb{Q}} < \infty$.

1. *Law of large numbers:* $f_{\boldsymbol{x}} \xrightarrow{a.s.} f_\lambda$ *as* $n \to \infty$
2. *Central limit theorem:*

$$\sqrt{n}(f_{\boldsymbol{x}} - f_\lambda) \xrightarrow{d} \mathcal{N}(0, Q) \quad \text{as } n \to \infty$$

with covariance operator $Q : \mathcal{H} \to \mathcal{H}$ *given by*

$$\langle Qh, h\rangle_{\mathcal{H}} = \mathbb{V}_{\mathbb{Q}}[(f - f_\lambda)(J^*J + \lambda)^{-1}h], \quad h \in \mathcal{H}.$$

Finite sample guarantee

Theorem 3.3.

Assume that

$$\|f\|_{\infty,\mathbb{Q}} < \infty \quad \text{and} \quad \|\kappa\|_{\infty,\mathbb{Q}} < \infty. \tag{3.1}$$

For any $\eta \in (0,1]$, *we have*

$$\|f_{\boldsymbol{x}} - f_\lambda\|_{\mathcal{H}} < \frac{2\sqrt{2\log(2/\eta)}\|(f - f_\lambda)\kappa\|_{\infty,\mathbb{Q}}}{\lambda\sqrt{n}}$$

with sampling probability of at least $1-\eta$.

Problem: Assumption (3.1) does not hold in finance in general!

Weighted sampling

- Equivalent sampling measure $\widetilde{\mathbb{Q}} \sim \mathbb{Q}$ with RN derivative $w = \frac{d\widetilde{\mathbb{Q}}}{d\mathbb{Q}}$
- Induces isometric isomorphism

$$\begin{array}{ccc} \widetilde{\mathcal{H}} & \xrightarrow{\widetilde{J}} & L^2_{\widetilde{\mathbb{Q}}} \\ {\scriptstyle \times\sqrt{w}}\Big\downarrow & & \Big\uparrow{\scriptstyle \times\frac{1}{\sqrt{w}}} \\ \mathcal{H} & \xrightarrow{J} & L^2_{\mathbb{Q}} \end{array}$$

$$\widetilde{f} = \frac{f}{\sqrt{w}} \text{ and } \widetilde{k}(x, y) = \frac{k(x,y)}{\sqrt{w(x)w(y)}}$$

Solution:

1. Choose $\widetilde{\mathbb{Q}}$ such that $\|\widetilde{f}\|_{\infty,\mathbb{Q}} < \infty$ and $\|\widetilde{\kappa}\|_{\infty,\mathbb{Q}} < \infty$.
2. Learn $\widetilde{f}_{\boldsymbol{x}}$ in $\widetilde{\mathcal{H}}$ and **define** $f_{\boldsymbol{x}} := \widetilde{f}_{\boldsymbol{x}} \times \sqrt{w}$
3. Obtain finite sample guarantee extending Theorem 3.3

$$\|f_{\boldsymbol{x}} - f_\lambda\|_{\mathcal{H}} < \frac{2\sqrt{2\log(2/\eta)}\|\frac{1}{w}(f - f_\lambda)\kappa\|_{\infty,\mathbb{Q}}}{\lambda\sqrt{n}}$$

Particular choice of sampling measure

Optimal sampling measure that minimizes $\|\widetilde{\kappa}\|_{\infty,\mathbb{Q}}$:

Lemma 3.4.

For any $\widetilde{\mathbb{Q}} \sim \mathbb{Q}$, we have $\|\widetilde{\kappa}\|_{\infty,\mathbb{Q}} \geq \|\kappa\|_{2,\mathbb{Q}}$, with equality if and only if

$$w = \frac{\kappa^2}{\|\kappa\|_{2,\mathbb{Q}}^2}, \quad \mathbb{Q}\text{-a.s.}$$

In this case, $\widetilde{\kappa} = \|\kappa\|_{2,\mathbb{Q}}$ is constant $\mathbb{Q}$-a.s.

Solution: Choose tractable kernel k such that $\frac{f}{\kappa}$ is bounded

Outline

Product structure

- Random driver $X = (X_1, \dots, X_T)$, e.g., Brownian motion
- Path space $E = E_1 \times \cdots \times E_T$, e.g., $E = \mathbb{R}^{d \times T}$
- Product measure $\mathbb{Q}(dx) = \mathbb{Q}_1(dx_1) \times \cdots \times \mathbb{Q}_T(dx_T)$
- Product kernel $k(x, y) = k_1(x_1, y_1) \times \cdots \times k_T(x_T, y_T)$
- RKHS tensor product $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_T$

Tractable kernels

- Assume for all t: k_t is tractable, i.e., kernel embeddings

$$m_t(y) = \int_{E_t} k_t(x, y)\mathbb{Q}_t(dx)$$

  in closed form

- Then k is tractable: conditional kernel embeddings given by

$$M_t(y) = \mathbb{E}_t^{\mathbb{Q}}[k(X, y)] = \prod_{s=1}^{t} k_s(X_s, y_s) \prod_{s=t+1}^{T} m_s(y_s)$$

Lévy models ($E_t \subseteq \mathbb{R}^d$)

- Bocher's theorem: for symmetric p.m. Λ on $\mathbb{R}^d$, $\beta \geq 0$, obtain kernel

$$k_t(x,y) = \mathrm{e}^{\beta x^\top y} \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}(x-y)^\top \lambda} \Lambda(d\lambda)$$

- Lévy driver: if $\mathbb{Q}_t$ infinitely divisible then kernel embedding

$$m_t(x) = \int_{\mathbb{R}^d} \widehat{\mathbb{Q}}_t(\beta x + \mathrm{i}\lambda) \mathrm{e}^{-\mathrm{i}x^\top \lambda} \Lambda(d\lambda)$$

  is in closed form subject to Λ-integration

- Lévy kernels: if Λ is infinitely divisible then kernel is in closed form

$$k_t(x,y) = \mathrm{e}^{\beta x^\top y} \mathrm{e}^{-\frac{1}{2}(x-y)^\top A(x-y) + \int_{\mathbb{R}^d} (\cos((x-y)^\top \xi) - 1)\nu(d\xi)}$$

- Example: **Gaussian-exponentiated kernel**, $\alpha \geq 0$,

$$k_t(x,y) = \mathrm{e}^{\beta x^\top y - \alpha \|x-y\|^2}$$

Outline

Gaussian white noise $\mathbb{Q} = N(0, I_{d\times T})$

Gaussian-exponentiated kernel, for $\alpha \geq 0$, $\beta \in [0, 1/2)$,

$$k(x, y) = \mathrm{e}^{-\alpha\|x-y\|^2+\beta x^\top y}$$

with RN derivative, for $\gamma < 1/2$,

$$w(x) = (1-2\gamma)^{dT/2}\mathrm{e}^{\gamma\|x\|^2}$$

Then:

- kernel is tractable: $m_s(y_s) = (1+2\alpha)^{-d/2}\mathrm{e}^{\frac{\beta^2+4\alpha\beta-2\alpha}{4\alpha+2}\|y_s\|^2}$
- sampling measure is normal: $\widetilde{\mathbb{Q}} = N(0, \frac{1}{1-2\gamma} I_{d\times T})$
- $\widetilde{k}$ is $L^2_{\widetilde{\mathbb{Q}}}$-universal: $\overline{\mathrm{Im}\,\widetilde{J}} = L^2_{\widetilde{\mathbb{Q}}}$ and hence $f_0 = f$
- $\|\widetilde{\kappa}\|_{\infty,\mathbb{Q}} < \infty \Leftrightarrow \beta \leq \gamma$ (Lemma 3.4: $\beta = \gamma$ optimal)

Black–Scholes model

- T time steps at length Δ_t [years], $t = 1, \ldots, T$
- $X = (X_1, \cdots, X_T) \sim N(0, I_{d \times T})$ Gaussian white noise
- Black–Scholes discrete time model with d nominal price processes

$$S_{i,t} = S_{i,t-1} \mathrm{e}^{(r+\sigma_i^2/2)\Delta_t + \sigma_i \sqrt{\Delta_t} X_{i,t}}, \quad S_{i,0} = 1$$

- Volatility parameters $\sigma_i = 0.2$
- Risk-free rate $r = 0$

European max-call and min-put options

- $T = 2$ time steps at length $\Delta_1 = 1/12$ and $\Delta_2 = 11/12$ [years]
- Max-call option has discounted payoff at T

$$f(X) = \mathrm{e}^{-rT}(\max_i S_{i,T} - K)^+$$

- Min-put option has discounted payoff at T

$$f(X) = \mathrm{e}^{-rT}(K - \min_i S_{i,T})^+$$

- Strike price $K = S_{i,0} = 1$
- Number of stocks $d = 6$
- **Total dimension** $d \times T = 12$

Barrier reverse convertible

- $T = 12$ time steps at length $\Delta_t = 1/12$ [years]
- Barrier reverse convertible has discounted payoff at T

$$f(X) = \mathrm{e}^{-rT} C + \mathrm{e}^{-rT} F \left(1 - 1_{\{\min_{i,t} S_{i,t} \le B\}} \left(1 - \min_i \frac{S_{i,T}}{S_{i,0} K} \right)^+ \right)$$

- Face value $F = 1$
- Coupon $C = 0$
- Barrier $B = 0.6$
- Strike price $K = S_{i,0} = 1$
- Number of underlying prices $d = 3$
- **Total dimension** $d \times T = 36$

Numerical results

- Training sample size n varies between $4,000$ and $20,000$
- Hyperparameters α, β, λ via GPR
- Compare true and estimated value process

$$V_t = \mathbb{E}_t^{\mathbb{Q}}[f(X)], \quad V_{\mathbf{x},t} = \mathbb{E}_t^{\mathbb{Q}}[f_{\mathbf{x}}(X)]$$

- Ground truth: nested MC with 10^5 outer and 10^3 inner simulations

Below we report

- normalized $L^2_{\mathbb{Q}}$-errors $\frac{\|V_t - V_{\mathbf{x},t}\|_{2,\mathbb{Q}}}{V_0}$ in % for $t = 0, 1, T$
- detrended Q-Q plots of $V_{\mathbf{x},t}$ against V_t for $t = 1, T$

Optimal hyperparameter values α, β, λ from GPR

Payoff	α	β	λ
Min-put	2.06×10^{-2}	0	1.86×10^{-8}
Max-call ($\gamma = 0$)	2.53×10^{-2}	0	3.33×10^{-8}
Max-call ($\gamma = 0.15$)	2.66×10^{-2}	4.10×10^{-9}	4.14×10^{-8}
Barrier reverse convertible	2.95×10^{-3}	0	9.19×10^{-8}

- Optimal values can differ across portfolios

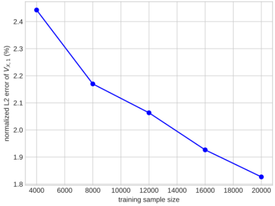
Normalized $L^2_{\mathbb{Q}}$-errors $\|V_t - V_{\mathbf{x},t}\|_{2,\mathbb{Q}}/V_0$ in %

Payoff	$t = 0$	1	T
Min-put	0.194	1.83	10.1
Max-call ($\gamma = 0$)	0.080	2.50	12.4
Max-call ($\gamma = 0.15$)	0.103	2.32	11.7
Barrier reverse convertible	0.022	0.25	5.8

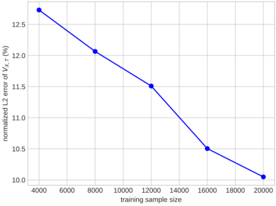
- In line with basic estimation error bound

$$\|V_t - V_{\mathbf{x},t}\|_{2,\mathbb{Q}} \leq \underbrace{\|f - f_{\mathbf{x}}\|_{2,\mathbb{Q}}}_{\text{estimate by MC}}$$

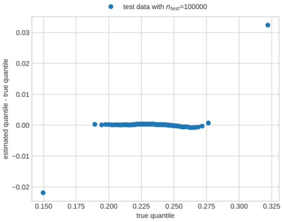
Results for the min-put



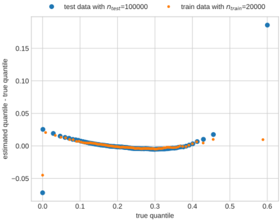
Normalized $L^2_{\mathbb{Q}}$-error of $V_{\boldsymbol{x},1}$



Normalized $L^2_{\mathbb{Q}}$-error of $V_{\boldsymbol{x},T}$

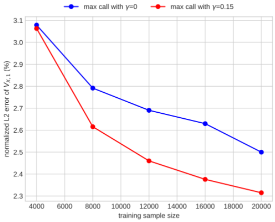


Detrended Q-Q plot of $V_{\boldsymbol{x},1}$

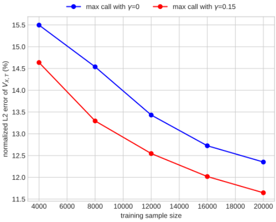


Detrended Q-Q plot of $V_{\boldsymbol{x},T}$

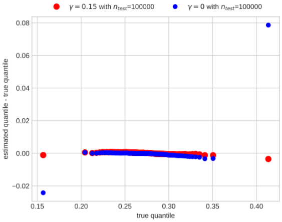
Results for the max-call with $\gamma = 0$ and $\gamma = 0.15$



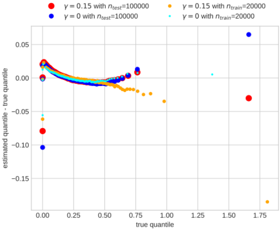
Normalized $L^2_{\mathbb{Q}}$-error of $V_{\boldsymbol{x},1}$



Normalized $L^2_{\mathbb{Q}}$-error of $V_{\boldsymbol{x},T}$

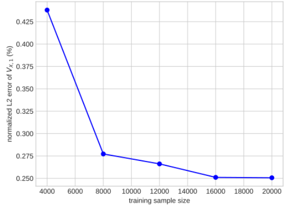


Detrended Q-Q plots of $V_{\boldsymbol{x},1}$

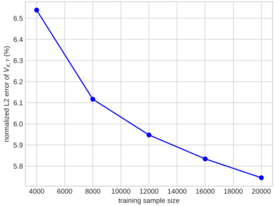


Detrended Q-Q plots of $V_{\boldsymbol{x},T}$

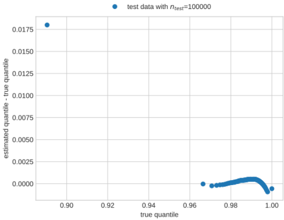
Results for the barrier reverse convertible



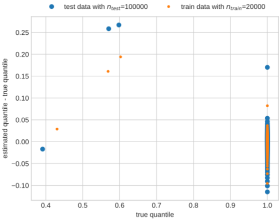
Normalized $L^2_{\mathbb{Q}}$-error of $V_{\boldsymbol{x},1}$



Normalized $L^2_{\mathbb{Q}}$-error of $V_{\boldsymbol{x},T}$



Detrended Q-Q plot of $V_{\boldsymbol{x},1}$



Detrended Q-Q plot of $V_{\boldsymbol{x},T}$

Outline

Regress-now

- Fix a risk horizon, e.g., $t = 1$
- Directly learn $V_1 = \mathbb{E}_1^{\mathbb{Q}}[f(X)]$ from "noisy" observations

$$f(X) = V_1 + \underbrace{f(X) - V_1}_{\text{noise } \epsilon}$$

- Straightforward adaptation: restrict to RKHS $\mathcal{H}_1$ with kernel k_1 on E_1
- Leads to regress-now estimator $V_{\boldsymbol{x},1}^{\text{now}} = \frac{1}{n}\sum_{j=1}^{n} k_1(X_1, x(j))\boldsymbol{g}_j^{\text{now}}$
- Advantage: smaller state space $\dim E_1 < \dim(E_1 \times \cdots \times E_T)$
- But: dimension of regression problem remains the same, $n = 20,000$
- **Result:** regress-later performs better than regress-now, both in terms of accuracy and computing times!

Computing times

Payoff	regress-now	regress-later
Min-put	3122	2756
Max-call	3288	2570
Barrier reverse convertible	3683	2756

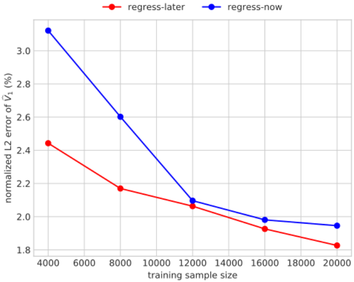
Table: Computing times (in seconds) for estimating $V_{x,1}^{\text{now}}$ and $V_{x,1}$, for sample size $n = 20,000$, on Skylake processors running at 2.3 GHz, using 14 cores and 100 GB of RAM.

Normalized $L^2_{\mathbb{Q}}$-errors $\|V_1 - \widehat{V}_1\|_{2,\mathbb{Q}}/V_0$ in %

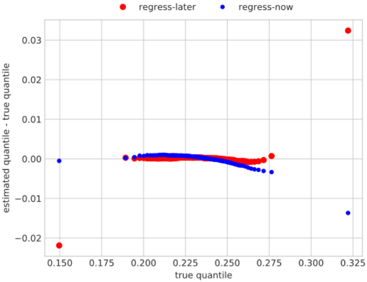
Payoff	regress-now	regress-later
Min-put	1.95	1.83
Max-call	2.61	2.50
Barrier reverse convertible	0.281	0.25

Table: Normalized $L^2_{\mathbb{Q}}$-error $\|V_1 - \widehat{V}_1\|_{2,\mathbb{Q}}/V_0$ in % for $\widehat{V}_1 \in \{V^{\text{now}}_{x,1}, V_{x,1}\}$

Results for the min-put

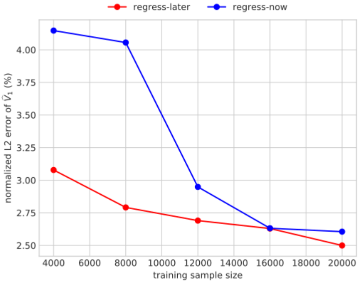


normalized $L^2_{\mathbb{Q}}$-errors of $\widehat{V}_1$ in %

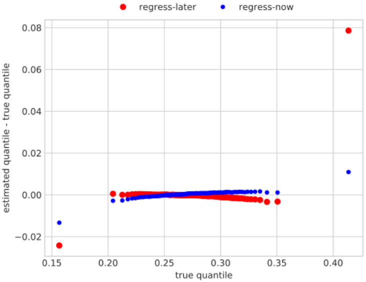


detrended Q-Q plots of $\widehat{V}_1$

Results for the max-call (with $\gamma = 0$)

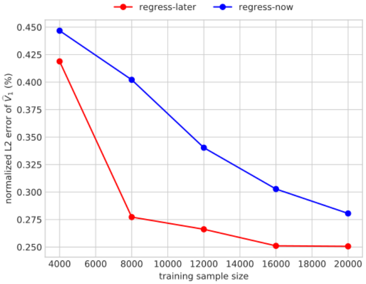


normalized $L^2_{\mathbb{Q}}$-errors of $\widehat{V}_1$ in %

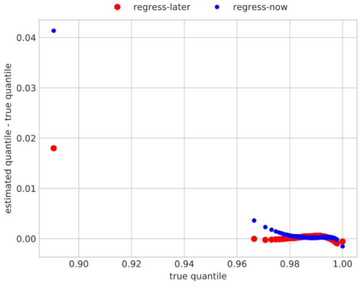


detrended Q-Q plots of $\widehat{V}_1$

Results for the barrier reverse convertible



normalized $L^2_{\mathbb{Q}}$-errors of $\widehat{V}_1$ in %



detrended Q-Q plots of $\widehat{V}_1$

Conclusion

- Computational framework for quantitative portfolio risk management
- Machine learning with kernels gives accurate estimates of the dynamic value process for relatively small training sample size
- Weighted sampling further improves results (finite sample guarantees)
- Regress-later outperforms regress-now
- Scalability limits at dimensions beyond $d \times T = 100$