Kernel-based nonparametric tests for shape constraints

Rohan Sen

Università della Svizzera italiana, Lugano

December 14, 2025

Motivation

- Economic theory motivates shape restrictions: monotone or convex stochastic discount factor.
- Learning without parametric or structural assumptions.
- Learn flexibly without imposing shape restrictions during estimation procedure.
- Balancing reward and risk: mean-variance / Sharpe-optimal objective.
- Need global, scalable inference applicable to large datasets.

Research goal

- Build a flexible learning setup under a mean-variance objective.
- Inference pipeline that can estimate economically relevant functions.
- Test their global shape restrictions (monotonicity/convexity) with statistical guarantees.

Contributions

- Nonparametric learning framework under mean-variance objective that depends on a function and its derivatives.
- Finite-dimensional representation space for the optimal solution.
- Statistical guarantees — consistency, asymptotic normality, finite-sample deviation bound (high-probability).
- Joint Wald-type test statistic for shape inference over grids.
- Scalable computational procedure applicable for large datasets.

Related literature

- *Shape inference*: Shapiro (1985), Wolak (1987), Ghosal et al. (2000), Juditsky and Nemirovski (2002), and Birke and Neumeyer (2013).

- *Shape-constrained estimation*: Seijo and Sen (2011), Groeneboom and Jongbloed (2014), Marteau-Ferey et al. (2020), Muzellec et al. (2021), and Aubin-Frankowski and Szabo (2022).

- *Statistical learning*: Schölkopf et al. (2001), Cucker and Smale (2001), Caponnetto and De Vito (2007), Zhou (2008), Alaoui and Mahoney (2015), and Filipović and Schneider (2025).

- *Financial econometrics*: Kozak (2020), Boudabsa and Filipović (2022), Filipović, Pelger, et al. (2022), Filipović and Schneider (2024), and Luzzi et al. (2025).

Outline

1. **Methodology**
2. **Statistical properties**
3. **Inference for shape properties**
4. **Application**
5. **References**

Methodology

- Specify admissible space of nonlinear functions as a Sobolev-type *reproducing kernel Hilbert space (RKHS)* $\mathcal{H}$ that contains sufficiently smooth functions.

$$\mathcal{H} \subset \mathscr{C}^s(\mathcal{X}), \qquad \langle f, g \rangle_{\mathcal{H}} := \sum_{|\boldsymbol{\alpha}| \leq s} \langle \partial^{\boldsymbol{\alpha}} f, \partial^{\boldsymbol{\alpha}} g \rangle_{L^2}.$$

- Let $\boldsymbol{z} = (x, y) \in \mathcal{X} \times \mathcal{Y}$ be distributed according to underlying probability $\mathbb{P}$.

- Target functional *linear* in unknown smooth function $h \in \mathcal{H}$ and its derivatives:

$$\mathcal{R}(h; \boldsymbol{z}) := \sum_{|\boldsymbol{\alpha}| \leq s} w_{\boldsymbol{\alpha}}(\boldsymbol{z})\, \partial^{\boldsymbol{\alpha}} h(\boldsymbol{x}).$$

Mean-variance optimization

- Population objective (Tikhonov regularized):

$$h_\lambda \;:=\; \operatorname*{argmin}_{h \in \mathcal{H}} \left\{ -\mathbb{E}[\mathcal{R}(h; \boldsymbol{z})] + \tfrac{1}{2}\mathbb{V}[\mathcal{R}(h; \boldsymbol{z})] + \tfrac{\lambda}{2}\|h\|_{\mathcal{H}}^2 \right\}. \tag{1}$$

- Data $\{\boldsymbol{z}_i = (x_i, y_i)\}_{i=1}^N \sim \mathbb{P}$; empirical distribution $\widehat{\mathbb{P}} := \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{z}_i}$.

- Empirical objective (Tikhonov regularized):

$$\widehat{h}_\lambda \;:=\; \operatorname*{argmin}_{h \in \mathcal{H}} \left\{ -\widehat{\mathbb{E}}[\mathcal{R}(h; \boldsymbol{z})] + \tfrac{1}{2}\widehat{\mathbb{V}}[\mathcal{R}(h; \boldsymbol{z})] + \tfrac{\lambda}{2}\|h\|_{\mathcal{H}}^2 \right\}. \tag{2}$$

- Problems 1 and 2 are convex: mean $\rightarrow$ reward, variance $\rightarrow$ risk penalty, regularization $\rightarrow$ complexity control.

Optimal representation

- Linearity implies mean and variance terms can be embedded in RKHS $\mathcal{H}$.
- *Representer theorem*: The optimal solution to Problem 2 has the form

$$\widehat{h}_\lambda = \sum_{i=1}^{N} \sum_{|\boldsymbol{\alpha}| \leq s} \widehat{c}_{i,\boldsymbol{\alpha}} \, \partial^{\boldsymbol{\alpha}} \phi(x_i). \tag{3}$$

- Optimal working subspace: $\text{span}\{\partial^{\boldsymbol{\alpha}} \phi(x_i) : |\boldsymbol{\alpha}| \leq s, \, 1 \leq i \leq N\}$.
- Nonparametric infinite-dimensional optimization becomes solving for *finitely* many coefficients — solved via quadratic program.

Asymptotic properties

Under some *regularity assumptions*, the following hold:

- *Asymptotic consistency*:

$$\widehat{h}_\lambda \xrightarrow{a.s.} h_\lambda \quad \text{as} \quad N \to \infty.$$

- *Asymptotic normality*:

$$\sqrt{N}\left(\widehat{h}_\lambda - h_\lambda\right) \xrightarrow{d} \mathcal{N}(0, \mathcal{C}_\lambda),$$

where $\mathcal{C}_\lambda \colon \mathcal{H} \to \mathcal{H}$ is a *covariance operator*.

Finite-sample properties

For any $\delta \in (0,1)$, it holds with sampling probability at least $(1-\delta)$:

- *Finite-sample deviation bound*:

$$\|\widehat{h}_\lambda - h_\lambda\|_{\mathcal{H}} \;\leq\; C_{FS}(\delta, \|h_\lambda\|)\,\lambda^{-1}\,N^{-1/2},$$

  where C_{FS} is a positive coefficient.

- High-probability bound matching Monte Carlo rate up to regularization.

- Estimation error depends on *confidence level* δ, size of true solution h_λ, *regularization hyperparameter* λ and sample size N.

Inference for shape properties

- Test if shape property holds jointly on random grid $\mathcal{Z} := \left\{\xi_j\right\}_{j=1}^{n} \subset \mathcal{X}$.
- Object of interest: $\boldsymbol{\theta} := \left[\partial^{\boldsymbol{\alpha}} h_\lambda(\xi_j)\right]_{j=1}^{n} \in \mathbb{R}^n$.
- Hypothesis test:

$$H_0 : \boldsymbol{\theta} \geq \mathbf{0} \qquad \text{vs} \qquad H_1 : \text{there exists } j \text{ such that } \boldsymbol{\theta}_j < 0. \tag{4}$$

- *Least favorable null*: $\boldsymbol{\theta} = \mathbf{0}$ (all inequalities binding).
- Examples: $\boldsymbol{\alpha} = 0 \rightarrow$ positivity, $\boldsymbol{\alpha} = 1 \rightarrow$ monotonicity, $\boldsymbol{\alpha} = 2 \rightarrow$ convexity.

Test statistic I

- Building blocks:
  - (i) Access to object of interest: $\widehat{\boldsymbol{\theta}} := \left[\partial^{\boldsymbol{\alpha}} \widehat{h}_\lambda(\xi_j)\right]_{j=1}^n \in \mathbb{R}^n$;
  - (ii) Consistent *plug-in* covariance matrix $\widehat{\boldsymbol{\Omega}}_\lambda$ of $\boldsymbol{\Omega}_\lambda \in \mathbb{R}^{n\times n}$.
- *Test statistic*:

$$W_N := \min_{\boldsymbol{c}\geq \mathbf{0}} N\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{c}\right)^\top \widehat{\boldsymbol{\Omega}}_\lambda^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{c}\right). \tag{5}$$

- *Asymptotic distribution*: Under least favorable null $\boldsymbol{\theta} = \mathbf{0}$,

$$W_N \xrightarrow{d} W \sim \chi_n^2 - \bar{\chi}^2(\boldsymbol{\Omega}_\lambda, \mathbb{R}_+^n).$$

Test statistic II

- Test statistic W_N is the scaled *distance-to-feasibility* for least favorable null.
- Measures the *projection error* under *whitening* by $\widehat{\boldsymbol{\Omega}}_\lambda^{-1/2}$.
- Computed via *non-negative least squares*:

$$W_N \;=\; N \min_{\boldsymbol{c} \geq \boldsymbol{0}} \left\| \widehat{\boldsymbol{\Omega}}_\lambda^{-1/2} \boldsymbol{c} - \boldsymbol{b} \right\|_2^2, \qquad \boldsymbol{b} \;:=\; \widehat{\boldsymbol{\Omega}}_\lambda^{-1/2} \widehat{\boldsymbol{\theta}}.$$

- The p-values are obtained by Monte Carlo calibration.

Application: asset pricing

- Fundamental challenge in asset pricing: understand investors' risk preferences and how these shape market dynamics.

- *Stochastic discount factor (SDF)* prices assets: under *no-arbitrage*, price of any asset is the expected value of its future payoff discounted by the SDF.

- Defined by connecting physical probability $\mathbb{P}$ to *risk-neutral probability* $\mathbb{Q}$:

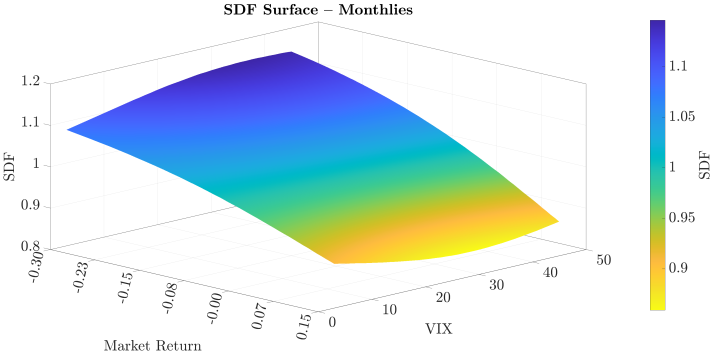
$$\widetilde{M}_t \; := \; \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t}, \qquad \mathcal{F}_t \; := \; \text{information at time } t.$$

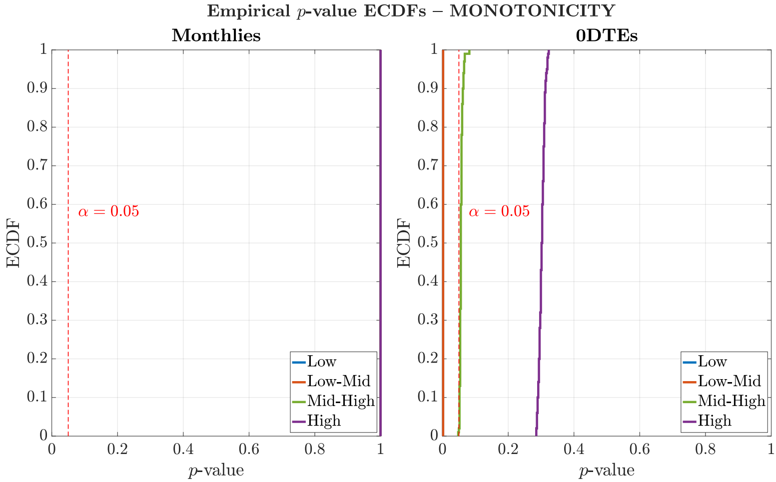
- Expected utility theory predicts SDF should be *monotonically decreasing* and *convex* in returns — proportional to marginal utility.

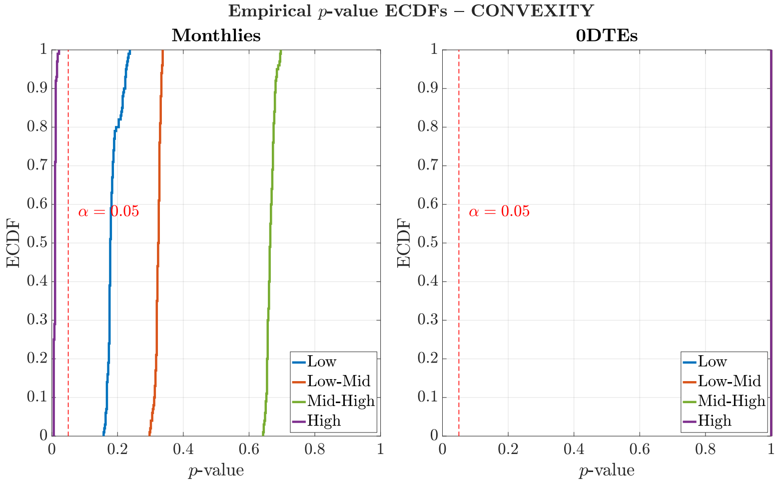
Learning the stochastic discount factor (Luzzi et al. (2025))

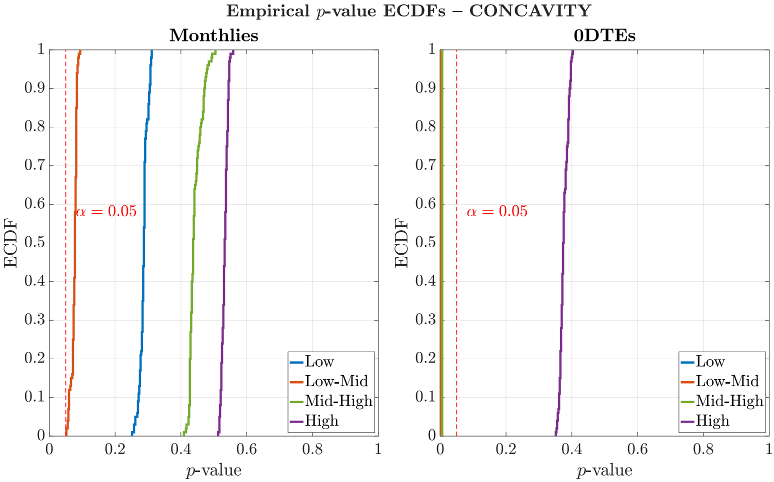
- Estimate SDF without parametric or structural assumptions using *options trading strategy* on the S&P 500.
- Trade SDF (projected onto returns) via *Carr-Madan* option portfolio.
- Equivalence between trading (shorting) the SDF and maximizing mean-variance portfolio (Hansen and Jagannathan (1991)).
- Optimal allocations in mean-variance sense identified by derivatives of SDF.
- Take random grids of market returns and volatility states: test whether estimated SDF satisfies monotonicity / convexity properties.

Plot of SDF surface (monthly options)




Empirical p-value ECDFs – MONOTONICITY
Monthlies
0DTEs
ECDF
p-value
$\alpha = 0.05$
Low
Low-Mid
Mid-High
High
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Empirical p-value ECDFs – CONVEXITY
Monthlies
0DTEs
ECDF
p-value
$\alpha = 0.05$
Low
Low-Mid
Mid-High
High
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Empirical p-value ECDFs – CONCAVITY
Monthlies
0DTEs
ECDF
p-value
$\alpha = 0.05$
Low
Low-Mid
Mid-High
High
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Key takeaways

- Maturity and volatility heterogeneity: SDF varies strongly across maturity horizons and also volatility states.
- Monthly options: SDF is near-linear and monotonically decreasing across volatility states.
- 0DTE options: Monotonicity almost always rejected; convexity is not rejected with very high *p*-values — consistent with U-shaped pattern.
- Results are robust across grids and grid sizes.

Thank you!

Link to the paper.

References I

Alaoui, A. E. and M. W. Mahoney (2015). "Fast randomized kernel ridge regression with statistical guarantees". In: *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1*. Montreal, Canada, pp. 775–783.

Aubin-Frankowski, P.-C. and Z. Szabo (2022). "Handling Hard Affine SDP Shape Constraints in RKHSs". In: *Journal of Machine Learning Research* 23.297, pp. 1–54.

Birke, M. and N. Neumeyer (2013). "Testing Monotonicity of Regression Functions – An Empirical Process Approach". In: *Scandinavian Journal of Statistics* 40.3, pp. 438–454.

References II

Boudabsa, L. and D. Filipović (2022). "Machine learning with kernels for portfolio valuation and risk management". In: *Finance and Stochastics* 26.2, pp. 131–172.

Caponnetto, A. and E. De Vito (2007). "Optimal Rates for the Regularized Least-Squares Algorithm". In: *Foundations of Computational Mathematics* 7.3, pp. 331–368.

Cucker, F. and S. Smale (2001). "On the mathematical foundations of learning". In: *Bulletin of the American Mathematical Society* 39, pp. 1–49.

Filipović, D., M. Pelger, and Y. Ye (2022). "Stripping the Discount Curve - a Robust Machine Learning Approach". In: *SSRN Electronic Journal*.

References III

Filipović, D. and P. G. Schneider (2024). "Joint Estimation of Conditional Mean and Covariance for Unbalanced Panels". In: *SSRN Electronic Journal*.

— (2025). "Kernel Density Machines". In: *SSRN Electronic Journal*.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000). "Testing Monotonicity of Regression". In: *The Annals of Statistics* 28.4, pp. 1054–1082.

Groeneboom, P. and G. Jongbloed (2014). *Nonparametric Estimation under Shape Constraints: Estimators, Algorithms and Asymptotics*. Cambridge University Press.

References IV

Hansen, L. P. and R. Jagannathan (1991). "Implications of security market data for models of dynamic economies". In: *Journal of political economy* 99.2, pp. 225–262.

Juditsky, A. and A. Nemirovski (2002). "On nonparametric tests of positivity/monotonicity/convexity". In: *The Annals of Statistics* 30.2.

Kozak (2020). "Kernel Trick for the Cross-Section". In: *SSRN Electronic Journal*.

Luzzi, E., P. G. Schneider, and R. Sen (2025). "Learning the Stochastic Discount Factor via Nonparametric Option Portfolios". In: *SSRN Electronic Journal*.

References V

Marteau-Ferey, U., F. Bach, and A. Rudi (2020). "Non-parametric models for non-negative functions". In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS '20. Vancouver, BC, Canada: Curran Associates Inc.

Muzellec, B., F. R. Bach, and A. Rudi (2021). "Learning PSD-valued functions using kernel sums-of-squares". In: *ArXiv* abs/2111.11306.

Schölkopf, B., R. Herbrich, and A. J. Smola (2001). "A Generalized Representer Theorem". In: *Computational Learning Theory*, pp. 416–426.

Seijo, E. and B. Sen (2011). "Nonparametric least squares estimation of a multivariate convex regression function". In: *The Annals of Statistics* 39.3, pp. 1633–1657.

References VI

Shapiro, A. (1985). "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints". In: *Biometrika* 72.1, pp. 133–144.

Wolak, F. A. (1987). "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model". In: *Journal of the American Statistical Association* 82.399, pp. 782–793.

Zhou, D.-X. (2008). "Derivative reproducing properties for kernel methods in learning theory". In: *Journal of Computational and Applied Mathematics* 220.1, pp. 456–463.