

Discount Curve Estimation Standalone to Transfer Learning

Sustainability in (Re-)Insurance Conference 2026
SCOR's Actuarial Awards

Nicolas Etienne Camenzind

École polytechnique fédérale de Lausanne

Supervisor: Prof. Damir Filipović

Zurich, 1st October 2026



Motivation: estimating discount curves under limited data

- Discount curves are fundamental for asset pricing, risk management, and insurance liability valuation.
- Market quotes can be noisy, sparse, and unevenly distributed across maturities.
- Long maturities are weakly identified by market data.

Central question

How can we construct robust, data-driven discount curves, especially for extrapolation?

Thesis overview

Chapter 1

Standalone

Swiss discount curve
Kernel ridge regression

[CF24] Published in EAJ¹

2024 Gauss Prize DGVFM & DAV

Chapter 2

Theory

Transfer learning
Vector-valued RKHS

[CF25] 1st revision

Management Science

Chapter 3

Empirics

Bonds and swaps
Improved extrapolation

[CF26] Submitted

EAJ

Motivated by and input to the SAV Working Group on Yield Curves

¹European Actuarial Journal

Outline

- 1 Chapter 1: Stripping the Swiss discount curve using kernel ridge regression
- 2 Chapter 2: Transfer Learning Across Fixed-Income Product Classes
- 3 Chapter 3: Transfer Learning of Discount Curves between Bonds and Swaps: An Empirical Study

Ingredients: what we observe in financial markets



Ingredients

- **Unobserved:** discount curve

$g(x)$ = value of a non-defaultable zero-coupon bond with time to maturity x

- **Observed:** M fixed income securities with

- ▶ cash flow dates $0 < x_1 < \dots < x_N$
- ▶ $M \times N$ cash flow matrix $C = [C_{ij}]$
- ▶ noisy ex-coupon prices $P = [P_1, \dots, P_M]^\top$

- **No-arbitrage pricing relation:**

$$P_i = \underbrace{C_i g(\mathbf{x})}_{\text{fundamental value}} + \underbrace{\epsilon_i}_{\text{pricing error}}$$

where $\mathbf{x} := [x_1, \dots, x_N]^\top$ and $g(\mathbf{x}) := [g(x_1), \dots, g(x_N)]^\top$

- **Pricing error** ϵ_i : due to data errors or market imperfections

Fundamental estimation problem: kernel-ridge regression (KR)

$$\min_{g \in \mathcal{G}_{\alpha, \delta}} \left\{ \underbrace{\sum_{i=1}^M \omega_i (P_i - C_i g(\mathbf{x}))^2}_{\text{pricing errors}} + \lambda \underbrace{\|g\|_{\alpha, \delta}^2}_{\text{smoothness}} \right\} \quad (1)$$

- $\mathcal{G}_{\alpha, \delta}$: **twice weakly differentiable** functions $g : [0, \infty) \rightarrow \mathbb{R}^2$
- **Smoothness parameter** $\lambda > 0$: balances pricing accuracy and smoothness
- **Duration weights** imply approximate yield-to-maturity (YTM) fitting

Representer Theorem

The KR solution $g = g_{KR}$ to fundamental problem (1) is given in closed-form by

$$g_{KR}(\cdot) = 1 + \sum_{j=1}^N k(\cdot, x_j) \beta_j, \quad \text{where} \quad \beta = C^\top (C \mathbf{K} C^\top + \Lambda)^{-1} (P - C \mathbf{1}), \quad (2)$$

where $\mathbf{K}$ is the kernel matrix induced by the kernel k . Λ the diagonal matrix consisting of λ/ω_i .

²Satisfying natural boundary conditions and hyperparameters α and δ specify the norm of $\mathcal{G}_{\alpha, \delta}$

Fundamental estimation problem: kernel-ridge regression (KR)

$$\min_{g \in \mathcal{G}_{\alpha, \delta}} \left\{ \underbrace{\sum_{i=1}^M \omega_i (P_i - C_i g(\mathbf{x}))^2}_{\text{pricing errors}} + \lambda \underbrace{\|g\|_{\alpha, \delta}^2}_{\text{smoothness}} \right\} \quad (1)$$

- $\mathcal{G}_{\alpha, \delta}$: **twice weakly differentiable** functions $g : [0, \infty) \rightarrow \mathbb{R}^2$
- **Smoothness parameter** $\lambda > 0$: balances pricing accuracy and smoothness
- **Duration weights** imply approximate yield-to-maturity (YTM) fitting

Representer Theorem

The KR solution $g = g_{KR}$ to fundamental problem (1) is given in closed-form by

$$g_{KR}(\cdot) = 1 + \sum_{j=1}^N k(\cdot, x_j) \beta_j, \quad \text{where} \quad \beta = C^\top (C K C^\top + \Lambda)^{-1} (P - C \mathbf{1}), \quad (2)$$

where K is the kernel matrix induced by the kernel k . Λ the diagonal matrix consisting of λ/ω_i .

²Satisfying natural boundary conditions and hyperparameters α and δ specify the norm of $\mathcal{G}_{\alpha, \delta}$

Fundamental estimation problem: kernel-ridge regression (KR)

$$\min_{g \in \mathcal{G}_{\alpha, \delta}} \left\{ \underbrace{\sum_{i=1}^M \omega_i (P_i - C_i g(\mathbf{x}))^2}_{\text{pricing errors}} + \lambda \underbrace{\|g\|_{\alpha, \delta}^2}_{\text{smoothness}} \right\} \quad (1)$$

- $\mathcal{G}_{\alpha, \delta}$: **twice weakly differentiable** functions $g : [0, \infty) \rightarrow \mathbb{R}^2$
- **Smoothness parameter** $\lambda > 0$: balances pricing accuracy and smoothness
- **Duration weights** imply approximate yield-to-maturity (YTM) fitting

Representer Theorem

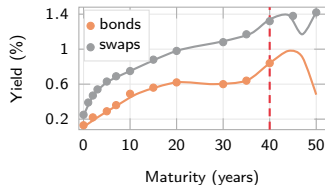
The KR solution $g = g_{KR}$ to fundamental problem (1) is given in closed-form by

$$g_{KR}(\cdot) = 1 + \sum_{j=1}^N k(\cdot, \mathbf{x}_j) \beta_j, \quad \text{where} \quad \beta = C^\top (C \mathbf{K} C^\top + \Lambda)^{-1} (P - C \mathbf{1}), \quad (2)$$

where $\mathbf{K}$ is the kernel matrix induced by the kernel k . Λ the diagonal matrix consisting of λ/ω_i .

²Satisfying natural boundary conditions and hyperparameters α and δ specify the norm of $\mathcal{G}_{\alpha, \delta}$

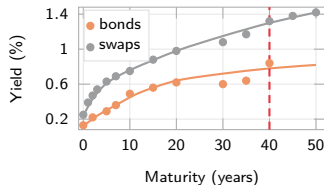
Smooth curves preferred: bias-variance trade-off



Overfit, high sensitivity

For pricing, market making

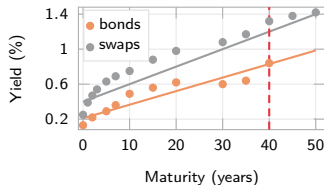
Fama–Bliss



Good fit, balanced sensitivity

For economic capital

KR method & transfer learning



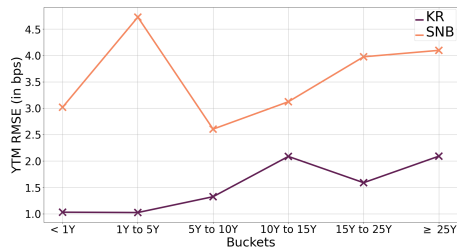
Underfit, low sensitivity

For monetary policy, macroeconomic analysis

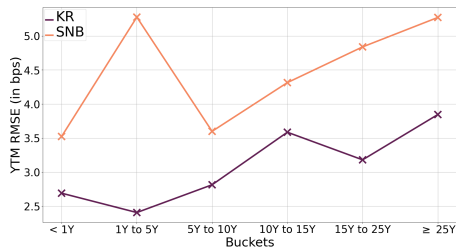
Nelson–Siegel–Svensson: SNB
Smith–Wilson: current SST

Figures are for illustration purposes only

Smooth curves preferred: bias-variance trade-off



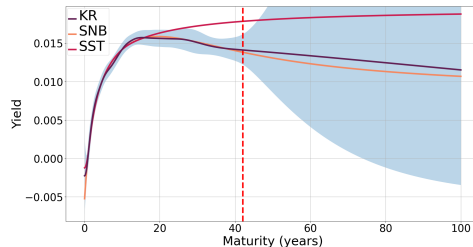
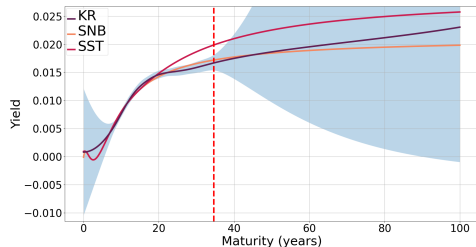
(a) in-sample



(b) out-of-sample

- Benchmark: SNB = Nelson–Siegel–Svensson (NSS) with parameter constraints
- In-sample and out-of-sample (next day) YTM RMSE with all bonds
- **KR dominates across all maturities**

Illustration: extrapolation is a choice



- Representative example days: 2014-06-16 (left) and 2022-06-15 (right)
- 3σ -confidence bands: wide for maturities with few observed or missing data

⇒ **Extrapolation is a choice and not verifiable on observed data**

Transition: extrapolation remains fundamentally difficult

- The standalone KR method performs well in regions supported by market instruments.
- However, extrapolation beyond the last available cash-flow information remains weakly identified.
- This limitation is not specific to KR: without long-dated market information, any method must impose external assumptions.

Idea

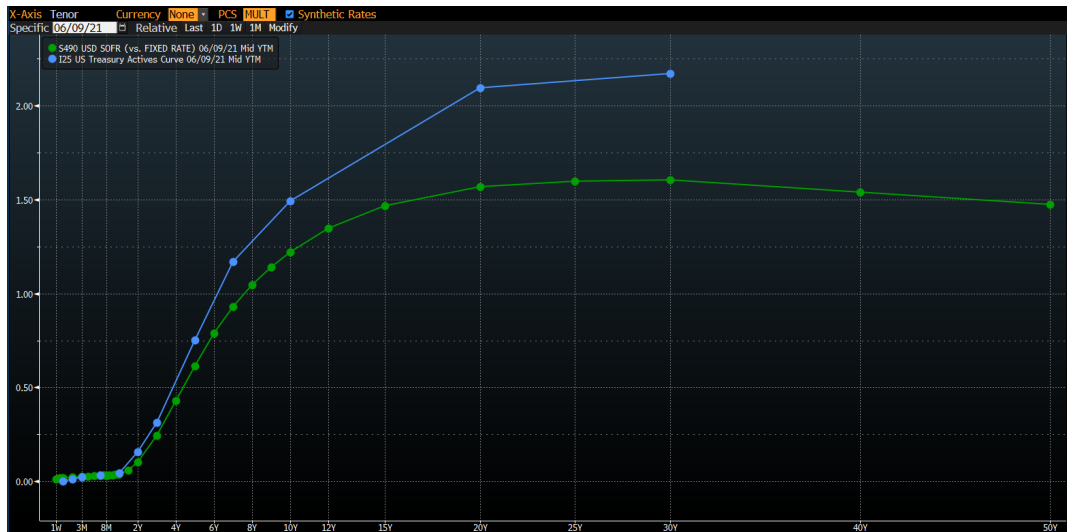
Use information from related fixed-income product classes to stabilize estimation in data-sparse regions.

Standalone KR $\Rightarrow$ Transfer learning

Outline

- 1 Chapter 1: Stripping the Swiss discount curve using kernel ridge regression
- 2 Chapter 2: Transfer Learning Across Fixed-Income Product Classes
- 3 Chapter 3: Transfer Learning of Discount Curves between Bonds and Swaps: An Empirical Study

Transfer learning: curves can be similar



Modular extension: from standalone to transfer learning

- Consider A standalone estimation problems

$$\min_{h_a \in \mathcal{H}_k} \left\{ \sum_{i=1}^{M_a} \omega_{a,i} \epsilon_{a,i}^2 + \gamma_a \|h_a\|_{\mathcal{H}_k}^2 \right\}, \quad a = 1, \dots, A,$$

with identical scalar kernel k and $\gamma_a > 0$ smoothness parameter for each product class a .

- Absent cross product class dependencies, this is equivalent to minimize the sum of the regularized objective functions,

$$\min_{h_1, \dots, h_A \in \mathcal{H}_k} \sum_{a=1}^A \left\{ \sum_{i=1}^{M_a} \omega_{a,i} \epsilon_{a,i}^2 + \gamma_a \|h_a\|_{\mathcal{H}_k}^2 \right\}.$$

- This simple reformulation turns A independent standalone estimation problems into **one optimization problem** over the product space $(\mathcal{H}_k)^A$.

Modular extension: from standalone to transfer learning

- Consider A standalone estimation problems

$$\min_{h_a \in \mathcal{H}_k} \left\{ \sum_{i=1}^{M_a} \omega_{a,i} \epsilon_{a,i}^2 + \gamma_a \|h_a\|_{\mathcal{H}_k}^2 \right\}, \quad a = 1, \dots, A,$$

with identical scalar kernel k and $\gamma_a > 0$ smoothness parameter for each product class a .

- Absent cross product class dependencies, this is equivalent to minimize the sum of the regularized objective functions,

$$\min_{h_1, \dots, h_A \in \mathcal{H}_k} \sum_{a=1}^A \left\{ \sum_{i=1}^{M_a} \omega_{a,i} \epsilon_{a,i}^2 + \gamma_a \|h_a\|_{\mathcal{H}_k}^2 \right\}.$$

- This simple reformulation turns A independent standalone estimation problems into **one optimization problem** over the product space $(\mathcal{H}_k)^A$.

Modular extension: from standalone to transfer learning

- Consider A standalone estimation problems

$$\min_{h_a \in \mathcal{H}_k} \left\{ \sum_{i=1}^{M_a} \omega_{a,i} \epsilon_{a,i}^2 + \gamma_a \|h_a\|_{\mathcal{H}_k}^2 \right\}, \quad a = 1, \dots, A,$$

with identical scalar kernel k and $\gamma_a > 0$ smoothness parameter for each product class a .

- Absent cross product class dependencies, this is equivalent to minimize the sum of the regularized objective functions,

$$\min_{h_1, \dots, h_A \in \mathcal{H}_k} \sum_{a=1}^A \left\{ \sum_{i=1}^{M_a} \omega_{a,i} \epsilon_{a,i}^2 + \gamma_a \|h_a\|_{\mathcal{H}_k}^2 \right\}.$$

- This simple reformulation turns A independent standalone estimation problems into **one optimization problem** over the product space $(\mathcal{H}_k)^A$.

Modular extension: from standalone to transfer learning

- Next step to lift this into transfer learning framework $\Rightarrow$ allows for dependencies across product classes!
- **Question:** How to achieve this? **Add spread smoothness terms of the form**

$$\sum_{a=1}^A \sum_{b>a} \Theta_{ab} \|h_a - h_b\|_{\mathcal{H}_k}^2$$

for spread parameters $\Theta_{ab} \geq 0$ for $a < b$.

- **Rational:** Arbitrage relationship across product classes enforce smooth spread curves. The more similar two product classes a and b the smoother their spread curve $h_a - h_b$, which is awarded by a larger spread parameter Θ_{ab} .

Transfer learning problem

- Jointly learn discount curves across $A \geq 2$ product classes
- Fundamental optimization problem (1) becomes

$$\min_{h_1, \dots, h_A \in \mathcal{H}_k} \sum_{a=1}^A \left\{ \sum_{i=1}^{M_a} \omega_{a,i} \underbrace{\epsilon_{a,i}^2}_{\text{pricing errors}} + \gamma_a \underbrace{\|h_a\|_{\mathcal{H}_k}^2}_{\text{smoothness}} + \sum_{b>a} \Theta_{ab} \underbrace{\|h_a - h_b\|_{\mathcal{H}_k}^2}_{\text{spread smoothness}} \right\} \quad (3)$$

- Spread parameters $\Theta_{ab} \geq 0$: tune between standalone and transfer learning
- Enables extrapolation: joint interpolation region = $\cup_{a=1}^A$ interpolation region_a

Fundamental estimation problem: vector-valued KR

Theorem 2.1.

For **regularization parameter** $\lambda > 0$, and weights $\omega_{a,i} > 0$, the unique solution of the **vector-valued kernel ridge regression** problem

$$\min_{h \in \mathcal{H}} \left\{ \underbrace{\sum_{a=1}^A \sum_{i=1}^{M_a} \omega_{a,i} \epsilon_{a,i}^2}_{\text{weighted pricing errors}} + \underbrace{\lambda \|h\|_{\mathcal{H}}^2}_{\text{regularization}} \right\}$$

is given in closed form by $\hat{h} = \sum_{j=1}^N K(\cdot, x_j) \beta_j$, where $\beta = (\beta_1, \dots, \beta_N) \in \mathbb{R}^{A \times N}$ is given by

$$\text{vec}(\beta^\top) := \mathbf{C}^\top (\mathbf{C} \mathbf{K} \mathbf{C}^\top + \mathbf{\Lambda})^{-1} (\mathbf{P} - \mathbf{C} \mathbf{1}),$$

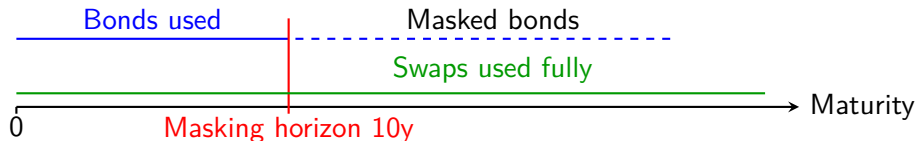
for the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\Lambda_1, \dots, \Lambda_A) \in \mathbb{R}^{M \times M}$ with $\Lambda_a = \text{diag}(\lambda/\omega_{a,1}, \dots, \lambda/\omega_{a,M_a})$, and we use the convention $\lambda/\infty = 0$. The corresponding discount curves are $\hat{g} = 1 + \hat{h}$.

Outline

- 1 Chapter 1: Stripping the Swiss discount curve using kernel ridge regression
- 2 Chapter 2: Transfer Learning Across Fixed-Income Product Classes
- 3 Chapter 3: Transfer Learning of Discount Curves between Bonds and Swaps: An Empirical Study

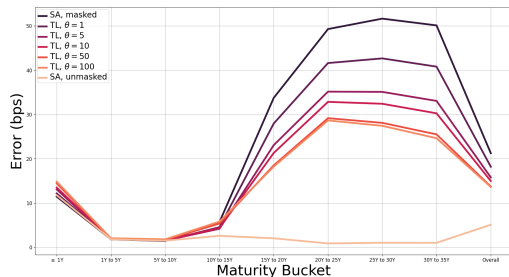
Transfer learning: masking experiment

- To assess benefits we test extrapolation ability with a masking experiment
- **Masking horizon** = maximum bond maturity used. All bonds maturing later are masked

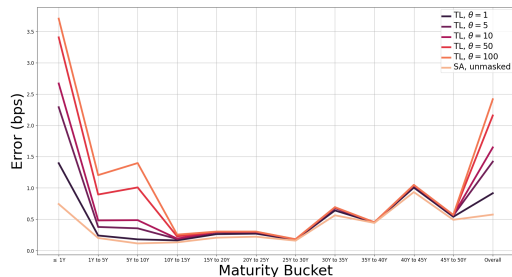


- Additional masking horizons (5 and 15 years) show similar behavior
- For various choices of θ we compare daily fitting errors with and without transfer learning
- Results show that transfer learning improves bond curve extrapolation in an economically meaningful way, with currency specific magnitudes

USD: time-averaged fitting errors by maturity bucket and overall



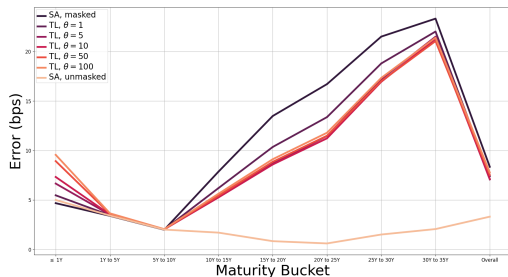
(a) US government bonds



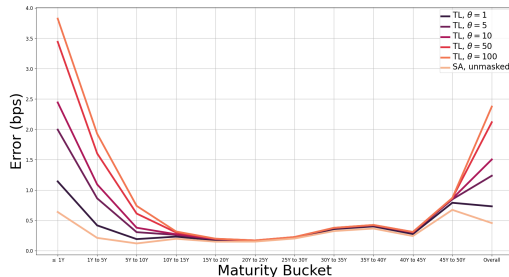
(b) SOFR swaps

- TL reduces average bond errors in masked maturity buckets
- Overall average bond error is minimized at $\theta = 100$
- TL raises average bond errors in unmasked buckets, but only moderately for all θ
- TL raises average swap errors marginally for $\theta \leq 10$, mainly in unmasked buckets

EUR: time-averaged fitting errors by maturity bucket and overall



(a) German government bonds



(b) ESTR swaps

- TL reduces average bond errors in masked maturity buckets
- Overall average bond error is minimized at $\theta = 5$
- TL raises average bond errors in unmasked buckets, but only moderately for $\theta \leq 10$
- TL raises average swap errors marginally for $\theta \leq 10$, mainly in unmasked buckets

Average errors by currency, bond and swap, and pooled

	CHF		EUR		GBP		USD		Pooled average	
	Bonds	Swaps	Bonds	Swaps	Bonds	Swaps	Bonds	Swaps	Bonds	Swaps
Unmasked										
SA	0.87	0.50	3.31	0.46	3.87	0.69	5.08	0.57	3.28	0.55
TL, $\theta = 1$	0.96	0.58	3.57	0.71	4.05	0.81	5.39	0.89	3.48	0.75
TL, $\theta = 5$	1.06	0.69	3.97	1.17	4.26	1.04	5.82	1.41	3.77	1.08
TL, $\theta = 10$	1.14	0.76	4.20	1.42	4.39	1.19	6.03	1.66	3.93	1.26
TL, $\theta = 50$	1.44	1.03	4.76	1.99	4.84	1.76	6.51	2.22	4.38	1.75
TL, $\theta = 100$	1.63	1.19	5.00	2.23	5.13	2.18	6.72	2.49	4.61	2.02
Masked										
SA	7.41	0.50	8.35	0.46	18.17	0.69	21.24	0.57	13.74	0.55
TL, $\theta = 1$	4.75	0.58	7.40	0.73	15.23	0.86	18.18	0.92	11.35	0.77
TL, $\theta = 5$	3.58	0.69	7.07	1.24	14.87	1.18	15.78	1.42	10.29	1.13
TL, $\theta = 10$	3.48	0.77	7.13	1.50	15.24	1.38	14.96	1.65	10.17	1.32
TL, $\theta = 50$	3.74	1.11	7.52	2.12	17.69	2.05	13.72	2.16	10.64	1.86
TL, $\theta = 100$	4.12	1.35	7.77	2.37	20.01	2.47	13.61	2.42	11.35	2.15

- Transfer learning significantly improves bond errors in the masked case
- Masked: $\theta = 10$ achieves the lowest pooled average error for bonds
- Unmasked: $\theta = 10$ moderately increases pooled average errors for bonds and swaps

THANK YOU

A tractable class of matrix-valued kernels: separable kernels

Definition 4.1.

A $\mathbb{R}^{A \times A}$ -valued kernel K on E is **separable** if

$$K(x, y) = B k(x, y),$$

where B is symmetric positive semi-definite and k is a scalar kernel.

Key result

$$\|h\|_{\mathcal{H}}^2 = \sum_{a=1}^A \gamma_a \|h_a\|_{\mathcal{H}_k}^2 + \sum_{a=1}^A \sum_{b>a}^A \Theta_{ab} \|h_a - h_b\|_{\mathcal{H}_k}^2$$

Is the norm of a vector-valued RKHS $\mathcal{H}$ induced by a separable kernel $K(x, y) = Bk(x, y)$ where B is parameterized by γ_a 's and Θ_{ab} 's

References I



Nicolas Camenzind and Damir Filipović.

Stripping the swiss discount curve using kernel ridge regression.
European Actuarial Journal, 14(2):371–410, June 2024.



Nicolas Camenzind and Damir Filipović.

Transfer learning across fixed-income product classes, 2025.
Available at SSRN: <https://ssrn.com/abstract=5251010>.



Nicolas Camenzind and Damir Filipović.

Transfer learning of discount curves between bonds and swaps: An empirical study, 2026.
Available at SSRN: <https://ssrn.com/abstract=6168607>.