Stripping the Discount Curve

A Robust Machine Learning Approach

Damir Filipović[1] Kay Giesecke[2] Markus Pelger[2] Ye Ye[2]

Risk Day
ETH Zurich, 17 September 2021

[1]EPFL and Swiss Finance Institute
[2]Stanford University

Preprint available online

- soon

What this paper is about

- Develop a data-driven, non-parametric discount curve estimator
- Learn discount curve from observable noisy market quotes
- Unifying, flexible machine learning framework based on a reproducing kernel Hilbert space, including, e.g., Nelson–Siegel, Svensson, Smith–Wilson curves
- Kernel ridge regression: trade off pricing error against curve regularity
- Bayesian view: Gaussian process regression
- Extensive empirical analysis of U.S. treasury data 1961 to present
- Compare with [Fama and Bliss, 1987], [Gürkaynak et al., 2007], [Svensson, 1994], [Liu and Wu, 2021] (ongoing)

Main results

- Highly tractable framework with closed-form discount curve estimator
- Hyper-parameters chosen via cross-validation (data driven)
- Estimated curves are stable over time and robust to outliers
- Inside market maturity range ("interpolation region"): curves are robust w.r.t. hyper-parameters
- Beyond market maturity range ("extrapolation region"): curves admit extrapolation, strongly dependent on hyper-parameters
- Compared to [Fama and Bliss, 1987], [Gürkaynak et al., 2007], [Svensson, 1994], KR has lowest pricing error on average, both in-sample and out-of-sample

Applications (and users)

- Study term structure of interest rates (economists)
- Predict bond returns (fixed income portfolio managers)
- Analyze monetary policy (central banks)
- Price assets, derivatives, and liabilities (actuaries and insurance regulators)

Outline

1. Discount curve by kernel ridge regression
2. A workable discount curve space
3. Comparison to Smith–Wilson method
4. Empirical study

Outline

1 Discount curve by kernel ridge regression

2 A workable discount curve space

3 Comparison to Smith–Wilson method

4 Empirical study

Ingredients

- **Unobserved ground truth**: $g(x)$ = fundamental value of a non-defaultable zero-coupon bond with time to maturity $x \in [0, \infty]$

- **Observed**: M fixed income securities with
  - cash flow dates $0 < x_1 < \cdots < x_N \leq \infty$
  - $M \times N$ cash flow matrix C
  - noisy ex-coupon prices $P = (P_1, \dots, P_M)^\top$

- No-arbitrage pricing relation

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

  where $\boldsymbol{x} = (x_1, \dots, x_N)^\top$ and $g(\boldsymbol{x}) = (g(x_1), \dots, g(x_N))^\top$

- ϵ_i: deviations from fundamental value, due to market imperfections (no deep, liquid, transparent market) and data errors

Hypothesis space

Model discount curve $g : [0, \infty] \to \mathbb{R}$ as

$$\boxed{g = p + h}$$

with

- exogenous **prior curve** $p : [0, \infty] \to \mathbb{R}$ with $p(0) = 1$, e.g., $p \equiv 1$
- **hypothesis function** h from a space $\mathcal{H}$ of functions $h : [0, \infty] \to \mathbb{R}$ with $h(0) = 0$

Good, flexible, unbiased choice of $\mathcal{H}$: reproducing kernel Hilbert space

Reproducing kernel Hilbert spaces

Definition 1.1.

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 1.2.

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 1.3 (Moore).

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

Kernel ridge regression (KR) problem

Solve

$$\min_{g=p+h,\, h\in\mathcal{H}} \left\{ \sum_{i=1}^{M} \underbrace{\omega_i (P_i - C_i g(\boldsymbol{x}))^2}_{\text{weighted pricing error}^2} + \lambda \underbrace{\|h\|_{\mathcal{H}}^2}_{\text{regularity}} \right\} \tag{1}$$

for

- exogenous weights $0 < \omega_i \le \infty$
- ω_i/λ is precision (1/variance) of price signal P_i, see Gaussian process view
- $\omega_i = \infty$ corresponds to exact pricing, $P_i = C_i g(\boldsymbol{x})$
- regularization parameter $\lambda > 0$

$\Rightarrow$ Trade-off between squared weighted pricing errors and regularity of h

Kernel representer theorem

Theorem 1.4.

KR problem (1) *has a unique solution,* $\bar{g} = p + \bar{h}$*, given by*

$$\bar{h} = \sum_{j=1}^{N} k(\cdot, x_j)\beta_j$$

where

$$\beta = C^\top (C\boldsymbol{K}C^\top + \Lambda)^{-1}(P - Cp(\boldsymbol{x})),$$

$\boldsymbol{K}_{ij} = k(x_i, x_j)$*, and* $\Lambda = \mathrm{diag}(\lambda/\omega_1, \dots, \lambda/\omega_N)$*, where we set* $\lambda/\infty = 0$*.*

Sketch of proof

Gaussian process view

$\Rightarrow$ Estimated discount curve is linear in data P

Yield-to-maturity (YTM) error

- Express price $P_i = P_i(y)$ of instrument i as function of its YTM y
- Modified duration (=Macaulay duration) $D_i = -\frac{1}{P_i}\frac{\partial P_i}{\partial y}$
- Express pricing error in terms of YTM error

$$\underbrace{P_i - C_i g(\boldsymbol{x})}_{\text{pricing error } \epsilon_i} = D_i P_i \underbrace{(y_i^g - y_i)}_{\text{YTM error}} + o(y_i - y_i^g)$$

  where y_i^g denotes the YTM of the fundamental value $C_i g(\boldsymbol{x})$

- This suggests to use weights, as in, e.g., [Gürkaynak et al., 2007],

$$\omega_i = \frac{1}{M}\frac{1}{(D_i P_i)^2}$$

  in objective function (1), so that up to first order

$$\boxed{\underbrace{\omega_i (P_i - C_i g(\boldsymbol{x}))^2}_{\text{weighted pricing error}^2} \approx \frac{1}{M}\underbrace{(y_i - y_i^g)^2}_{\text{YTM error}^2}}$$

Exogenous target yield at maturity

- We can enforce $\bar{g}(\infty) = 0$ by incrementing the number of securities

$$M \leftarrow M + 1$$

and the number of cash flow dates

$$N \leftarrow N + 1$$

and setting

$$C_M = (0, \ldots, 0, 1), \quad P_M = 0, \quad \omega_M = \infty, \quad x_N = \infty$$

- Similarly, we can enforce $\bar{g}(x_{target}) = \mathrm{e}^{-x_{target} \times y_{target}}$ for some exogenous target yield y_{target} at maturity x_{target}

Infinite-maturity yield

Outline

1 Discount curve by kernel ridge regression

2 A workable discount curve space

3 Comparison to Smith–Wilson method

4 Empirical study

A workable discount curve space

RKHS $\mathcal{H}$ consisting of twice weakly differentiable functions $h : [0, \infty] \to \mathbb{R}$ with $h(0) = h'(\infty) = 0$ and finite norm

$$\|h\|_{\mathcal{H}}^2 = \int_0^\infty (\delta h'(x)^2 + (1-\delta) h''(x)^2) w(x)\, dx$$

for

- weight function $w : [0, \infty) \to (0, \infty)$ such that $\int_0^\infty \frac{1+x^2}{w(x)} dx < \infty$
- shape parameter $\delta \in [0, 1)$

Norm $\|h\|_{\mathcal{H}}^2$ penalizes large values of

- $h'(x)^2$ to avoid oscillations, forcing h to be flat (**tension**)
- $h''(x)^2$ to avoid kinks, forcing h to be straight (**smoothness**)

$\Rightarrow$ Trade-off between tension and smoothness of h

Advantage of increasing weight function w

- Tension and smoothness penalty is maturity-dependent increasing, allowing for greater pricing flexibility at shorter maturities, while enforcing a smooth long end (as suggested by [Bliss, 1996])

Reproducing kernel: structural equation

Fix $y \geq 0$, find kernel function $\psi(\cdot) = k(\cdot, y)$:

- Kernel property

$$\langle \psi, h \rangle_{\mathcal{H}} = h(y) = \int_0^\infty 1_{[0,y]}(x) h'(x)\, dx$$

- Integration by parts, for $h \in \mathcal{H}$ with compactly supported h',

$$\begin{aligned} \langle \psi, h \rangle_{\mathcal{H}} &= \int_0^\infty (\delta \psi'(x) h'(x) + (1-\delta)\psi''(x) h''(x)) w(x)\, dx \\ &= \int_0^\infty (\delta \psi'(x) w(x) - (1-\delta)(\psi'' w)'(x)) h'(x)\, dx \end{aligned}$$

Gives structural equation for $\psi(\cdot) = k(\cdot, y)$:

$$\boxed{\delta \psi' w - (1-\delta)(\psi'' w)' = 1_{[0,y]}} \tag{2}$$

Reproducing kernel: no tension

General weight

Lemma 2.1.

Assume zero tension, $\delta = 0$, then reproducing kernel of $\mathcal{H}$ is given by

$$k(x,y) = \int_0^\infty (t \wedge x)(t \wedge y) \frac{1}{w(t)}\, dt.$$

Exponential weight $w(x) = \mathrm{e}^{\alpha x}$ for some $\alpha > 0$

Corollary 2.2.

Assume zero tension, $\delta = 0$, and exponential weight $w(x) = \mathrm{e}^{\alpha x}$, then

$$k(x,y) = -\frac{x \wedge y}{\alpha^2} \mathrm{e}^{-\alpha(x \wedge y)} + \frac{2}{\alpha^3} \left(1 - \mathrm{e}^{-\alpha(x \wedge y)}\right) - \frac{x \wedge y}{\alpha^2} \mathrm{e}^{-\alpha(x \vee y)}.$$

Reproducing kernel: tension, exponential weight

Lemma 2.3.

Assume $\delta \in (0,1)$ *and exponential weight* $w(x) = \mathrm{e}^{\alpha x}$, *then*

$$k(x,y) = -\frac{\alpha}{\delta \ell_2^2}\left(1 - \mathrm{e}^{-\ell_2 x} - \mathrm{e}^{-\ell_2 y}\right) + \frac{1}{\alpha\delta}\left(1 - \mathrm{e}^{-\alpha(x\wedge y)}\right) + \frac{1}{\delta\sqrt{D}}\left(\frac{\ell_1^2}{\ell_2^2}\mathrm{e}^{-\ell_2(x+y)} - \mathrm{e}^{-\ell_1(x\wedge y)-\ell_2(x\vee y)}\right)$$

where $\ell_1 = \frac{\alpha-\sqrt{D}}{2}$, $\ell_2 = \frac{\alpha+\sqrt{D}}{2}$, *and* $D = \alpha^2 + 4\delta/(1-\delta)$.

Sketch of proof

Infinite-maturity yield

Boundary case Smith–Wilson: tension, constant weight

Let $\alpha \to 0$ in Lemma 2.3, we obtain boundary case of our framework:

Corollary 2.4.

Assume $\delta > 0$ and constant weight $w(x) = 1$, then

$$k(x, y) = \frac{1}{\delta}(x \wedge y) - \frac{1}{\delta\rho} \mathrm{e}^{-\rho(x \vee y)} \sinh(\rho(x \wedge y))$$

where $\rho = \sqrt{\delta/(1-\delta)}$.

Functions h in $\mathcal{H}$ are only defined on $[0, \infty)$ and unbounded in general, with maximal growth rate $h(x) \sim \sqrt{x}$. E.g., $h(x) = (1+x)^{1/2-\epsilon} - 1$, $\epsilon > 0$. Hence we cannot impose constraint $g(\infty) = 0$.

Remark ([Smith and Wilson, 2001]).

The Wilson function is given by $W(x, y) = \mathrm{e}^{-y_\infty (x+y)} \delta\rho k(x, y)$, where $y_\infty = \log(1 + UFR)$, with the ultimate forward rate UFR.

Outline

1 Discount curve by kernel ridge regression

2 A workable discount curve space

3 Comparison to Smith–Wilson method

4 Empirical study

Smith–Wilson method

[Smith and Wilson, 2001] model discount curve as

$$g(x) = \mathrm{e}^{-y_\infty x}(1 + h(x))$$

with

- ultimate forward rate parameter $y_\infty > 0$
- hypothesis function h from above RKHS with constant weight $w = 1$, as in Corollary 2.4

Solve exact pricing problem with regularization

$$\begin{aligned} &\min \|h\|^2_{\mathcal{H}} \\ \text{s.t. } &P = Cg(\boldsymbol{x}), \\ &g(x) = \mathrm{e}^{-y_\infty x}(1 + h(x)), \\ &h \in \mathcal{H} \end{aligned} \tag{3}$$

Smith–Wilson estimator

Pricing equation in (3) reads

$$P = Cg(\boldsymbol{x}) = \hat{C}(1 + h(\boldsymbol{x}))$$

where we define $\hat{C} = C\,\mathrm{diag}(\mathrm{e}^{-y_\infty \boldsymbol{x}})$

⇒ (3) is special case of Theorem 1.4 with prior $p = 1$, weight $w = 1$, and **minimal regularization** $\Lambda = 0$

Corollary 3.1.

Problem (3) *has a unique solution,* $\hat{g}(x) = \mathrm{e}^{-y_\infty x}(1 + \hat{h}(x))$*, given by*

$$\hat{h}(x) = \sum_{j=1}^{N} k(x, x_j)\hat{\beta}_j = \mathrm{e}^{y_\infty x} \sum_{j=1}^{N} W(x, x_j) \underbrace{\frac{1}{\delta\rho} \mathrm{e}^{y_\infty x_j} \hat{\beta}_j}_{=\zeta_j}$$

where

$$\hat{\beta} = \hat{C}^\top (\hat{C}\boldsymbol{K}\hat{C}^\top)^{-1}(P - \hat{C}1).$$

Smith–Wilson penalty term

- KR: substitute $h(x) = g(x) - 1$ in $\mathcal{H}_\alpha$-norm with weight $w(x) = \mathrm{e}^{\alpha x}$

$$\|h\|^2_{\mathcal{H}_\alpha} = \int_0^\infty \left(\delta g'(x)^2 + (1-\delta) g''(x)^2\right) \mathrm{e}^{\alpha x}\, dx$$

directly measures tension and smoothness of g

- SW: substitute $\hat{h}(x) = \mathrm{e}^{y_\infty x} g(x) - 1$ in norm

$$\|\hat{h}\|^2_{\mathcal{H}_0} = \int_0^\infty \Big(\delta(y_\infty g(x) + g'(x))^2 + (1-\delta)(y_\infty^2 g(x) + 2y_\infty g'(x) + g''(x))^2\Big) \mathrm{e}^{2y_\infty x}\, dx$$

does **not (directly) measure** tension and smoothness of g

Consequence [Willems, 2017]

SW discount curves are less smooth than our KR curves

Related literature: [Lagerås and Lindholm, 2016], [Viehmann, 2019]

Outline

1. Discount curve by kernel ridge regression
2. A workable discount curve space
3. Comparison to Smith–Wilson method
4. Empirical study

Data

- U.S. Treasury securities from the CRSP Treasury data file
- end of month, ex-dividend bid-ask averaged mid-prices
- June 1961 to December 2017 (679 months)
- total of 4,976 issues of Treasury securities and 100,258 price quotes
- market maturity range $\leq$ 10 years

Implemented baseline specifications

- Prior curve $p = 1$, i.e., model discount curve as $g(x) = 1 + h(x)$
- Weights $\omega_i = \frac{1}{M}$ and all prices normalized to $P = 100$, i.e., we measure equally-weighted price MSE $\frac{1}{M}\|P - Cg(\boldsymbol{x})\|^2_{\mathbb{R}^M}$
- Currently: zero tension, $\delta = 0$, and exponential weight $w(x) = \mathrm{e}^{\alpha x}$, as in Corollary 2.2, i.e., we measure smoothness $\int_0^\infty g''(x)^2 \mathrm{e}^{\alpha x} dx$
- Out-of-sample pricing error: next business day

Benchmark models

- FB [Fama and Bliss, 1987]: bootstrapped piecewise constant forward rates, only available June 2016 to December 2013 ("short sample")
- NSS [Svensson, 1994]: yield curves

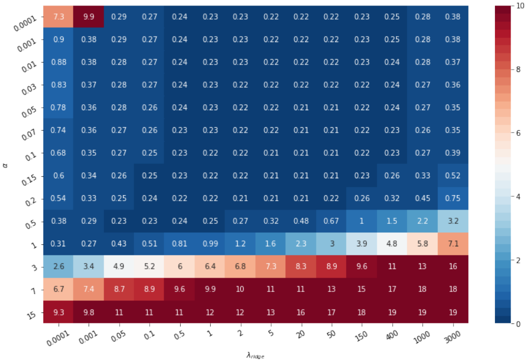
$$y_{NSS}(x) = \beta_0 + \beta_1 \frac{1 - \mathrm{e}^{-x/\tau_1}}{x/\tau_1} + \beta_2 \left(\frac{1 - \mathrm{e}^{-x/\tau_1}}{x/\tau_1} - \mathrm{e}^{-x/\tau_1} \right) + \beta_3 \left(\frac{1 - \mathrm{e}^{-x/\tau_2}}{x/\tau_2} - \mathrm{e}^{-x/\tau_2} \right)$$

- GWS [Gürkaynak et al., 2007]: NSS yield curves, more restricted data set: excluding all bonds close to maturity, and on-the-run and first off-the-run issues.

Lemma 4.1.

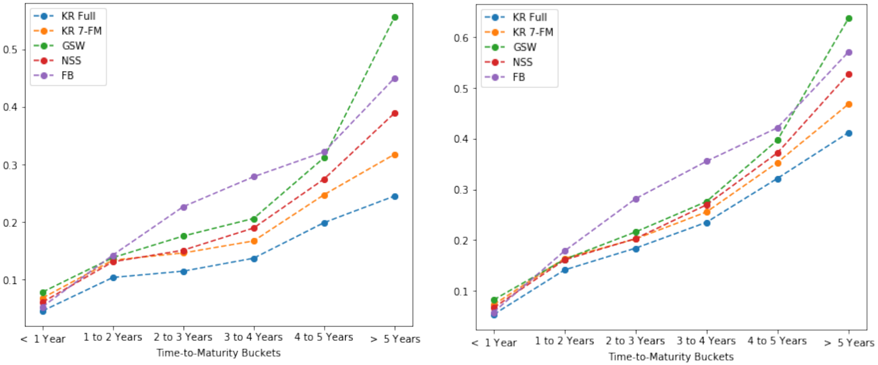
$h_{NSS}(x) = 1 - \mathrm{e}^{-x y_{NSS}(x)} \in \mathcal{H}$ *if* $\beta_0 > \alpha/2$

Average LOO cross-validation of hyperparameters



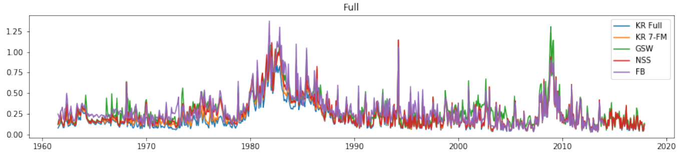
Results are robust to the choice of hyper-parameters

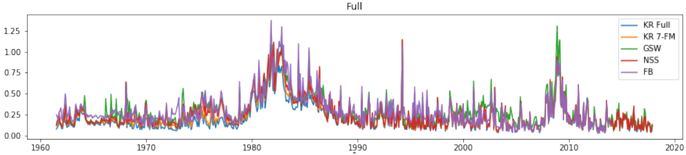
Average in-sample and out-of-sample pricing RMSE



KR model ("KR Full") outperforms

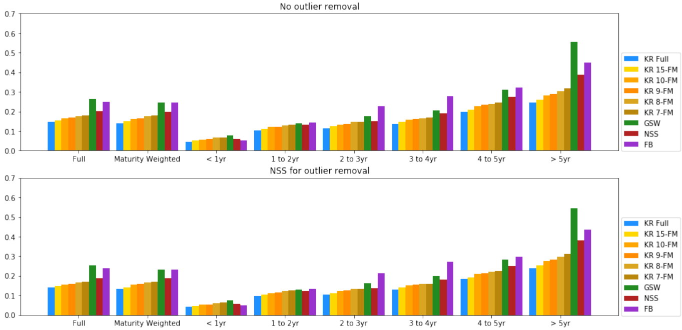
In-sample and out-of-sample pricing RMSE time series





KR model ("KR Full") outperforms

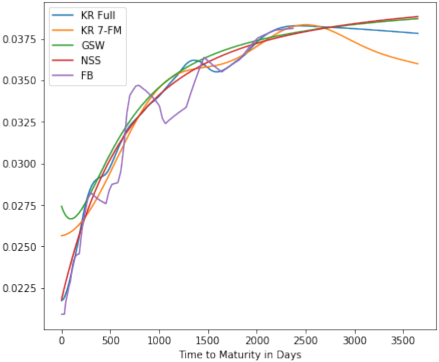
Pricing RMSE: role of outliers

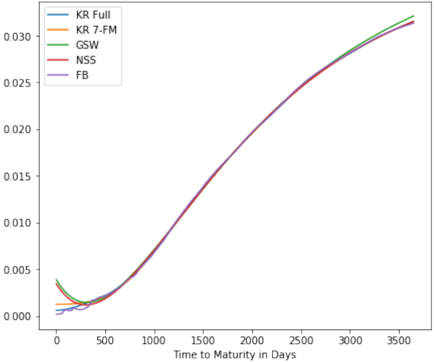


- Outlier: (YTM fitting error - CS-average YTM fitting error) $>$ 3 std.

KR Results are not driven by over-fitting to outliers

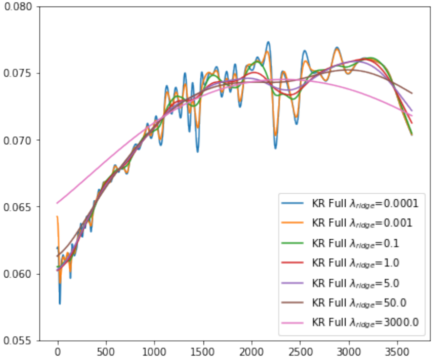
Yield curve estimates: examples 1961-06 and 2013-12





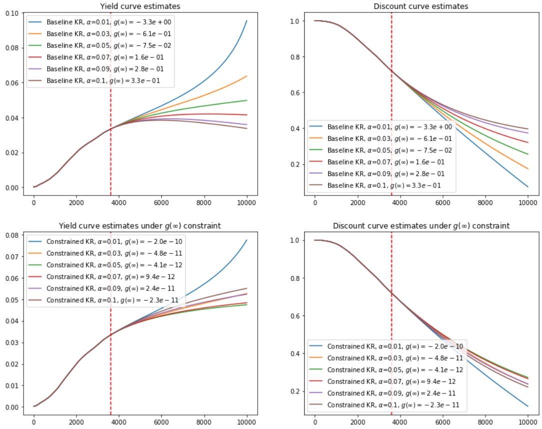
- FB curves not smooth
- GSW and NSS curves can have excessive curvature in the short end

Yield curve with varying regularization: example 1986-06



Smoother for larger λ

Extrapolation depends on α: example 2011-06



- Curve inside market maturity range is robust to choice of α
- Extrapolation of curve depends strongly on α

$\Rightarrow$ Extrapolation requires judgement call: exogenous target yield

Conclusion and outlook

- Our method is simple, fast, transparent and more robust and precise than other existing methods.

New method of choice for insurance and banking industry, regulators and central banks, to estimating the discount curve.

Extended empirical studies ongoing, including:

- U.S., EUR, and CHF data on maximum market maturity range
- Target yield $y_{target}(x_{target})$: judgement call
- Tension: more flexible curve shape and extrapolation

References I

Bliss, R. R. (1996).
Testing term structure estimation methods.
Working Paper 96-12a, Atlanta, GA.

Fama, E. F. and Bliss, R. R. (1987).
The Information in Long-Maturity Forward Rates.
The American Economic Review, 77(4):680–692.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007).
The U.S. Treasury yield curve: 1961 to the present.
Journal of Monetary Economics, 54(8):2291–2304.

Lagerås, A. and Lindholm, M. (2016).
Issues with the Smith–Wilson method.
Insurance: Mathematics and Economics, 71:93–102.

References II

Liu, Y. and Wu, J. C. (2021).
Reconstructing the yield curve.
Journal of Financial Economics.
forthcoming.

Smith, A. and Wilson, T. (2001).
Fitting yield curves with long term constraints.
Technical report, unpublished.

Svensson, L. E. (1994).
Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994.
Technical report, National Bureau of Economic Research, Cambridge, MA.

Viehmann, T. (2019).
Variants of the Smith-Wilson method with a view towards applications.

References III

Willems, S. (2017).
An alternative to the Smith-Wilson method.
Technical report, unpublished.

Appendix

Outline

5 Discount curve by kernel ridge regression (backup)

6 Gaussian process view

7 A workable discount curve space (backup)

8 KR Factor models

9 Empirical study (backup)

Sketch of proof of Theorem 1.4

- Sampling operator $S : \mathcal{H} \to \mathbb{R}^N$, $Sh = h(\boldsymbol{x})$, has adjoint

$$S^*\beta = \sum_{j=1}^N k(\cdot, x_j)\beta_j$$

and $SS^* : \mathbb{R}^N \to \mathbb{R}^N$ has matrix representation $\boldsymbol{K}$

- Rewrite KR problem (1) in operator form

$$\min_{h\in\mathcal{H}} \left\{ \sum_{i=1}^M \omega_i (P_i - C_i p(\boldsymbol{x}) - C_i Sh)^2 + \lambda \|h\|_{\mathcal{H}}^2 \right\} \tag{4}$$

- Solution h of (4) must be orthogonal to null space of CS:

$$h = S^* C^\top q, \quad \text{some } q \in \mathbb{R}^M$$

- Problem (4) becomes quadratic in $q \in \mathbb{R}^M$, solve by FOC

Back to main

Infinite-maturity yield

Zero-coupon yield $\bar{y}(x) = -\frac{\log \bar{g}(x)}{x}$, define $\bar{y}(\infty) = \lim_{x\to\infty} \bar{y}(x)$

Lemma 7.1.

Assume that there exists a function $q > 0$ *and parameter* $\alpha > 0$ *such that*

$$\lim_{x\to\infty} \frac{1}{x} \log q(x) = 0,$$
$$\lim_{x\to\infty} (p(x) - p(\infty))q(x)\mathrm{e}^{\alpha x} = \gamma_0,$$
$$\lim_{x\to\infty} (k(x,x_j) - k(\infty,x_j))q(x)\mathrm{e}^{\alpha x} = \gamma_j, \quad j = 1,\dots,N,$$

for some real $\gamma_0,\dots,\gamma_N$ *such that* $\gamma_0 + \sum_{j=1}^N \beta_j\gamma_j > 0$. *Then*

$$\boxed{\bar{g}(\infty) = 0 \iff \bar{y}(\infty) = \alpha.}$$

Sketch of proof of Lemma 7.1

- Decompose

$$\begin{aligned}\bar{g}(x) &= \bar{g}(\infty) + \bar{g}(x) - \bar{g}(\infty)\\ &= \bar{g}(\infty) + \underbrace{p(x) - p(\infty)}_{\to 0} + \sum_{j=1}^{N} \underbrace{(k(x,x_j) - k(\infty,x_j))}_{\to 0}\beta_j\end{aligned}$$

- Hence: $\bar{g}(\infty) = \lim_x \bar{g}(x) = 0$ if $\bar{y}(\infty) = \alpha > 0$
- Now assume $\bar{g}(\infty) = 0$ and decompose

$$\bar{y}(x) = \underbrace{\frac{-\log(\bar{g}(x)q(x)\mathrm{e}^{\alpha x})}{x}}_{\to 0} + \underbrace{\frac{\log q(x)}{x}}_{\to 0} + \alpha$$

Back to main

Outline

5 Discount curve by kernel ridge regression (backup)

6 Gaussian process view

7 A workable discount curve space (backup)

8 KR Factor models

9 Empirical study (backup)

Gaussian processes

- Assume $g : [0, \infty] \to \mathbb{R}$ is a **Gaussian process** with **mean** function m and **covariance** kernel k. That is, for any choice $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$,

$$g(\boldsymbol{x}) \sim \mathcal{N}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}^\top))$$

- Assume $m(0) = 1$ and $k(0, 0) = 0$, so that $g(0) = 1$

GP pricing model

Model pricing relation as

$$P_i = C_i g(\boldsymbol{x}) + \epsilon_i$$

with

- pricing errors $\epsilon = (\epsilon_1, \dots, \epsilon_M)^\top \sim \mathcal{N}(0, \Sigma)$ independent of g
- variance parameters $\Sigma = \mathrm{diag}(\sigma_1^2, \dots, \sigma_M^2)$, for $\sigma_i \geq 0$
- $\sigma_i = 0$ corresponds to exact pricing, $P_i = C_i g(\boldsymbol{x})$

GP posterior distribution

- $\bar{N}$ arbitrary dates $\bar{\boldsymbol{x}} = (\bar{x}_1, \dots, \bar{x}_{\bar{N}})^\top$
- **Fact**: Conditional distribution of $g(\bar{\boldsymbol{x}})$ given $(P, \boldsymbol{x})$ is Gaussian, $\mathcal{N}(\bar{m}, \bar{\Sigma})$, with mean vector

$$\bar{m} = m(\bar{\boldsymbol{x}}) + k(\bar{\boldsymbol{x}}, \boldsymbol{x}^\top) C^\top (C\boldsymbol{K}C^\top + \Sigma)^{-1}(P - Cm(\boldsymbol{x}))$$

and covariance matrix

$$\bar{\Sigma} = k(\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}^\top) - k(\bar{\boldsymbol{x}}, \boldsymbol{x}^\top) C^\top (C\boldsymbol{K}C^\top + \Sigma)^{-1} C k(\boldsymbol{x}, \bar{\boldsymbol{x}}^\top)$$

$\Rightarrow$ obtain confidence ranges for prices $\gamma^\top g(\bar{\boldsymbol{x}}) \sim \mathcal{N}(\gamma^\top \bar{m}, \gamma^\top \bar{\Sigma} \gamma)$

GP posterior prediction

- **Consequence**: Obtain the posterior **predicted discount curve** $\bar{g}$, given the observed prices P, for a generic cash flow date $\bar{\boldsymbol{x}} = x$,

$$\bar{g}(x) = \mathbb{E}[g(x) \mid P, \boldsymbol{x}] = m(x) + k(x, \boldsymbol{x}^\top)\beta$$

for coefficients

$$\beta = C^\top (C\boldsymbol{K}C^\top + \Sigma)^{-1}(P - Cm(\boldsymbol{x})).$$

$\Rightarrow$ Equivalent to KR with prior $m = p$ and variance weights $\omega_i = \lambda/\sigma_i^2$.

Back to main

Outline

5 Discount curve by kernel ridge regression (backup)

6 Gaussian process view

7 A workable discount curve space (backup)

8 KR Factor models

9 Empirical study (backup)

Sketch of proof of Lemma 2.3

Structural equation (2),

$$\delta \psi' w - (1-\delta)(\psi'' w)' = 1_{[0,y]},$$

becomes non-homogeneous linear differential equation with constant coefficients for $f(\cdot) = \psi'(\cdot)$:

$$\delta f(t) - (1-\delta)\alpha f'(t) - (1-\delta) f''(t) = 1_{[0,y]}(t)\mathrm{e}^{-\alpha t}.$$

Solve by the variation of constants method:

- characteristic equation, $\delta/(1-\delta) - \alpha t - t^2 = 0$, has roots $t = \ell_1, \ell_2$, so that
$$f(t) = c_1(t)\mathrm{e}^{\ell_1 t} + c_2(t)\mathrm{e}^{\ell_2 t}$$
- boundary conditions and cumbersome integration (by parts) gives the result

Back to main

Infinite-maturity yield: exponential weight

Theorem 9.1.

Assume $w(x) = \mathrm{e}^{\alpha x}$*, constraint* $\bar{g}(\infty) = 0$*, and* $\tilde{g}(\infty) > 0$ *where* $\tilde{g}$ *is the estimated curve without constraint on* $\tilde{g}(\infty)$*. Then*

$$\boxed{\bar{y}(\infty) = \alpha.}$$

Back to main

Outline

5 Discount curve by kernel ridge regression (backup)

6 Gaussian process view

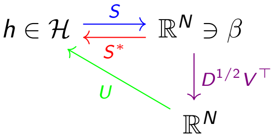
7 A workable discount curve space (backup)

8 KR Factor models

9 Empirical study (backup)

Discount curve compression in $\mathcal{H}$

- Sampling operator $Sh = (h(x_1), \dots, h(x_N))^\top$
- Adjoint given by $S^*\beta = \sum_{j=1}^N k(\cdot, x_j)\beta_j$
- Hence we can write $g = p + S^*\beta$
- $SS^* : \mathbb{R}^N \to \mathbb{R}^N$ has matrix representation $\boldsymbol{K}$



Spectral decomposition $\boldsymbol{K} = VDV^\top$ for

- $V = (v_1, \dots, v_N)$ orthonormal eigenvectors v_i of $\boldsymbol{K}$
- $D = \mathrm{diag}(\mu_1, \dots, \mu_N)$ with eigenvalues $\mu_1 \geq \dots \geq \mu_N \geq 0$

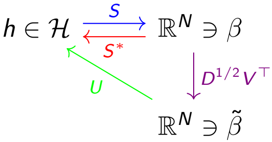
Singular value decomposition $S^* = UD^{1/2}V^\top$ for

- $U = (u_1, \dots, u_N) : \mathbb{R}^N \to \mathcal{H}$ orthonormal eigenfunctions $u_i = \frac{1}{\sqrt{\mu_i}} S^* v_i$ of $S^*S : \mathcal{H} \to \mathcal{H}$ with eigenvalues μ_i, i.e., $S^*Su_i = \mu_i u_i$.

Discount curve compression in $\mathcal{H}$

- Write

$$h = S^* \beta = U \underbrace{D^{1/2} V^\top \beta}_{=\tilde\beta} = U \tilde\beta$$



with **principal components** $\tilde\beta$

- Note that $\|h\|_{\mathcal{H}} = \|\tilde\beta\|_{\mathbb{R}^N}$

→ Obtain the low rank approximation (compression)

$$h \approx \sum_{j=1}^{d} \tilde\beta_j u_j$$

for the first d PCs $\tilde\beta_i$ corresonding to largest singular values $\sqrt{\mu_i}$

KR d-factor models

- By Theorem 1.4, the KR problem (1) is equivalent to

$$\min_{g=p+U\tilde{\beta},\,\tilde{\beta}\in\mathbb{R}^N} \left\{ \sum_{i=1}^{M} \omega_i (P_i - C_i g(\boldsymbol{x}))^2 + \lambda \|\tilde{\beta}\|^2_{\mathbb{R}^N} \right\}$$

- Obtain d-factor model by solving

$$\min_{g=p+\sum_{j=1}^{d}\tilde{\beta}_j u_j,\,\tilde{\beta}\in\mathbb{R}^d} \left\{ \sum_{i=1}^{M} \omega_i (P_i - C_i g(\boldsymbol{x}))^2 + \lambda \|\tilde{\beta}\|^2_{\mathbb{R}^d} \right\}$$

- Sparsity check: run LASSO, selecting d factors,

$$\min_{g=p+U\tilde{\beta},\,\tilde{\beta}\in\mathbb{R}^N} \left\{ \sum_{i=1}^{M} \omega_i (P_i - C_i g(\boldsymbol{x}))^2 + \lambda \|\tilde{\beta}\|^2_{\mathbb{R}^N} + \lambda_{LASSO} \|\tilde{\beta}\|_{\mathbb{R}^N} \right\}$$

- Question: does LASSO select first principal components $\tilde{\beta}_1, \ldots, \tilde{\beta}_d$?

Outline

5 Discount curve by kernel ridge regression (backup)

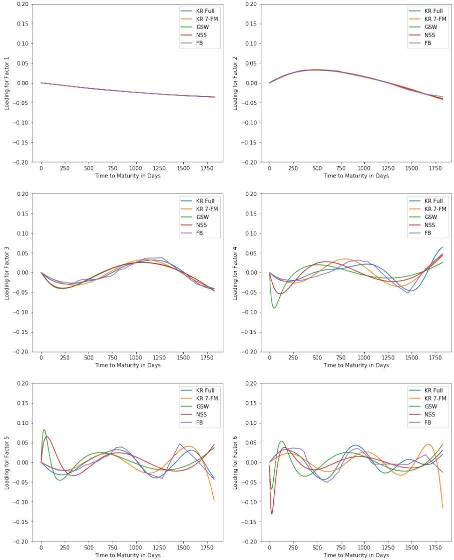
6 Gaussian process view

7 A workable discount curve space (backup)

8 KR Factor models

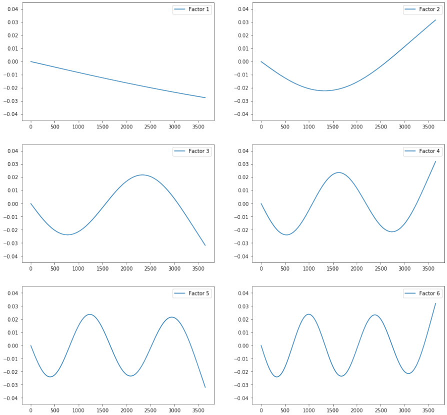
9 Empirical study (backup)

First 6 PCA loadings estimated on panels



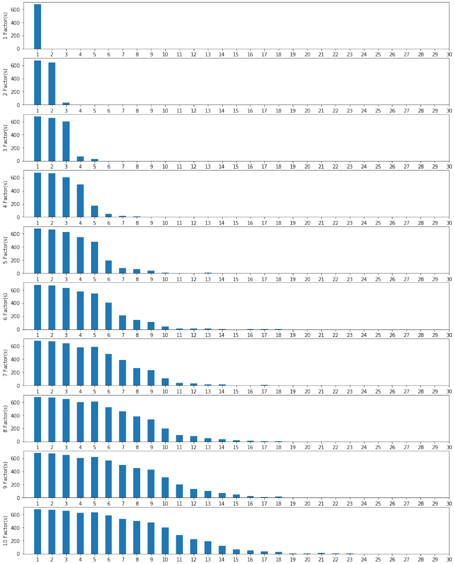
- GSW and NSS are unstable in the short end

Eigenvectors of kernel matrix (=loadings of factor model)

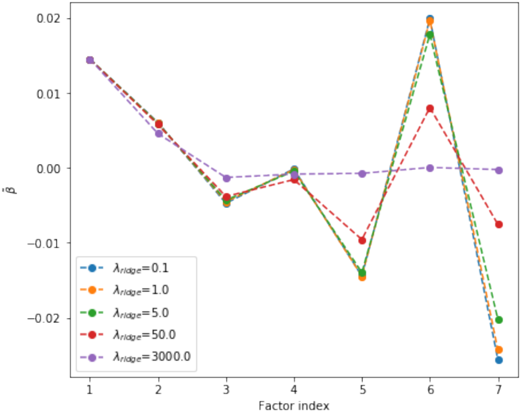


- Striking similarity to panel PCA

Sparsity check: LASSO selects first principal components



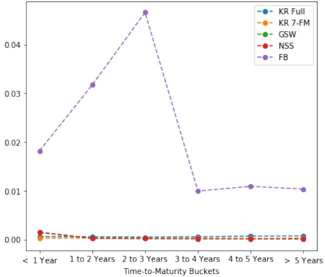
Regularization shrinks magnitude of principal components

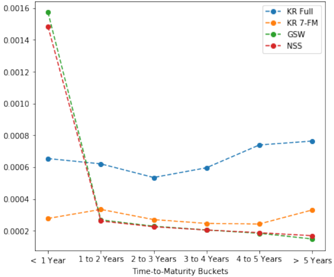


- Increasing regularization λ shrinks principal components $\tilde{\beta}$

$\Rightarrow$ Ridge has similar effect as LASSO

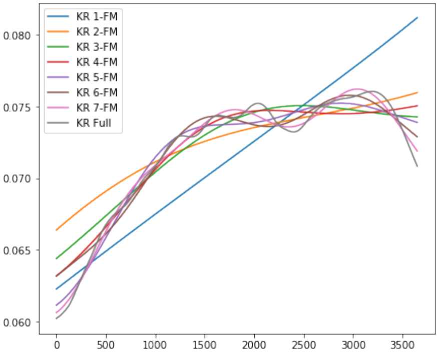
Smoothness measure $\int g''(x)^2 dx$ comparison

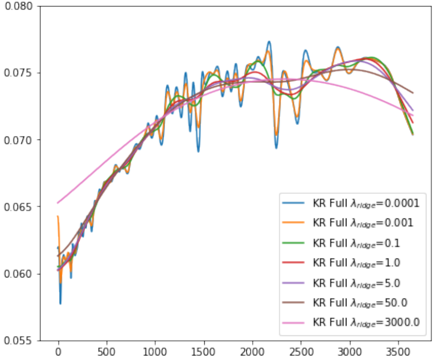




- FB least smooth
- Short ends of GSW and NSS are not stable (excessive curvature)

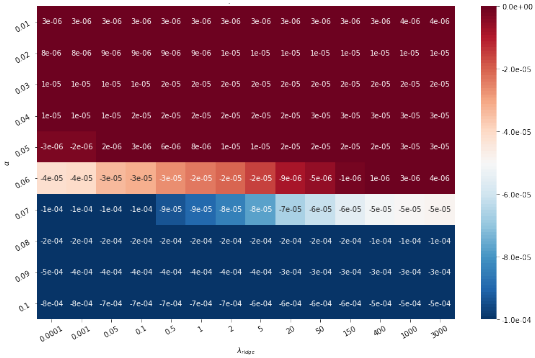
Yield curve for KR FMs and full: example 1986-06





- Smoother for less factors and larger λ

Extrapolation to infinite-maturity yield α



- Average β_N for constrained KR model $\bar{g}(\infty) = 0$
- Extrapolation to infinite-maturity yield $\boxed{\alpha = \bar{y}(\infty) \text{ only if } \beta_N < 0}$
- Including tension, $\delta > 0$, should improve the results (ongoing)