Stripping the Discount Curve
A Robust Machine Learning Approach

Damir Filipović\textsuperscript{1} Markus Pelger\textsuperscript{2} Ye Ye\textsuperscript{2}

Conference in memory of Tomas Björk
Stockholm, 10 October 2022

\textsuperscript{1}EPFL and Swiss Finance Institute
\textsuperscript{2}Stanford University
This talk is based on

- Work of Tomas and many other researchers who have been inspired by him
Outline

1 Past to Present: 1990s – today

2 Kernel Ridge (KR) Method

3 Empirical results
Outline

1. Past to Present: 1990s – today
3. Empirical results
Past: the roaring 1990s

- 1990s: from PDEs to Stochastic PDEs: HJM–Musiela equation

\[
df_t(x) = \left( \frac{\partial}{\partial x} f_t(x) + \sigma(f_t, x, \omega) \int_0^x \sigma(f_t, \xi, \omega) \, d\xi \right) \, dt + \sigma(f_t, x, \omega) \, dW_t
\]

- Heath, Jarrow, and Morton (HJM): every arbitrage-free interest rate model is of this form
- Da Prato and Zabczyk et al.: analyze (1) as stochastic equation in Hilbert space \( \mathcal{H} \)
- Parametric families \( \mathcal{M} \) of forward curves: e.g., Nelson–Siegel and Svensson (NSS)

\[
f_{NSS}(x) = \gamma_0 + \gamma_1 e^{-\frac{x}{\tau_1}} + \gamma_2 \frac{x}{\tau_1} e^{-\frac{x}{\tau_1}} + \gamma_3 \frac{x}{\tau_2} e^{-\frac{x}{\tau_2}}, \quad \gamma_0, \ldots, \tau_2 \in \mathbb{R}
\]

to represent and estimate forward curves from data
Consistency problems

1999: Tomas and B. Christensen formulate consistency problems:

- Problem I. Given an interest rate model (1) and a family of forward curves \( M \), what are necessary and sufficient conditions for invariance (i.e. consistency)?
- Problem II. Take as given a specific family \( M \) of forward curves. Does there exist any interest rate model (1) that is consistent with \( M \)?
- Problem III. Take as given a specific interest rate model (1). Does there exist a finitely parametrized family of forward curves \( M \) that is consistent with (1); in other words, does there exist a finite dimensional factor realization?

Set the ground for many (and my own PhD) research projects
State space $\mathcal{H}$ for HJMM equation (1)

- Björk and Christensen (1999): assume a strong solution of (1) in $\mathcal{H}$ with norm
  \[ \| f \|^2 = \int_0^\infty |f(x)|^2 e^{-\gamma x} \, dx + \int_0^\infty |f'(x)|^2 e^{-\gamma x} \, dx, \quad \gamma > 0 \]

- Björk and Svensson (2001): define Hilbert space $\mathcal{H}$ with norm
  \[ \| f \|^2 = \sum_{n=0}^\infty \beta^{-n} \int_0^\infty |f^{(n)}(x)|^2 e^{-\gamma x} \, dx, \quad \gamma > 0, \beta > 1 \]
  $\frac{\partial}{\partial x}$ is bounded on $\mathcal{H}$. Space is very “small”: contains only real analytic functions

- DF (2000) analyzed HJM equation (1) on weighted Sobolev space $\mathcal{H} = H_{\alpha,1}$
  \[ \| f \|^2 = |f(0)|^2 + \int_0^\infty |f'(x)|^2 e^{\alpha x} \, dx, \quad \alpha > 0 \]
  $\frac{\partial}{\partial x}$ generates strongly continuous shift semigroup on $\mathcal{H}$.

- DF and Teichmann (2003): work on larger convenient spaces
Some main findings

- Björk and Christensen (1999): Nelson–Siegel is not consistent with (deterministic) HJM
- DF (1999): Nelson–Siegel(–Svensson) is not consistent with any (nontrivial) HJM model
- Björk and Svensson (2001): finite-dimensional invariant submanifolds are affine
- DF and Teichmann (2003) generalized this to convenient spaces and stochastic volatility
Fact check: estimation methods used by several central banks


<table>
<thead>
<tr>
<th>Central bank</th>
<th>Method</th>
<th>Minimized error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>S or NS</td>
<td>wp</td>
</tr>
<tr>
<td>Canada</td>
<td>Exponential spline</td>
<td>wp</td>
</tr>
<tr>
<td>Finland</td>
<td>NS</td>
<td>wp</td>
</tr>
<tr>
<td>France</td>
<td>S or NS</td>
<td>wp</td>
</tr>
<tr>
<td>Germany</td>
<td>S</td>
<td>yields</td>
</tr>
<tr>
<td>Italy</td>
<td>NS</td>
<td>wp</td>
</tr>
<tr>
<td>Japan</td>
<td>Smoothing spline</td>
<td>prices</td>
</tr>
<tr>
<td>Norway</td>
<td>S</td>
<td>yields</td>
</tr>
<tr>
<td>Spain</td>
<td>S</td>
<td>wp</td>
</tr>
<tr>
<td>Sweden</td>
<td>Smoothing spline or S</td>
<td>yields</td>
</tr>
<tr>
<td>Switzerland</td>
<td>S</td>
<td>yields</td>
</tr>
<tr>
<td>UK</td>
<td>Smoothing spline</td>
<td>yields</td>
</tr>
<tr>
<td>USA</td>
<td>Smoothing spline</td>
<td>bills: wp, bonds: prices</td>
</tr>
</tbody>
</table>

**Table:** Nelson–Siegel (NS), Svensson (S), weighted prices (wp)
Present: current research

- Discount or yield curve critical economic quantity
- Discount curve has to be estimated from sparse and noisy Treasury prices
- Statistical estimation of the discount curve was considered in fundamental work by [Fama and Bliss, 1987], follow-up by Nelson and Siegel, [Svensson, 1994], [Gürkaynak et al., 2007], and [Liu and Wu, 2021], a.o.

Problems:
- Parametric models lack flexibility and are misspecified (consistency)
- Nelson–Siegel–Svensson fitting leads to non-convex optimization problem
- Non-parametric models (e.g., FB) prone to overfitting and irregular curve shapes
Outline

1. Past to Present: 1990s – today
3. Empirical results
Ingredients

- **Unobserved ground truth: discount curve** $g(x) = \text{fundamental value of a non-defaultable zero-coupon bond with time to maturity } x$

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

- No-arbitrage pricing relation

\[
    P_i = C_i g(x) + \epsilon_i
\]

where $x = (x_1, \ldots, x_N)^\top$ and $g(x) = (g(x_1), \ldots, g(x_N))^\top$

- $\epsilon_i = \text{deviations of } P_i \text{ from fundamental value, due to market microstructure effects, convenience yields, or data errors}$
Estimation problem

Minimize weighted pricing errors for some exogenous weights $\omega_i$:

$$\min_{g \in G} \left\{ \sum_{i=1}^{M} \omega_i (P_i - C_i g(x))^2 \right\}$$

Problems:

- Underdetermined problem: we observe only $M \approx 300$ U.S. treasuries, need to estimate $N \approx 10,000$ (30 years x 365 days) discount bond prices
- Every estimation approach needs to reduce degrees of freedom
- Existing approaches impose ad-hoc assumptions $\Rightarrow$ misspecified form
- How to choose hypothesis set $G$?
- How to choose the weights $\omega_i$?
Arbitrage-free discount curves are generically twice differentiable

- Any arbitrage-free discount curve is the form $g(x) = \mathbb{E}_Q\left[e^{-\int_0^x r_u \, du}\right]$.
- For some short rate $Q$-dynamics $dr_t = \mu_t \, dt + dM_t$.
- Stochastic Taylor expansion

$$e^{-\int_0^x r_u \, du} = 1 - \int_0^x e^{-\int_0^r r_u \, du} \, r_t \, dt$$

$$e^{-\int_0^t r_u \, du} \, r_t = r_0 - \int_0^t e^{-\int_0^s r_u \, du} \left(r_s^2 - \mu_s\right) \, ds + \int_0^t e^{-\int_0^s r_u \, du} \, dM_s$$

- Under mild moment assumptions on $r_t$, $\mu_t$, $M_t$, we obtain that

$$g(x) = 1 - r_0 x + \int_0^x \int_0^t \mathbb{E}_Q\left[e^{-\int_0^s r_u \, du} \left(r_s^2 - \mu_s\right)\right] \, ds \, dt$$

is twice weakly differentiable.
Measure of smoothness

- We obtain natural space of arbitrage-free discount curves

\[ G = \text{space of twice weakly differentiable functions } g : [0, \infty) \to \mathbb{R} \text{ with } g(0) = 1 \]

- Define measure smoothness of \( g \in G \) by

\[
\|g\|_{2,\alpha,\delta}^2 = \int_0^\infty \left( \delta g'(x)^2 + (1 - \delta) g''(x)^2 \right) e^{\alpha x} \, dx
\]

- **Curvature** \( g''(x)^2 \): penalizing avoids kinks
- **Tension** \( g'(x)^2 \): penalizing avoids redundant oscillations
- **Maturity weight** \( \alpha \geq 0 \): distributes smoothness measure across maturities
- **Tension parameter** \( \delta \in [0, 1] \) balances tension and curvature
Appendix: two lemmas

**Lemma 2.1 (Equivalence of norms).**

For fixed $\alpha > 0$, for all $0 \leq \delta_1, \delta_2 < 1$ there exists $C_{12}$ such that $\|g\|_{\alpha, \delta_1} \leq C_{12} \|g\|_{\alpha, \delta_2}$.

**Lemma 2.2 (Relation to forward curves).**

Assume the forward curve $f$ lies in the forward curve space $H_{\alpha, 1}$, that is,

$$\int_0^\infty f'(x)^2 e^{\alpha x} \, dx < \infty.$$

Then the discount curve $g(x) = e^{-\int_0^x f(t) \, dt}$ satisfies $\|g\|_{\alpha, \delta} < \infty$ for any $\alpha < 2f(\infty)$.

**Corollary 2.3.**

The discount curve space $G_{\alpha, \delta} \subset \mathcal{G}$ is consistent with essentially all HJM models.
Regularized estimation problem

- Add smoothness measure as regularization term to estimation problem:

\[
\min_{g \in \mathcal{G}} \left\{ \sum_{i=1}^{M} \omega_i(P_i - C_i g(x))^2 + \lambda \int_{0}^{\infty} \left( \delta g'(x)^2 + (1 - \delta) g''(x)^2 \right) e^{\alpha x} \, dx \right\} \tag{2}
\]

- **Smoothness parameter** \( \lambda \geq 0 \): trade-off between pricing errors and smoothness
- **Weights** \( 0 < \omega_i \leq \infty \) (\( \omega_i = \infty \) is exact pricing): we set \( \sqrt{\omega_i} \propto \frac{1}{P_i \text{duration}_i} \) \( \Rightarrow \) yield fitting
- **Technical remark**: \( \lambda = 0 \) corresponds to \( \omega_i = \infty \) for all \( i \)
- **Select** \( \alpha, \delta, \lambda \) empirically via cross-validation to minimize pricing errors out-of-sample
Theorem 2.4 (Kernel-Ridge (KR) Solution).

The regularized problem (2) has a unique solution \( g = \hat{g} \) given in closed form

\[
\hat{g}(x) = 1 + \sum_{j=1}^{N} k(x, x_j) \beta_j,
\]

where \( \beta = (\beta_1, \ldots, \beta_N)^\top \) is given by

\[
\beta = C^\top (C K C^\top + \Lambda)^{-1} (P - C1), \quad \Lambda = \text{diag} \left( \frac{\lambda}{\omega_1}, \ldots, \frac{\lambda}{\omega_M} \right)
\]

for the \( N \times N \)-kernel matrix \( K_{ij} = k(x_i, x_j) \), where kernel function \( k : [0, \infty) \times [0, \infty) \rightarrow \mathbb{R} \) is given in closed form.
KR solution

\[ \hat{g}(x) = 1 + \sum_{j=1}^{N} k(x, x_j) \beta_j, \quad \text{where} \quad \beta = C^T (CKC^T + \Lambda)^{-1} (P - C1) \]

- Simple closed-form solution, easy to implement
- Kernel ridge regression with smoothness as ridge penalty
- Discount bonds are portfolios of coupon bonds \( \Rightarrow \) Immunization
- Basis functions \( k(\cdot, x_j) \) are determined by smoothness measure \( (\alpha, \delta) \)
Outline

1 Past to Present: 1990s – today
2 Kernel Ridge (KR) Method
3 Empirical results
Data

Out-of-sample analysis on U.S. Treasury securities:
- U.S. Treasury securities from the CRSP Treasury data file
- End of month, ex-dividend bid-ask averaged mid-price
- Sampling period: June 1961 to December 2020 (715 months)
- Total of 5,422 issues of Treasury securities and 121,088 price quotes

Estimation and evaluation:
- Out-of-sample evaluation on next business day $t + 1$ for model estimated on day $t$
- Cross-sectional out-of-sample with stratified sampling to keep maturity distribution
- Root-mean-squared errors (RMSE) for yields and percentage price errors
Cross-validation for hyper-parameters $\alpha$, $\delta$, $\lambda$

- Figures show average cross-validation YTM fitting error (in bps)
- Optimal $\lambda \approx 1$: ground truth discount curve exhibits minimal curvature
- Optimal $\delta \approx 0$: ground truth discount curve exhibits systematic slope pattern

**Figure**: Left panel: fix $\alpha = 0.05$, vary $\delta$, $\lambda$. Right panel: fix $\delta = 0$, vary $\alpha$, $\lambda$. 
Conclusion: KR method satisfies all principles of yield curve estimation

- Simple and fast to implement
- Transparent and reproducible
- Data-driven
- Precise representation of the term structure, taking into account all market signals
- Robust to outliers and data selection choices (shown in paper)
- Flexible for integration of external views: exogenous points, choice of weights
- Consistent with absence of arbitrage
Appendix: Bayesian perspective

Assume Gaussian prior distribution

\[ g(x) \sim \mathcal{N}(m(x), k(x, x^\top)) \]

with pricing errors \( \epsilon \sim \mathcal{N}(0, \Sigma^\epsilon) \) for \( \Sigma^\epsilon = \text{diag}(\sigma_1^2, \ldots, \sigma_M^2) \).

**Theorem 5.1 (Bayesian perspective).**

If prior mean function \( m(x) = 1 \) and pricing error variance \( \sigma_i^2 = \frac{\lambda}{\omega_i} \), then

1. the posterior mean function of the Gaussian Process equals the KR estimated discount curve,
2. the posterior distribution is a Gaussian process with known posterior variance.

\( \Rightarrow \) We obtain a confidence range for the discount curve and securities
References


