Stripping the Discount Curve - a Robust Machine Learning Approach

Damir Filipović, Markus Pelger, Ye Ye

Stanford University
Fundamental Problem

Research problem:
- Discount or yield curve critical economic quantity
- Discount curve has to be estimated from sparse and noisy Treasury prices

This paper: Methodology
- Develops a data-driven, non-parametric discount curve estimator
- Combines financial theory with modern machine-learning
- Closed-form solution as simple ridge regression
- Gaussian process view gives confidence intervals
⇒ Simple, fast, transparent, robust, and precise estimator for yields

This paper: Empirics
- Extensive out-of-sample study on U.S. Treasury securities
- Our method dominates all parametric and non-parametric benchmarks
- Substantially smaller out-of-sample yield and pricing errors
- Robust to outliers and data selection choices
- Document systematic biases and instabilities of popular methods
⇒ Our method sets the new standard for yield curve estimation
⇒ Publicly available data sets
Model
Fundamental Problem

**Unobserved discount curve** $g(x)$

- fundamental value of a non-defaultable zero-coupon bond with time to maturity $x$

**Observed:** $M$ coupon bonds securities with

- cash flow dates $0 < x_1 < \cdots < x_N$
- $M \times N$ cash flow matrix $C$
- Observed coupon bond prices $P = (P_1, \ldots, P_M)^\top$

**No-arbitrage pricing relation:**

$$P_i = C_i g(x) + \epsilon_i$$

- Evaluated at cash flow dates: $x = (x_1, \ldots, x_N)^\top$ and $g(x) = (g(x_1), \ldots, g(x_N))^\top$
- Errors $\epsilon_i$: deviations from fundamental value, due to market imperfections (no deep, liquid, transparent market) and data errors
Estimation Problem

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

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

- Observe only $M \approx 300$ treasuries, need to estimate $N \approx 10,000$ (30 years $\times$ 365 days) discount bond prices
- Any estimation approach imposes regularizing assumptions to limit the number of parameters
- Restrict the class of potential discount curves either in terms of their functional form or their smoothness
- Existing approaches ad-hoc assumptions $\Rightarrow$ misspecified form

Our approach: Smoothness regularization

- Smoothness of the discount curve is motivated by economic principles.
- Spikes in discount curve can imply extreme payoffs
- Intuition: Similar bonds should have similar payoffs
Smooth Discount Curves

General measure of smoothness for functions

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

- **Curvature** $g''(x)^2$: penalizing avoids kinks
- **Tension** $g'(x)^2$: penalizing avoids oscillations
- **Maturity weight** $\alpha \geq 0 \Rightarrow$ corresponds to infinity maturity yield
- **Tension parameter** $\delta \in [0, 1]$ balances tension and curvature

$\Rightarrow$ We study the extremely large space of discount curves given by the set $G_{\alpha,\delta}$ of twice differentiable functions $g : [0, \infty) \rightarrow \mathbb{R}$ with $g(0) = 1$ and finite smoothness measure

**Theorem (Space of arbitrage-free discount curves)**

*Arbitrage-free discount curves are twice weakly differentiable under mild technical assumptions.*

$\Rightarrow$ Space $G_{\alpha,\delta}$ essentially without loss of generality

$\Rightarrow$ What is the right level of smoothness?
Fundamental Estimation Problem

Fundamental optimization problem:

$$\min_{g \in G_{\alpha, \delta}} \left\{ \sum_{i=1}^{M} \omega_i (P_i - C_i g(x))^2 + \lambda \left\| g \right\|_{\alpha, \delta}^2 \right\}$$  \hspace{1cm} (1)

- **Smoothness parameter** $\lambda > 0$:
  Trade-off between pricing errors and smoothness
- **Exogenous weights** $0 < \omega_i \leq \infty$ ($\omega_i = \infty$ is exact pricing):
  Set $\omega_i$ to duration weights $\Rightarrow$ approx. yield fitting
- **Problem completely determined up to the three parameters** $\alpha, \delta, \lambda$ selected empirically via cross-validation to minimize pricing errors out-of-sample.

Increasing the smoothness $\lambda$ has three effects:

- reduce excessive oscillations and curvature of the estimated function
- regularize the problem as smooth curve needs fewer parameters
- more robust estimates by penalizing outliers that cause to deviations
General Solution

Novel perspective:

- We provide a simple closed-form solution to the general problem (1)
- Usual approach: (1) Specify ad-hoc set of basis functions for a non-parametric or parametric estimation, (2) study properties of solution
- We reverse the order: Solution to fundamental problem implies the optimal choice of basis functions for the non-parametric estimation

Reproducing kernel Hilbert space (RKHS) approach:

- Builds on insights in functional analysis and machine learning
- Celebrated representer theorem: Simplifies an infinite dimensional optimization problem to a finite dimensional one
  \[ \Rightarrow \] Objective function (pricing error and smoothness measure) and function space (twice differentiable) uniquely pins down basis functions
- The solution is linear in these basis functions and a simple regression.
Theorem (Kernel-Ridge (KR) Solution)

The fundamental problem (1) has the unique solution \( \hat{g} \) in closed form

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

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

\[
\beta = C^T (CKC^T + \Lambda)^{-1}(P - C1),
\]

for the \( N \times N \)-kernel matrix \( K_{ij} = k(x_i, x_j) \), and \( \Lambda = \text{diag}(\lambda/\omega_1, \ldots, \lambda/\omega_M) \).

The kernel function \( k : [0, \infty) \times [0, \infty) \to \mathbb{R} \) is given in closed form.

Simple closed-form solution, easy to implement

- Ridge regression with smoothness as ridge penalty
- Estimated discount curve is linear in the observed prices.

\( \Rightarrow \) Discount bonds are portfolios of coupon bonds \( \Rightarrow \) Immunization

Special cases:

- Smoothing splines are special case with suboptimal parameters
- Fama-Bliss, Nelson-Siegel-Svensson, Smith-Wilson in function space
Assume Gaussian prior distribution

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

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

**Theorem (Bayesian perspective)**

If prior mean function \( m(x) = 1 \) and pricing error variance \( \sigma_i^2 = \lambda/\omega_i \), 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
Empirical Analysis
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
Benchmark models

Parametric models:

- **NSS**: Nelson-Siegel-Svensson model
- **GSW**: Gurkaynak, Sack, and Wright (2007)
  
  NSS yield curves on more restricted dataset that excludes Treasury bills

Non-parametric models:

- **FB**: Fama and Bliss (1987): piecewise constant forward curve
- **LW**: Liu and Wu (2021): kernel-smoothing method
  
  pre-specified normal kernel with specific adaptive bandwidth
  (dominates spline estimates)
Data: Maturity ranges

- Unequal maturity distribution: long maturities underrepresented
- Unbalanced panel: 30 years only available after 1985
- Sparse representation of middle maturity ranges
Cross-validated out-of-sample yield errors (in BPS)

- Vary smoothness penalty $\lambda$ and maturity weight $\alpha$ (set tension $\delta = 0$)
- Baseline choice: $\lambda = 1$, $\alpha = 0.05$, $\delta = 0$.
- Results are robust to the choice of tuning parameters.
Cross-validated out-of-sample yield errors (in BPS)

- Vary smoothness penalty $\lambda$ and tension $\delta$ (set maturity weight $\alpha = 0.05$)
- Baseline choice: $\lambda = 1$, $\alpha = 0.05$, $\delta = 0$.
- Results are robust to the choice of tuning parameters
Illustration: Yield curve estimates as a function of parameters

Varying $\lambda$, fixed $\alpha = 0.05$ and $\delta = 0$

- Representative example day: 1986-06-30
- Effect of $\lambda$: Less curvature $\Rightarrow$ bias-variance tradeoff
- Effect of $\alpha$: only affects long maturities $\Rightarrow$ $\alpha =$ infinite maturity yield

Varying $\alpha$, fixed $\lambda = 1$ and $\delta = 0$
Out-of-sample pricing errors for different maturities

- Out-of-sample evaluation on next business day
- **KR** uniformly dominates all benchmark methods along all maturities
- **KR** has smallest yield and pricing errors for all bonds
- Results also hold for cross-sectional out-of-sample evaluation
- Weaknesses: GSW short maturities, FB long maturities, LW sparse middle maturities

⇒ **KR** best uniform out-of-sample fit ⇒ closest to ground truth yield curve
Aggregated out-of-sample pricing errors

- Out-of-sample evaluation on next business day
- KR dominates all benchmark methods along all evaluation metrics
- Results are robust to various filters: full data, 3-Month filter, KR outlier filter, NSS outlier filter
- KR is most robust estimator

Full Data (no outlier removal)
Aggregated out-of-sample pricing errors

3-Month Filter

- Out-of-sample evaluation on next business day
- **KR** dominates all benchmark methods along all evaluation metrics
- Results are robust to various filters: full data, 3-Month filter, **KR** outlier filter, **NSS** outlier filter
- **KR** is most robust estimator
Illustration: Yield curve estimates of different methods

Yield curve estimate on representative example day 1961-06-30

- FB curves not smooth and overfit outliers
- GSW and NSS curves not flexible and excessive curvature in the short end
Smoothness for different curves

Discretized smoothness measure \( \int g^2(x) ds \) evaluated on discount curves

- Non-parametric methods: FB and LW are not smooth
- Parametric methods: excessive curvature of GSW and NSS in the short end due to the underlying functional form
- KR is uniformly smooth across all maturity buckets

\( \Rightarrow \) KR provides optimal tradeoff between smoothness and pricing errors
Robustness: Yield estimation with outlier contamination

- Observed and fitted yields with outlier contamination on 1963-06-28
  - Observed and implied yields of coupon bonds on representative day
  - Single outlier point: increase price by 3, 5 or 10%
  - Non-parametric kernel estimators like the LW method inherently local:
    Takes weighted average of yields of 8 nearby bonds
  - LW prone to overfitting singular outliers
  - KR benefits from global smoothness reward:
    Regularization provides robustness
  - KR estimator combines flexibility with robustness to outliers
Extrapolated KR yield curves

Varying \( \lambda \), fixed \( \alpha = 0.05 \) and \( \delta = 0 \)

Varying \( \alpha \), fixed \( \lambda = 1 \) and \( \delta = 0 \)

Yield estimation with extrapolation on example day 1986-06-30

- Extrapolation to up to 50-year maturity
- Smoother for larger \( \lambda \)
- Effect of varying \( \alpha \) is negligible for interpolation region
- Extrapolation depends on tuning parameter choices

\( \Rightarrow \) Extrapolation is a choice and not verifiable on observed data
3-standard-deviation confidence bands for yield curve estimates on 1961-06-30

- 99% confidence intervals for estimate yield curves
- Wider confidence intervals for maturity regions with less observed prices
- Confidence intervals also increase for larger price dispersion
- Observed data (up to 30y) little info about very long maturity (50y).
3-standard-deviation confidence bands for yield curve estimates on 1986-06-30

- 99% confidence intervals for estimate yield curves
- Wider confidence intervals for maturity regions with less observed prices
- Confidence intervals also increase for larger price dispersion
- Observed data (up to 30y) little info about very long maturity (50y).
Short and long maturity rate estimates over time

- One-month yield is important for research and as economic indicator
- GSW and NSS 30-day yields cannot be used
- Long maturities for FB cannot be used
- LW is not stable and prone to outliers
- Short and long maturity rates sensitive to choice of estimator
- KR provides the most reliable and stable dynamics for interest rates
Short and long maturity rate estimates over time

- One-month yield is important for research and as an economic indicator.
- GSW and NSS 30-day yields cannot be used.
- Long maturities for FB cannot be used.
- LW is not stable and prone to outliers.
- Short and long maturity rates sensitive to choice of estimator.
- KR provides the most reliable and stable dynamics for interest rates.
Short and long maturity rate estimates over time

10-Year forward rate estimates (only KR and LW)

- One-month yield is important for research and as economic indicator
- GSW and NSS 30-day yields cannot be used
- Long maturities for FB cannot be used
- LW is not stable and prone to outliers
- Short and long maturity rates sensitive to choice of estimator
- **KR provides the most reliable and stable dynamics for interest rates**
Basis Functions: Eigenvectors of the KR kernel matrix

- Eigenvectors correspond to portfolio weights in discount bonds
- Optimal basis functions: Polynomial shapes (level, slope, curvature,...)
Basis Functions: PCA of panel of discount bonds

- Similar shapes of PCs for first three PCs
- Misspecified estimators like NSS and GSW have distorted higher PCs
  ⇒ Our paper *Shrinking the Term Structure* studies asset pricing implications...
How does it matter?
Economic Importance

**Yield curve crucial input for economics and finance:**

- term structure effects and bond term premia
- forecasting, exchange rates, monetary policy
- broadly asset pricing and derivates

⇒ More precise and robust estimation benefits downstream applications
⇒ Economic implications depend on application

**Systematic biases and instabilities over time for popular methods:**

- Study *forward rates* to isolate the effect for different maturities:
  yields are essentially averages of forward rates of different maturities
- *Returns* of discount bonds:
  Term structure premium and investment implications
- Dependence on *economic states*
Forward rate curve over time

Average 1-month forward rates over time

Systematic bias of imprecise estimation methods

- KR on average lower long maturity rates (10 years, >20 years)
- FB extremely unstable time-series
- GSW distorted long maturity estimates

⇒ Potential bias for parameters estimated from KR, FB or GSW time-series
Forward rate curve over time

(a) 10-year maturity (1971/11 - 2013/12)

(b) 28-year maturity (1985/11 - 2000/01)

Standard deviation of 1-month forward rates over time

Systematically higher variation and instability for imprecise methods

- Times-series very unstable for FB and LW
- GSW small variation because of extreme bias
- KR has the most stable time-series

⇒ Smaller confidence intervals for parameters estimated with KR
Forward rates conditional on yield spread

Figure 9: Time-series of 10Y - 1Y yield spread

Average 1-month forward rates conditional on yield spread:
- Average over low, medium and high yield spread terciles
- Here large sample up to 10 year maturity, similar for longer maturities
- Systematic overestimation bias of long maturities for imprecise methods
- Bias strongest for times of low yield spread (e.g. inverted yield curves)

⇒ Precision of KR even more important for conditional analysis
Forward rates conditional on yield spread

Average 1-month forward rates conditional on yield spread:

- Average over low, medium and high yield spread terciles
- Here large sample up to 10 year maturity, similar for longer maturities
- Systematic overestimation bias of long maturities for imprecise methods
- Bias strongest for times of low yield spread (e.g. inverted yield curves)

⇒ Precision of KR even more important for conditional analysis
Forward rate curve conditional on FOMC dates

(a) 10-year maturity, unscheduled FOMC dates
(b) 28-year maturity, unscheduled FOMC dates

Average forward rate curve condition on FOMC dates

Average 1-month forward rates conditional on FOMC dates:
- FOMC meetings announce changes in monetary policy
- Here unannounced FOMC dates, but similar for announced dates
- FB and LW extremely unstable time-series
- Systematic overestimation bias of long maturities for imprecise methods
- Bias more pronounced for conditional analysis

⇒ Precision of KR even more important for conditional analysis
Returns of discount bonds

(a) 10-year maturity (1971/11 - 2013/12)
(b) 28-year maturity (1985/11 - 2000/01)

Average monthly returns of discount bonds:
- Return of discount bond fundamental for research and investment
- Only KR returns are based on tradeable portfolios of assets
- Returns of GSW, FB and LW are “artificial numbers”
- FB returns extremely unstable
- Systematic overestimation bias of long maturities for imprecise methods

⇒ Precision of KR matters for investment analysis
Discount returns conditional on yield spread

Average one-day returns conditional on yield spread:

- Average returns over low, medium and high yield spread terciles
- Here large sample up to 10 year maturity, similar for longer maturities
- Return distortion for FB and GSW more extreme for times of low and medium yield spreads
  ⇒ Precision of KR matters even more for conditional investment analysis
Broader Economic Implications

Economic implications of using non-parametric vs. parametric estimators:

- GSW yields reduce predictability because of their overly simplistic form (Cochrane and Piazzesi (2009) and Gurkaynak, Sack, and Wright (2010))
- GSW has implications for forecasting regressions and excess volatility of long-term bond prices (Liu and Wu (2021))
- GSW short-term rates are not suitable for asset pricing

Implications of KR vs. LW:

- Economic implications when (1) extremely precise estimations of the yield curve or (2) stable time-series are needed
- Differences in fourth to sixth principal components
- Filipovic, Pelger and Ye (2022) study the term structure risk premium and connection between non-parametric estimate and bond risk factors.
- Link bond risk factors to higher order PCs with substantial risk premium.
- Precise estimation of yields results in exploitable trading strategies.
- More benefits possible for sparse and noisy data of other countries
Conclusion
Conclusion

Methodology:
- Simple, fast, transparent, robust, and precise estimator for yields
- Nonparametric estimator with optimal tradeoff between flexibility and smoothness
- Learns optimal basis functions in RKHS with smoothness reward
- Closed-form solution as simple ridge regression
- Gaussian process view gives confidence ranges
- Flexible for integration of external views: priors, choice of weights

Empirical results:
- Extensive out-of-sample study on U.S. Treasury securities
- KR strongly dominates all parametric and non-parametric benchmarks
- Substantially smaller out-of-sample yield and pricing errors
- Robust to outliers and data selection choices
⇒ Method of choice for industry, regulators, central banks and researchers
⇒ Publicly available data sets: https://www.discount-bond-data.org
Appendix
Simulation

- The ground-truth discount curve is set to the KR or NSS estimates on 2013-12-31.
- Keep the maturity distribution of observed securities on the example date and assume that all securities are zero-coupon bonds.
- Obtain 10 sets of simulated noisy prices by adding independent Gaussian noise with mean zero and standard deviation one to the implied ground-truth prices that are normalized to 100.
- Estimate KR, NSS, LW, and FB from each set of noisy prices,
Simulation

In-Sample YTM RMSE (BPS)

YTM RMSE (BPS)

Discount Curve Fitting RMSE (BPS)

Yield Curve Fitting RMSE (BPS)

KR discount curve

NSS discount curve