

On the estimation of the yield curve - A simulation study

Bachelorarbeit zur Erlangung des Grades

BACHELOR OF SCIENCE (B.Sc.)

im Studiengang Volkswirtschaftslehre

an der Rheinischen Friedrich-Wilhelms-Universität Bonn

Themensteller:

Prof. Dr. Joachim Freyberger

vorgelegt im Februar 2023 von:

Marvin Lob

Matrikelnummer:

Contents

1	Introduction	1
2	Setup	1
2.1	Models	2
2.1.1	Fama-Bliss (FB)	2
2.1.2	Kernel-Ridge (KR)	3
2.2	Data-generating processes	5
2.2.1	Yield curve	5
2.2.2	Bonds	7
3	Simulations	8
3.1	Evaluation metrics	8
3.2	Evaluation setups	10
3.2.1	In-sample	10
3.2.2	Out-of-sample	10
3.3	Simulation designs	10
3.3.1	Normal yield curve	11
3.3.2	Humped yield curve	11
3.3.3	90-day filter	12
3.3.4	Shorter time-frame	12
3.3.5	Unsmooth yield curve	12
3.3.6	Varying sample size	12
3.3.7	Bias-variance decomposition	13
3.3.8	Smooth Fama-Bliss	13
4	Results	14
4.1	Base setup	14
4.2	Variations	16
4.3	Sample size and decomposition	16
4.4	Smooth FB	18
5	Conclusion	19
5.1	Economic implications	19
5.2	Limitations and possible extensions	20
Appendix A	Proofs and definitions	21
Appendix B	Additional figures	24
References		28

1 Introduction

In the years following up on the financial crisis of 2008 interest rates were at an all-time low. Consequently, assets like bonds fell out of investors' favor meanwhile occupying oneself with the underlying yields was less than unglamorous. However, the sharp interest rate increases in the wake of historical inflation surges have put a larger emphasis on yields and bonds. Although there are countless bond markets and subsequent yield curves, the focus in research and industry remains on US treasuries. This stems from multiple factors. First and foremost, US treasuries are considered to be an investment that is as risk-free as it can get. Hence, the yields implied by US treasuries are often seen as a risk-free rate for investing (Berk and DeMarzo, 2019). Its importance can not be overstated as widely recognized models in portfolio theory like CAPM are based on this risk-free rate (Fama and French, 2004). Additionally, financial products with interest like payments, i.e. swaps or options, use rates that are closely tied to those of US treasuries (Hull, 2003). Lastly, a better estimation of the yield curve could provide various useful trading strategies since, for example, the prices of bonds are directly connected to the yield curve (Filipović et al., 2022). Estimating the yield curve is therefore of fundamental importance for various applications. Previous papers have proposed many different parametric and non-parametric estimators for the yield curve, but the comparisons were almost exclusively only focused on historical data for US treasuries thus, limiting the explanatory power of its results (Bliss, 1996; Jeffrey et al., 2006; Liu and Wu, 2021).

In my thesis, I will test the classic method by Fama and Bliss (1987) against a newly proposed method by Filipović et al. (2022). Unlike other comparisons studies, I establish data-generating processes (dgp) that allow me to test both methods in distinct scenarios which offer me more profound assertions of the estimators' abilities. The dgp are constructed in a way that mimics the current listings of US treasuries while granting enough flexibility to establish various simulation setups. This thesis is structured in the following way: In section 2, I will describe the challenges in estimating the term structure and introduce both estimators. Additionally, my data-generating processes for yield curves and bonds are presented. I proceed in section 3 by defining evaluation metrics and the simulation designs. Section 4 unveils and asserts the results. And last but not least, section 5 concludes with a discussion of the gathered insights and potential extensions.

2 Setup

The zero-coupon yield curve for US treasuries is a function of time that describes the hypothetical yield of a US treasury bond without coupon payments. I note this zero-coupon yield curve as $y(t)$. Its estimation is not straightforward because most US treasury bonds have coupon payments and are not directly expressed by the yield curve itself. Therefore, estimating the yield curve requires additional assumptions and related concepts. Two related curves are the discount curve $g(t)$ and the forward rate curve $f(t)$. The former expresses the price of a zero-coupon bond with t years to maturity while the latter can be interpreted as expectations about future yields

(Svensson, 1994). They are related by the equations:

$$g(t) = \exp\left(-\int_0^t f(x) dx\right), \quad (1)$$

$$g(t) = \exp\left(-y(t) \times t\right). \quad (2)$$

I consider daily values for the yield curve and calculations will be based on the assumption that interests are continuously compounded. Primarily because this is standard in the finance literature, but also because it is sufficiently accurate since the days are measured on a yearly time scale t , meaning the intervals are small enough for expressing the yield, discount, and forward rate curve with the limits defined by continuous compounding. Ultimately, yield curve estimators exploit the relationship between cash flows and prices of observed bonds. In theory, it is often assumed that markets offer no opportunity for arbitrage, and therefore, the fair price P_{true}^i of a security i should be equal to its discounted future cash flows (Berk and DeMarzo, 2019). If this assumption holds then the observed price P_{obs}^i equals the true price P_{true}^i which can be expressed directly by the discount curve and the bond's J^i future cash flows $\{c_j^i\}_{j=1}^{J^i}$:

$$P_{true}^i = \sum_{j=1}^{J^i} c_j^i g(t_j^i). \quad (3)$$

Here, $\{t_j^i\}_{j=1}^{J^i}$ indicate the time at which cash flows occur. However, in practice, it can not be assumed that P_{true}^i is actually observed due to market imperfections (Filipović et al., 2022). Hence, the observed price P_{obs}^i is often assumed to be contaminated by a noise component ϵ^i :

$$P_{obs}^i = P_{true}^i + \epsilon^i. \quad (4)$$

Now, even with a large number of observable bonds, most dates in the time frame for which the yield curve should be estimated have little to no cash flow occurring. For example, on a 30-year time frame there are almost 11000 days for which the yield curve must be estimated but cash flow will only occur on a small fraction of those days. Consequently, yield curve estimators that use equation 3 have to make additional assumptions because the number of equations is far less than the unknown values. While parametric methods like those proposed by Nelson and Siegel (1987) or Svensson (1994) postulate a specific parametric form, non-parametric methods like in Fama and Bliss (1987) or Filipović et al. (2022) solely assume general properties about the curves. My analysis tests and compares both of these non-parametric methods.

2.1 Models

2.1.1 Fama-Bliss (FB)

Fama and Bliss's method (FB) is a relatively old but still widely used non-parametric method. It uses the relationship between the forward rate and discount curve (equation 1) by assuming that

the forward rate is a piece-wise constant function with intervals determined by the maturities of observed bonds (Fama and Bliss, 1987). It is an iterative approach where the forward rate for an interval is calculated by taking into account the forward rate estimations of previous intervals. That is why it is sometimes referred to as the "bootstrap" method for curve estimation (Bliss, 1996). Many papers use the estimations made by FB which can be downloaded from data vendors like CRSP¹. My implementation is based on Jeffrey et al. (2006) who use slightly different notation. In order to understand their algorithm, let the number of payments of security i be expressed by m^i and its cash flows by $\{b_j^i\}_{j=1}^{m^i}$ then the estimation of the forward rate curve is described by Jeffrey et al. (2006) as follows:

[...] let the sequence of observed bonds $\{P_j^i\}_{j=1}^{N^i}$ be ordered from the shortest maturity to longest maturity and let τ^i denote the time-to-maturity of the i^{th} bond. Let F^i denote the constant forward rate on the time-to-maturity interval $(\tau^{i-1}, \tau^i]$ where $\tau^0 = 0$, that is $f(\tau) = F^i$ for $\tau \in (\tau^{i-1}, \tau^i]$. The discount function now takes the form $d(\tau) = \exp(-F^K(\tau - \tau^{K-1}) - \sum_{k=1}^{K-1} F^k(\tau^k - \tau^{k-1}))$ where K is chosen so that $\tau \in (\tau^{K-1}, \tau^K]$. To extract (bootstrap) the forward rate curve proceed as follows. First determine F^1 by considering the shortest maturity instrument and solve for F^1 in $P^1 = \sum_{j=1}^{m^1} b_j^1 \exp(-F^1 \times \tau_j^1)$. Now consider the second shortest instrument and solve for F^2 in $P^2 = \sum_{j=1}^{m^2} b_j^2 d(\tau_j^2)$ given F^1 , and so on. In general, to bootstrap F^i use the i^{th} observed bond and find the F^i that solves $P^i = \sum_{j=1}^{m^i} b_j^i d(\tau_j^i)$ where the sequence $\{F^j\}_{j=1}^{i-1}$ has been computed from previous bonds in the same fashion.

Evidently, at the i^{th} iteration, F^i is only taken into account by the payments taking place after τ^{i-1} . As a result, the estimation could be extremely variable and susceptible to outliers which might be seen in the later simulations. Additionally, the last day to which the forward rate is estimated is determined by the longest maturity instrument. Thus, FB can only be used for interpolation and not for extrapolating on the time frame after the longest observed maturity. My implementation differs from Jeffrey et al.'s proposal and other implementations, as I strictly allow for multiple bonds with the same time-to-maturity. This allows the estimator to use more available data from the underlying yield structure. If there are no identical maturities then FB fits each bond's observed price exactly, meaning the price for bond i implied by the estimation P_{imp}^i equals the observed price P_{obs}^i (Jeffrey et al., 2006). By allowing for duplicates, this property is not always true for bonds that share the same time-to-maturity. In the presence of duplicates, I estimate the forward rate by minimizing the implied, unweighted, and squared pricing error for bonds with the same time-to-maturity (see definition 1). Hence, exact pricing is possible but not guaranteed as it depends on the mispricing of bonds.

2.1.2 Kernel-Ridge (KR)

In contrast to Fama and Bliss (1987), the estimator proposed by Filipović et al. (2022) is not attempting to price each bond exactly. It minimizes a global mispricing criterion while providing a smoothness reward thus, preventing multicollinearity. They assume that the discount curve is

¹ Center for Research in Security Prices

a smooth function for which smoothness can be measured. Specifically, they introduce the norm

$$\|g\|_{\alpha,\delta} = \left(\int_0^\infty \left(\delta g'(t)^2 + (1 - \delta) g''(t)^2 \right) \exp(\alpha t) dt \right)^{\frac{1}{2}} \quad (5)$$

as a measurement of smoothness for the discount curve g with tuning parameters $\alpha \geq 0$ and $\delta \in [0, 1]$. This means that the considered functions have to be twice weakly differentiable (Filipović et al., 2022). Evidently, larger values for α increase the norm and thus, the smoothness penalty. The weighting increases with the time horizon since α is multiplied by t . Therefore, α allows for a weighting based on time where early flexibility is less punished than a late one. Additionally, Filipović et al. (2022) show that α can be interpreted as the yield of a discount bond with infinite maturity. This offers KR a limiting property that allows for extrapolating over time-frames that lie after the longest observed maturity. So, KR has a direct advantage compared to FB which can not extrapolate without additional assumptions. The second tuning parameter δ trades off the first two derivatives, meaning how much the function is allowed to oscillate versus the prevalence of kinks. In addition, they use

$$\omega_{obs}^i = \left[M (D_{obs}^i P_{obs}^i)^2 \right]^{-1} \quad (6)$$

as weights for the pricing error implied by the estimation. Here, M is the number of observed bonds and D_{obs}^i is the observed duration of security i (definition 2). The duration offers information about how sensitive the price of a bond reacts to the yield curve (Brealey et al., 2012). Therefore, it accounts for the maturity of a bond because prices of bonds with longer maturities behave more extremely to changes in the yield curve. Filipović et al. (2022) use these weights as a means to account for possibly larger pricing errors of bonds with longer maturities. Their minimization problem is formulated by

$$\min_{g \in \mathcal{G}_{\alpha,\delta}} \left\{ \sum_{i=1}^M \omega_{obs}^i \left(P_{obs}^i - P_{imp}^i(g) \right)^2 + \lambda \|g\|_{\alpha,\delta}^2 \right\}, \quad (7)$$

where $P_{imp}^i(g)$ is the implied price for the bond i and $\lambda > 0$ is a smoothness penalty. The space of functions $\mathcal{G}_{\alpha,\delta}$ with its associated norm (5) is additionally restricted by setting $g(0) = 1$ and $g : [0, \infty) \rightarrow \mathbb{R}$. A solution is derived by using the properties of the induced space $\mathcal{G}_{\alpha,\delta}$. The authors show that this space is a reproducing kernel Hilbert space (RKHS) with a corresponding reproducing kernel (RK). Filipović et al. (2022) then use a variation of the representer theorem (Wahba and Wang, 2019) that states how the solution \hat{g} can be expressed as a linear combination of the reproducing kernel k over each day in the time frame:

$$\hat{g}(t) = 1 + \sum_{j=1}^N k(t, t_j) \beta_j. \quad (8)$$

Here, N is the last day in the considered time frame, and for my purpose t is measured in years, thus $t_j = j/365$.

This solution is linearly dependent on $\beta = (\beta_1, \dots, \beta_N)$ which Filipović et al. (2022) show is the solution to a kernel-ridge regression and so estimated by

$$\hat{\beta} = C^T (CKC^T + \Lambda)^{-1} (P - C\mathbf{1}), \quad (9)$$

where $C \in \mathbb{R}^{M \times N}$ is the cash flow matrix with rows for each bond and columns for each day. The kernel matrix $K \in \mathbb{R}^{N \times N}$ has elements $\{k(t_l, t_j)\}_{l,j=1}^N$, where $t_l = l/365$ and $t_j = j/365$ are the days in the time frame measured in years. Next, the distinctive ridge-penalization is characterized by the diagonal weight matrix Λ with elements $\{\lambda/\omega_{obs}^i\}_{i=1}^M$. Lastly, the bonds' observed prices are grouped in the price vector $P \in \mathbb{R}^M$ and $\mathbf{1}$ denotes a vector of N ones. The RK implied by the norm (5) depends on the values chosen for α and δ . Filipović et al. (2022) examine optimal tuning parameters and propose $\alpha = 0.05$ and $\delta = 0$. These values imply the RK

$$k(x, y) = -\frac{\min(x, y)}{\alpha^2} \exp(-\alpha \times \min(x, y)) + \frac{2}{\alpha^3} (1 - \exp(-\alpha \times \min(x, y))) - \frac{\min(x, y)}{\alpha^2} \exp(-\alpha \times \max(x, y)).$$

Other tuning parameters induce other RK and are therefore associated with other RKHS. They are not used in my thesis but can be found in Filipović et al. (2022). In addition to $\alpha = 0.05$ and $\delta = 0$, my simulations use their recommended penalty of $\lambda = 1$. One might argue that a separate analysis is necessary in order to determine optimal tuning parameters for each later used simulation setup. But, since Filipović et al. (2022) advertise their method as "easy to implement", using only their recommended tuning parameters tests this alleged advantage. Fine-tuning α , δ , and λ for each setup would require extensive efforts which could not be seen as an easy way of implementing their method.

2.2 Data-generating processes

The challenge in testing estimators often comes down to the available data. In the case of US treasuries, it is not only difficult and expensive to acquire the necessary bond data but also, the available data is just 'one history'. Results solely based on historical data risk being due to chance and are, therefore, not well suited for general statements. My simulations, build on two data-generating processes (dgp) that together provide a wide range of possible scenarios to test both introduced models. First, a yield curve is sampled with a specific shape after which a portfolio of coupon bonds is then built which mimics the maturity distribution of US treasury bonds.

2.2.1 Yield curve

The main focus of this thesis is the evaluation of yield curve estimators. Thus, the sampling process for yield curves is of fundamental importance for my analysis. I envisioned a process that allows for testing different scenarios as well as being challenging to estimate. In addition, I created a process that samples ever so slightly different yield curves for the same set of tuning parameters as this allows me additional interpretations of the results later on (3.2.2). The sampling process has two components. First, my yield curve starts at zero and for each following

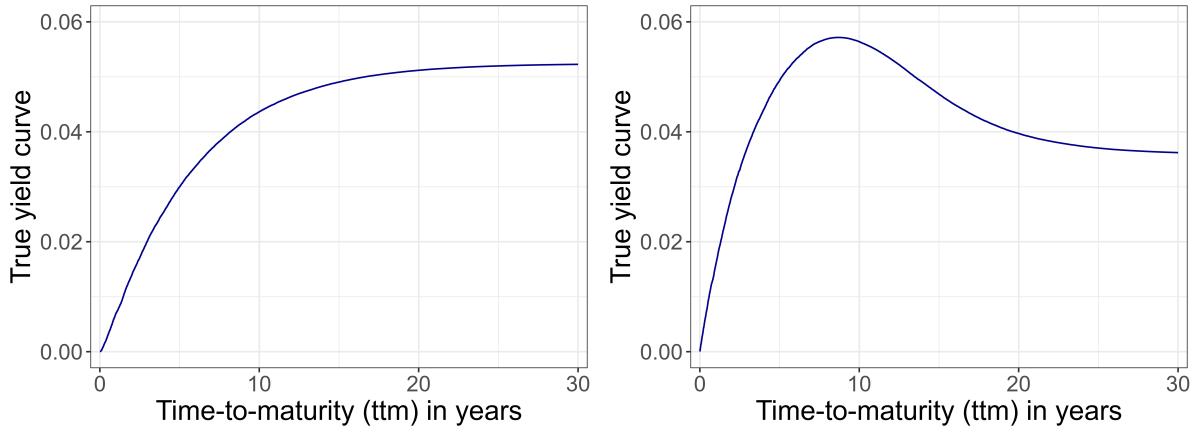


Figure 1. Normal and humped yield curve samples from the data-generating process.

day, a draw from a uniform distribution is added. Second, each addition is weighted by a weight function that depends on the time and a set of tuning parameters. Formally, let N be the last day to which the yield curve should be sampled and $\mathcal{N} = \{1, 2, \dots, N\}$ the set of days in the time frame. The data-generating process for the yield curve $\{Y_n : n \in \mathcal{N}\}$ is defined by

$$Y_n = \sum_{x=1}^n U_x \times \omega(x|\theta), \quad (10)$$

where $U_x \sim U(0, 0.1)$ and $\theta = \{a, b, c, d, e\}$ is a set of tuning parameters. The weight function $\omega(x|\theta)$ is given by

$$\omega(x|\theta) = a \left(\exp \left(-\frac{0.15 \times x}{b} \right) - c \times f_{LN}(x | \mu, \sigma^2) \right), \quad (11)$$

where $f_{LN}(x | \mu, \sigma^2)$ is the density function of a log-normal distribution (definition 3) with $\mu = \log(d)$ and $\sigma = \log(\sqrt{e})$. This setup allows for sampling slightly different yield curves even when θ remains fixed. The weight function itself comprises an exponential and a positively skewed function. The former allows for a classical increasing but converging yield curve, while the latter permits special shapes. Theoretically, there are many functions that could be used for the skewed function. My process uses the density function of a log-normal distribution because of its shape and easy implementation in *R*. The tuning parameters in θ jointly define the converging speeds, overall yield levels, and tail shapes of the sampled yield curves. Their values claim no financial interpretation and are simply set to produce compelling shapes for the estimation problem. Two sampled yield curves from this dgp with different tuning parameters are seen in figure 1. Here the normal-shaped yield curve is produced with $\theta_{norm} = \{0.0006, 300, 250, 500, 70\}$ and the humped shape with $\theta_{humped} = \{0.001, 200, 600, 5000, 2\}$. Ultimately, this dgp creates yield curves that are stochastic processes and have no fixed parametric form. Hence, it could pose a real challenge for estimators.

2.2.2 Bonds

In order to compare FB and KR, a portfolio of bonds with prices and coupon payments is needed. My dgp for bonds samples a certain number of bonds with a realistic payment structure and maturity distribution while allowing for different tuning parameters to create various simulation setups. I implement the number of sampled bonds, the intensity of pricing errors, and the usage of a 90-day filter as tuning parameters. The 90-day filter allows only for bonds with longer maturities than 90 days. This follows standards in the yield curve estimation literature and stems from the observation that bonds with extremely short maturities are more susceptible to larger mispricing (Gürkaynak et al., 2007). To ensure a realistic maturity distribution, I collected a current listing of all US treasuries from *The Wall Street Journal* (2023) and performed a kernel density estimation with a Gaussian kernel and bandwidth selected by Sheather and Jones's (1991) proposed plug-in method. The resulting estimation can be seen in figure 2. Here, the vertical blue lines indicate the actual maturity of the listed bonds and show the relative lack of bonds maturing 10 to 20 years from now. The absence of bonds in this time interval is also seen in other recent papers like Liu and Wu (2021) or Filipović et al. (2022) and poses a severe interpolation challenge for the estimators. The whole process of creating a portfolio of M bonds is now conducted as follows. For every bond, a maturity is sampled by drawing from a normal distribution with the mean equal to a maturity randomly chosen from the observed listings and the standard deviation equal to the bandwidth estimation. As a result, the maturity distribution of the sampled bonds has the kernel density estimator as its density function (see theorem 6). Admittedly, the actual density will not be exactly equal because I only allow maturities longer than one day (or 90 days if the 90-day filter is used) and shorter than a certain maximal maturity (in most cases 30 years). Afterward, the coupon of a bond is determined by sampling from a uniform distribution with a range inspired by the listings used for the density estimation¹. Payments are set semi-annually, in line with the payment structure of US treasuries (U.S. Department of the Treasury, 2023a). This amounts to J^i payments for each bond i . These payments $\{c_j^i\}_{j=1}^{J^i}$ occur on $\{t_j^i\}_{j=1}^{J^i}$. Next, a sampled yield curve is used to calculate the true prices $\{P_{true}^i\}_{i=1}^M$ of the bonds. This is achieved by transforming the yield curve via its relationship with the discount curve (see equation 2) which can then be used for discounting the payments. I allow the observed prices $\{P_{obs}^i\}_{i=1}^M$ to be contaminated in an additive way as depicted in equation 4. The noise is sampled from a Gaussian distribution with mean zero and standard deviation chosen as a tuning parameter. However, before adding the noise I multiply it by a scaling function that increases

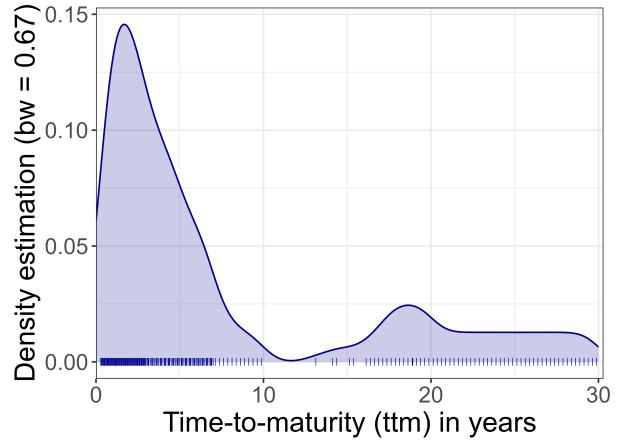


Figure 2. Kernel density estimation of the maturities of listed US treasuries.

¹ It is sampled from a $U(2, 4)$ -distribution.

with the bond's maturity. This prohibits unrealistic large price and subsequently yield deviations for bonds with short maturities while allowing for still significant effects on bonds with long maturities. An illustration of the sampling process for different noise levels is given in figure B.1 showing the observed yield-to-Maturity (YTM) for a fixed set of 200 zero-coupon bonds. The observed YTM of bond i is calculated by solving for ytm_{obs}^i in

$$P_{obs}^i = \sum_{j=1}^{J^i} c_j^i \exp(-t_j^i \times ytm_{obs}^i), \quad (12)$$

which simplifies for zero-coupon bonds because they only have a principal. The YTM of a bond equalizes the bond's discounted cash flows with its price. It expresses the yield an investor gets if she holds the bond until maturity with all payments paid on time (Berk and DeMarzo, 2019). I specifically illustrated the YTM for zero-coupon bonds because, in absence of noise, the YTM of zero-coupon bonds lie exactly on the yield curve which is not true for coupon bonds. Thus, all fluctuation seen in the figure results from the inflicted noise. In the end, a sampled portfolio consists of M bonds each having constant, semi-annual cash flows and an observed price.

3 Simulations

The simulations were implemented with *R* (R Core Team, 2022) using *RStudio* (Posit team, 2022). Plots were made with the *ggplot2* package (Wickham, 2016) which is part of the *tidyverse* package (Wickham et al., 2019) that was generally used as well. The implementation of the estimators required a solver for nonlinear equations for which I used the *nleqslv* package (Hasselman, 2022). Dates were managed with the *lubridate* package (Grolemund and Wickham, 2011). The complete code used for my simulations and plots can be found in my GitHub repository¹.

3.1 Evaluation metrics

Simulations require metrics to determine how well a method performs. An intuitive metric for curve estimation would be the difference between the true yield curve and its estimation. However, the prices implied by the yield curve estimation comprise estimations for many days and so, could offer more insights into the estimators' performances than just the yield curve estimation itself. More importantly, FB suffers by construction from a large variability, and thus its performance would deteriorate on any criteria that evaluate differences between the estimated and true yield curve (see figure 4 later on). I introduce two versions of what Filipović et al. (2022) call "root-mean-squared errors" (RMSE). The first metric uses observed prices $\{P_{obs}^i\}_{i=1}^M$ while the second uses true prices $\{P_{true}^i\}_{i=1}^M$. Both metrics determine the weighted differences between their used and implied prices. Here, the implied prices $\{P_{imp}^i\}_{i=1}^M$ are calculated by discounting the cash flows with the yield curve estimation. So for a given set of M bonds, the two metrics are defined by

¹ https://github.com/malo4300/bachelor_thesis

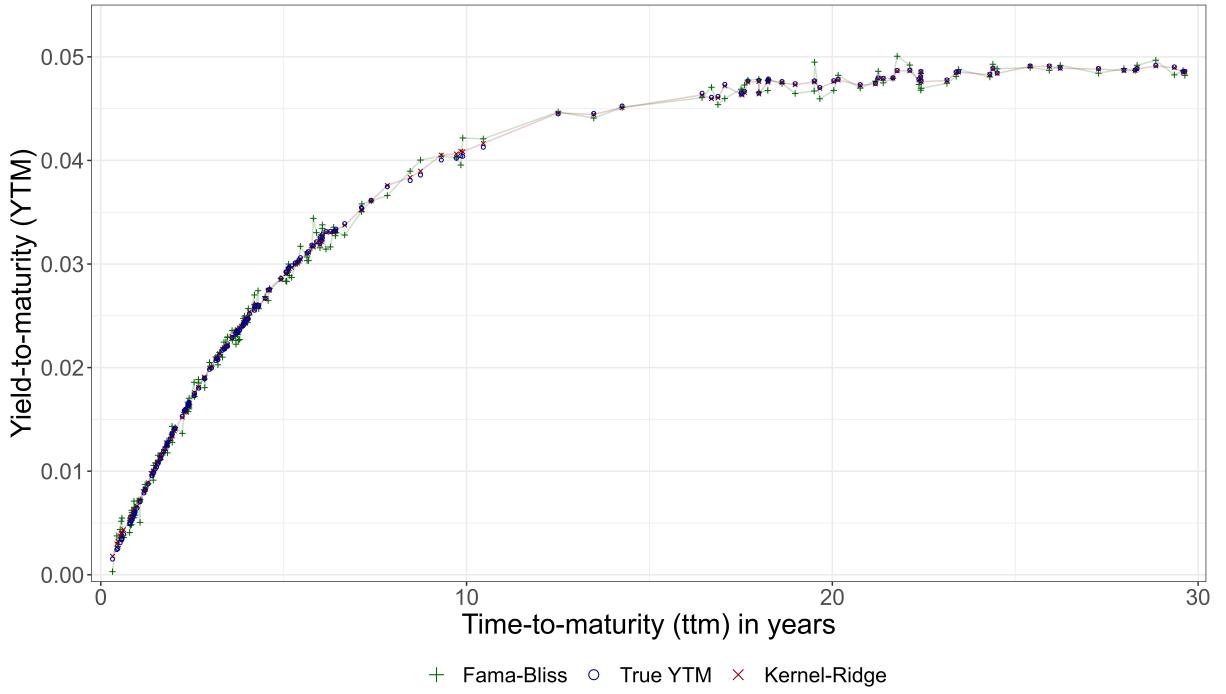


Figure 3. True YTM and the estimations for a 200-bond portfolio with an underlying normal yield curve. The noise parameter was set to 1.

$$RMSE_{obs} = \sqrt{\sum_{i=1}^M \omega_{obs}^i (P_{obs}^i - P_{imp}^i)^2}, \quad (13)$$

$$RMSE_{true} = \sqrt{\sum_{i=1}^M \omega_{true}^i (P_{true}^i - P_{imp}^i)^2}. \quad (14)$$

Noticeably, my definition distinguishes between true and observed weights which is due to the way they are calculated. A true weight ω_{true}^i is based on the true price, while an observed weight ω_{obs}^i is calculated with the observed price that was treated with noise. For my simulations, I use the weights that were already defined in section 2.1.2. There are two reasons why I use the weights proposed by Filipović et al. (2022). First, the scaling with the duration limits the effect of mispricing bonds with longer maturities as those are more susceptible to discrepancies between the true and estimated yield curve. Since FB's performance can deteriorate for long maturities, scaling could help FB in the evaluation (Filipović et al., 2022). Second, unlike other proposed metrics (Liu and Wu, 2021), these weights include the bonds' prices. Filipović et al. (2022) show that the inclusion of prices induces an evaluation metric that is approximately equivalent to the average yield-to-maturity error (definition 7). Thus, these weights allow me to evaluate the estimators from an additional perspective. Figure 3 illustrates an example fit by both estimators to the YTM. Here, I sampled a normal-shaped yield curve, but an example fit using a humped yield curve can be found in the appendix (B.2). Keep in mind that the curve shape seen here is not identical to the yield curve itself. The YTM can be different for bonds with the same maturity since the coupon payments can vary. I use both evaluation metrics in two distinct ways that I call in-sample and out-of-sample evaluation.

3.2 Evaluation setups

3.2.1 In-sample

The first form of evaluation is based on the same prices and cash flows that both methods used for estimating the yield curve. This means that on the one hand, I use the observed prices $\{P_{obs}^i\}_{i=1}^M$ for calculating the $RMSE_{obs}$. On the other hand, I use the prices without noise $\{P_{true}^i\}_{i=1}^M$ for calculating the $RMSE_{true}$. Both evaluation metrics combined can hint at the degree to which the estimators are fitting the noise i.e. overfitting.

3.2.2 Out-of-sample

As previously mentioned, my dgp always generates slightly different yield curves for the same tuning parameters. Nevertheless, a sampled yield curve always falls around some common trend. In my case, the yield curves are falling around a mean determined by the tuning parameters of the weight function¹. In my opinion, there is an argument to be made that in the real world, true yield curves might behave in a similar way. Given the central bank rates, expectations, and the economic as well as political circumstances, there are probably still remaining factors left that influence the yield curve. Assuming those influences, e.g. marked imperfections, occur at random, the real-world yield curve might be resampled from an underlying stochastic process at certain times. Actually, there already exists research that uses stochastic processes to model yield curves like Brennan and Schwartz (1979) or Vasicek (1977). My out-of-sample evaluation assumes a daily resampling and is conducted by first sampling a yield curve and with it, a set of bonds. Subsequently, the models are fitted before a new yield curve with the same tuning parameters is sampled. Next, the cash flows are shifted by one day and discounted with the new yield curve. Here, the true prices $\{P_{true}^i\}_{i=1}^M$ are the shifted and discounted cash flows using the new yield curve, whereas the observed prices $\{P_{obs}^i\}_{i=1}^M$ were again treated with random noise. Consequently, the $RMSE_{obs}$ could be seen as a measurement of how well the methods can predict the next day's prices while the $RMSE_{true}$ could offer insights into how well the estimators fit the underlying stochastic process. Of course, these interpretations depend on the previously made assumptions about the stochastic behavior of yield curves. Regardless of these assumptions, my out-of-sample evaluation can still be used to evaluate an estimator's tendency to overfit since the noise is uncorrelated and inflicted at random. Thus, methods that fit the noise will experience deteriorating performances.

3.3 Simulation designs

After introducing the evaluation metrics it is necessary to consider the scenarios estimators are going to face. The idea behind creating data-generating processes is to allow repeated sampling to approximate the average performance for each estimator in a given scenario. So, as a base parameter setting, the number of iterations for each simulation, meaning how often a new yield curve and bond portfolio are sampled, models are fitted, and evaluation metrics calculated is

¹ A visualization of this pattern can be found in the appendix (figure B.3).

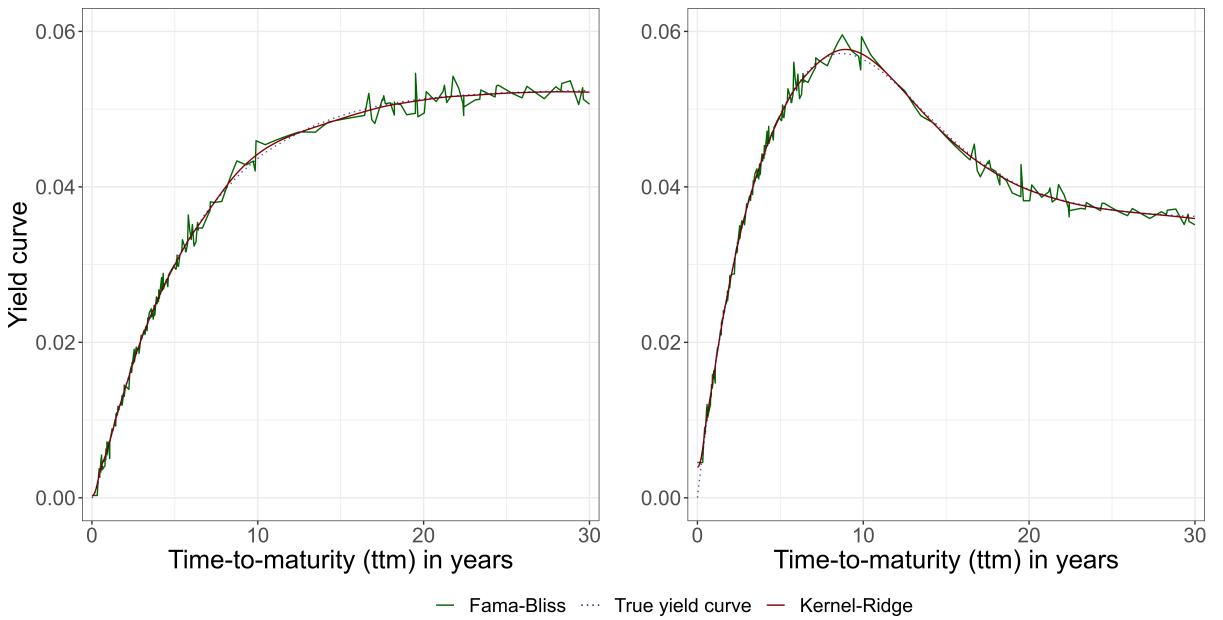


Figure 4. Yield curve estimations for a normal and humped yield curve.

set to 100. The average results for the 100 iterations are discussed in section 4. If not otherwise indicated, the number of sampled bonds will be 200 in a 30-year time frame. This is also the maturity interval for which real US treasuries are observed. Additionally, most simulations only allow for bonds with maturities longer than 90 days. Except for the 'Varying sample size' simulation (3.3.6), each setup runs on a grid of different standard deviations for the Gaussian noise added in the pricing process. This grid ranges from 0 (no noise) to 2 (high noise). The impact of the noise parameter was already depicted in figure B.1.

3.3.1 Normal yield curve

My base simulation uses a normal yield curve with the same set of tuning parameters θ_{norm} that was defined in section 2.2.1 and visualized on the left side of figure 1. The resulting shape of the yield curve is generally seen as the 'standard' shape as maturity and yield increase together over time. This simulation is later used as a baseline for the other scenarios. I illustrate the estimations made by the models for one such sampled yield curve in figure 4 next to a fit for a humped yield curve.

3.3.2 Humped yield curve

Contrary to a normal shape, a yield curve can have various unusual shapes, for example, a humped shape which this setup introduces. Here, the medium-term yields are higher than the long-term yields. This shape might not be seen regularly in financial markets. However, exactly this uniqueness might pose a problem for estimators. Especially KR which rewards smoothness might struggle with this unexpected shape since its tuning parameters remain fixed. Here, the yield curves are sampled with the same set of tuning parameters θ_{humped} that was used for the humped yield curves in figures 1 and 4. This special shape might also be interesting because a humped yield curve can arise when investors expect interest rates to rise in the medium future

while falling long-term (Nag and Ghose, 2000). Recently, prominent investors like Michael Burry expressed their belief this might happen when further interest rate increases cause an economic downturn in which rates have to be decreased again (Mohamed, 2023).

3.3.3 90-day filter

In my base setup, I only allow for bonds with maturities longer than 90 days. This follows remarks from Gürkaynak et al. (2007) that note how soon-maturing "securities often seem to behave oddly" due to their lack of liquidity. Hence, practical implementations of yield curve estimators often exclude these securities. In addition, the noise infliction when determining the observed prices could induce unrealistic large variations in the prices of soon maturing bonds and thereby in their yield-to-maturities despite the noise-weighting by maturity (see 2.2.2). Nevertheless, the inclusion of bonds with shorter maturities might offer additional insights into the general practicability of the estimators. Liu and Wu (2021) also point out that bonds with shorter maturities could entail important information that should be included. Furthermore, in the original paper, Fama and Bliss (1987) do not exclude soon maturing bonds either. Thus, the setup without a 90-day filter might be a more appropriate setting for FB.

3.3.4 Shorter time-frame

Fama and Bliss (1987) stems from a time when not only the number of bonds was far lower than today (see 3.3.6) but also the average maturity was significantly lower. As an example, the first non-callable 30-year bond was only issued two years prior to the publication of the original paper (U.S. Department of the Treasury, 2023b). Consequently, it would come as no surprise to observe deteriorating performances of FB when many bonds with long maturities are included. So in this setup, I will only allow for bonds with maturities of up to 10 years while also decreasing the number of bonds to 100 as this number should respond to the shorter time frame.

3.3.5 Unsmooth yield curve

Thus far, the sampled yield curves have one yield increase for each day in the time horizon. Although this results in a relatively smooth shape, the forward rate induced by this yield curve is constant for each day. Thus, FB's constant forward rate assumption on maturity intervals would hold if we had daily observations. As this is not the case, FB struggles in finding the true yield curve while daily jumps seem to be smooth enough for KR (see figure 4). A new less smooth yield curve might challenge KR's ability to fit its smooth structure to the yield curve. My altered dgp for unsmooth yield curves is constructed by changing the frequency and height of yield increases. In detail, my altered dgp produces yield curves that have increases only every seven days in exchange for being on average seven times larger than before. The resulting yield curve is much less smooth and the performances of FB and KR are not obvious.

3.3.6 Varying sample size

Up until now, the simulations only considered a fixed sample size of 200 bonds. However, a comparison of FB and KR should also include an analysis of the methods' behavior for a varying

number of bonds. To accomplish this, I set the standard deviation of the Gaussian pricing noise to 1 and perform simulations for an increasing number of bonds starting at 50 and leading up to 200. Two notes on why this setup could be especially interesting. First, Fama and Bliss (1987) is a relatively old method that was proposed at a time when the number of US treasuries was much lower than today. Thus, a smaller number of observations might benefit its performance in comparison to the new KR method. Second, since I do not change KR's tuning parameters it could be intriguing to see how this affects the threat of overfitting as KR's weights put a larger emphasis on every single bond for a smaller sample size M . In general, better performance of both estimators should be expected because more observed bonds imply more available information about the true, underlying yield curve.

3.3.7 Bias-variance decomposition

My simulations only calculated the RMSE as a whole. Since it is the root of a 'mean-squared error' (MSE) normally, it could be decomposed into the squared bias and variance of the estimator (Hastie et al., 2009). Considering this, an estimator can improve its performance measured by the MSE either by decreasing the bias or the variance which is known as a bias-variance trade-off. Now, the bias-variance decomposition in the context of curve estimation is not straightforward because neither FB nor KR performs simple point estimation. However, I show that the expected MSE_{true} which is just the squared $RMSE_{true}$ can be decomposed into a variance and bias part (theorem 8). An accurate estimation of both parts requires multiple iterations with the same yield curve and bonds. Otherwise, each iteration would have different true prices and so, calculations of variance and bias over all iterations would be nonsense. I conduct a simulation where I fix cash flows and a normal-shaped yield curve. Consequently, only the random mispricing induces a change in the estimations. Building on this, I calculate the variance and average squared bias of the implied prices for each fixed bond separately. Next, variances and biases are weighted with the usual weights before adding them together. This results in an estimation of the variance and bias part as seen in theorem 8. Unfortunately, the results only describe an estimator's behavior to the noise-inflicting process, not the randomness induced by the coupon, maturity, or yield curve sampling considering those remain fixed. Nevertheless, I think the results may still provide appealing insights into the driving factors of the estimators' performances.

3.3.8 Smooth Fama-Bliss

Even before regarding the results, FB's lack of smoothness is something of concern. Subsequently, even Bliss himself considered ways of smoothening the estimator (Bliss, 1996). However, his approach was to fit a pre-defined function to FB's estimations which makes it a parametric method. My 'smooth FB' model retains the previous flexibility of the non-parametric model by choosing a cubic smoothing spline for fitting the estimations made by FB. A cubic smoothing spline consists of piece-wise defined cubic functions with intervals marked by 'knots'. These are normally set at every observation¹. The first and second derivatives of two neighbor-

¹ For computational purposes, I allow R's smoothing spline function to choose a smaller number of knots without an impact on performance.

ing piece-wise functions are forced to be equal at the knots. This ensures a smooth transition from one to another interval. Consequently, an observer can not detect kinks. To avoid 'weird' behavior at the extrapolation regions, the functions before the first and after the last knot are forced to be linear. Generally, this method would badly overfit the observations because of the high degrees of freedom but this can be controlled by a smoothness penalty on the second derivative (Hastie et al., 2009). Since knots are placed automatically, the model only requires the penalty as input which I optimally set by leave-one-out cross-validation (LOOCV). There have been similar approaches for fitting the yield curve with smoothing splines (Fisher et al., 1995). My implementation differs from other proposals as I use FB's estimations as input where each bond is one observation with its time-to-maturity and yield curve estimation at that date. There are two important considerations. First, the penalty selection by LOOCV results in fine-tuning the model to the used setup, something I specifically prohibited for KR. Hence, smooth FB gets an unfair advantage. Second, by smoothing FB with this model, I introduce almost the same regularization as KR which hurts the main goal of comparing both methods on their own. Actually, for $\alpha = 0$ and $\delta = 0$ KR's penalty (5) is equal to the smoothness penalty used for determining the smoothing spline (Hastie et al., 2009). That is why smooth FB and KR might produce similar results. But, smooth FB has the advantage that a user can implement it much easier and faster. KR is quite new and so, its implementation is left to the user. On the contrary, FB's estimations are widely available. Additionally, smoothing splines are easily trained thanks to existing implementations like in plain *R*. Consequently, if smooth FB and KR produce similar results then a pragmatic user might prefer smooth FB over KR.

4 Results

The simulation results are presented in four stages. First I will look at the base setup and small variations before considering more elaborate setups and finally smooth FB. All results are presented with plots either in the main text or in the appendix. The RMSE has only a limited economical interpretation, but since the weights include the bond's prices it can be seen as a percentage. An additional interpretation is possible as the RMSE is approximately equal to the yield-to-maturity error.

4.1 Base setup

Starting off with the in-sample evaluation of the base setup (3.3.1), some main characteristics of both models are directly visible (left side of figure 5). FB's observed RMSE remains relatively low regardless of the noise level, whereas KR's ability to fit the observed prices quickly deteriorates. Its global smoothness reward restricts large local changes in the curve estimation which ends up in mispricing the observed bonds. But, FB's performance starts to worsen with the increase of noise as well. This is expected as finding forward rates that price each bond exactly becomes impossible when there are duplicates with noisy prices. Recall that in presence of duplicates, the forward rate that minimizes the pricing error is chosen (definition 1). This error becomes non-zero if observed prices diverge from the true prices. However, when looking at the true

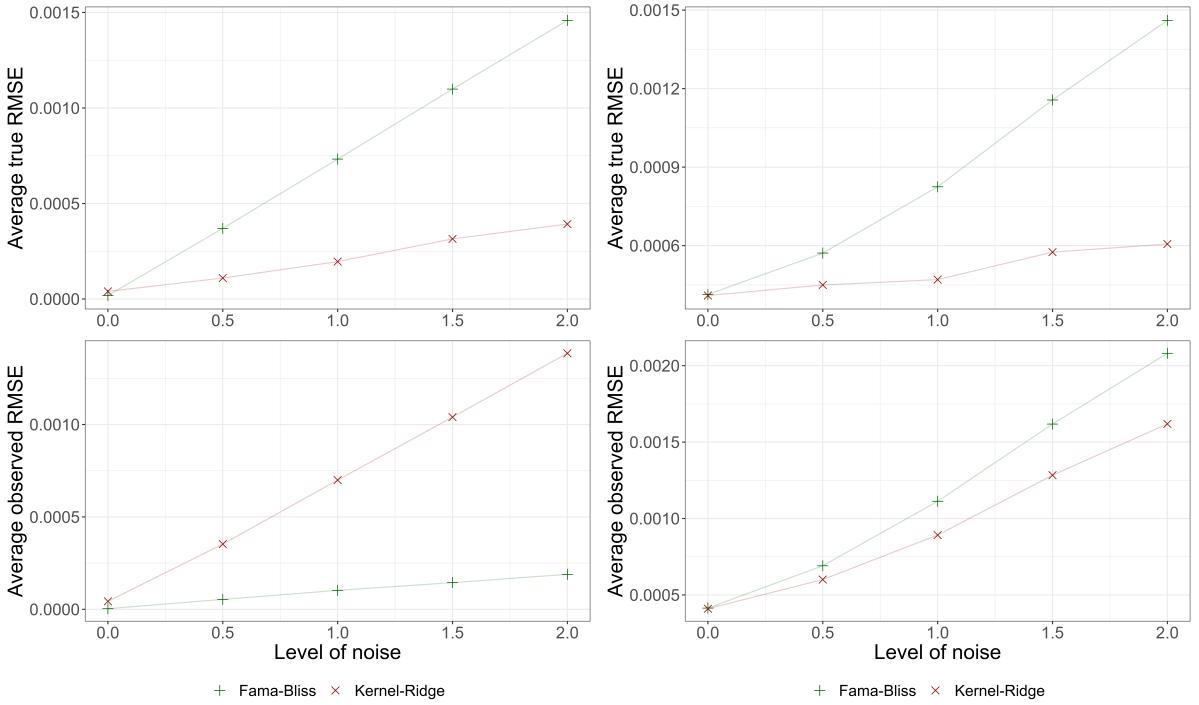


Figure 5. Average RMSE for the in-sample (left) and out-of-sample (right) evaluation with the base setup (3.3.1).

RMSE the downfalls of fitting the observations as precise as FB does become apparent. Here, KR beats FB in finding the true structure which is consistent with the example fits seen in figure 4. KR avoids overfitting individual bonds and remains relatively stable for noise increases as compared to FB.

The in-sample findings for the true RMSE translate to the performances in the out-of-sample evaluation (right side of figure 5). Although both estimators struggle at predicting the observed prices of the next day, KR is consistently better. It seems that because FB fitted the noise, it is biased with regards to the new day's prices with new, uncorrelated noise. In contrast, since KR does not fit individual bonds it does not produce predictions that are as biased which results in lower observed prediction errors. KR's stability is impressively seen in the small increase of the true RMSE when comparing the noise levels 0 and 2. Whereas FB's true RMSE almost triples, KR's increases by only one-third. Generally, KR seems to be much more robust to noise when fitting the underlying yield structure and as a result, in predicting the next day's true prices. KR's consistent superiority in the out-of-sample evaluation comes as no surprise considering the reasons for its in-sample performance. Since KR is not strongly affected by noise, the differences between its implied and true prices remain relatively stable on average. KR's global smoothness reward imposed by the norm makes the estimation of β robust to individual outliers.

As mentioned in section 3.1, the RMSE can be seen as an approximation of the YTM error (definition 7). Therefore, all results have an additional interpretation. Especially the true in- and out-of-sample results are interesting as they imply KR's superiority in determining the true YTM of a bond. This suggests that investment decisions determined by the YTM might be better

based on estimations made by KR. Such implications are coherent with the example estimations of the yield-to-maturities of a bond portfolio seen in figure 3 and the appendix (B.2).

4.2 Variations

The interpretation of the RMSE as a yield error could entice the hypothesis that the bonds' yields could be driving factors of the results. Meaning, that it is a determining factor at which point in the time frame which yield can be observed. Especially when estimators perform differently for different time-intervals. Therefore, a new shape that introduces different yield levels, as well as a shorter time frame that limits the maximum yield, could offer an answer to this hypothesis. However, neither the simulation with a humped yield curve (setup 3.3.2) nor with a shorter time frame (setup 3.3.4) reported significantly different results (see B.4 and B.6). Evidently, FB's struggle to fit the true yield curve is not driven by bonds with longer maturities or different yield levels. In contrast, KR is surprisingly consistent when considering its tuning parameters remained untouched. An explanation for this might be that despite the global smoothness, KR's kernel provides the ability to react to different shapes of the yield curve. Following up, the relative performance is still seen when turning off the 90-day filter (setup 3.3.3 and figure B.5). But, the overall errors are much higher which offers some validation for the use of the filter in general. Clearly, the bonds with extremely short maturities and hence, higher susceptibility to mispricing, are especially threatening for the estimators' performances. This simulation does not answer if the error increases stem only from a bad fit of the yield curve's short end or an overall worse fit when including bonds with extremely short maturities. Thus, whether or not the filter should be turned on, has to be examined in future simulations where, for example, errors are calculated for different time intervals. Finally, for less smooth yield curves (setup 3.3.5), the in-sample results show much larger differences between KR and FB for zero noise when compared to the other setting (figure B.7). This demonstrates how KR's smoothness imposes difficulties when trying to fit this setup's seven-day constant yield curve. Nevertheless, as soon as noise is inflicted or the fit is evaluated on the next day KR outperforms FB again. Meaning that KR's inability to fit the extremely unsmooth yield curve is outweighed by the mispricing due to noise. In the out-of-sample evaluation, overall error levels are slightly higher than in the base evaluation showing the struggle both estimators have with this unsmooth yield curve. KR's apparent improvement in finding the true next days prices for higher noise values seems odd. But seeing the general stability KR showed in this evaluation, it might simply be due to chance.

4.3 Sample size and decomposition

A general pattern emerged in the previous sections. FB heavily overfits individual bonds which leads to a good observed fit but, the true and next day fit suffer because of noise. This pattern is consistent when changing the sample size. The left side of figure 6 shows the in-sample results for an increasing number of bonds (setup 3.3.6). Evidently, as the number of bonds increases so does the probability of having bonds with the same maturity. As a consequence, the observed RMSE increases for both estimators. The true RMSE interestingly only decreases

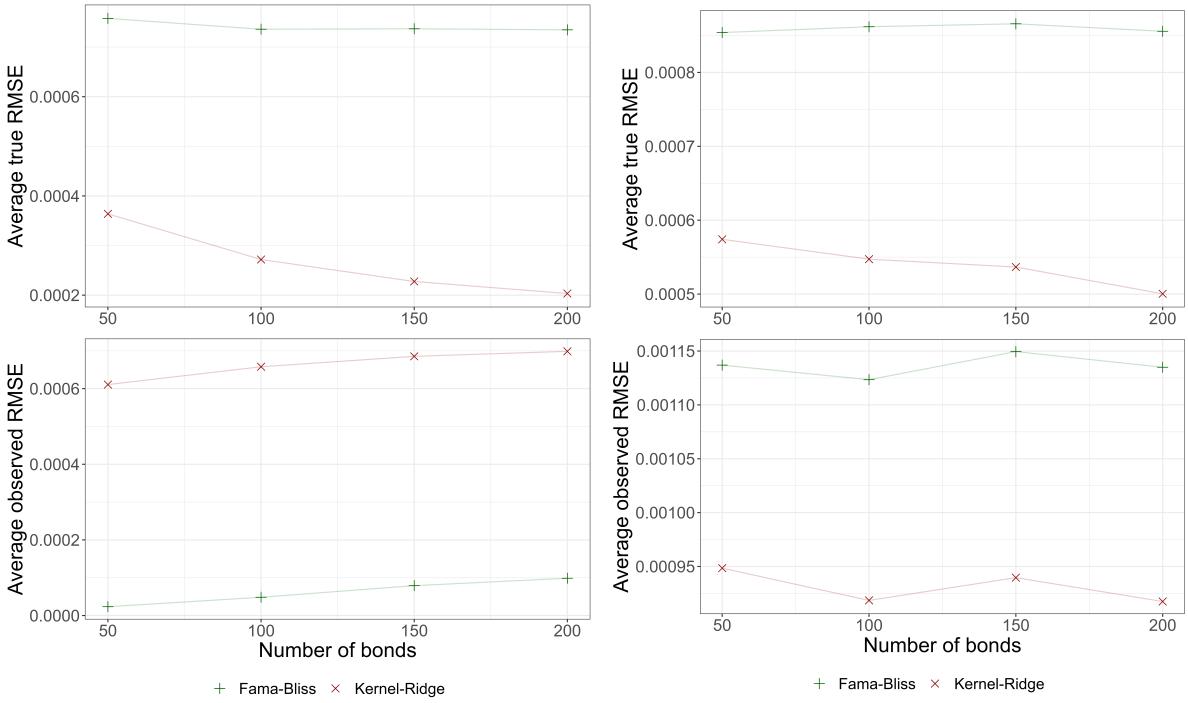


Figure 6. Average RMSE for the in-sample (left) and out-of-sample (right) evaluation with a varying number of bonds (setup 3.3.6).

for KR, meaning only it is able to use the increasing amount of information. In contrast, FB's noise-fitting seems to prevent any meaningful adoption of new information and as a result, the average true RMSE remains roughly constant. This means that with FB each bond's true price is on average equally mispriced regardless of the number of observations. Normally, an estimator is expected to improve for an increasing number of observations but FB's iterative and thus, local approach prohibits this. Recall that FB calculates the forward rate on intervals determined by the bond's maturities. While the increasing number of bonds means that those intervals get smaller, the overfitting that occurs remains the same. As the calculation of a forward rate at an iteration considers the previous calculations, FB has to compensate for previously made overfitting while also trying to fit the noise of the current bond that determines the current interval. Hence, FB's local and iterative approach restricts using more available information of the true yield curve as it is not considering the 'big picture'. This pattern persists in the out-of-sample evaluation where KR's estimations for the next day again uniformly dominate FB's overfitted estimations.

In order to better understand the driving factors of the estimators' performances one might look at a bias-variance decomposition (setup 3.3.7). Overall FB seems to have an unfavorable trade-off between bias and variance. The left side of figure 7 shows the variance part of each estimator. Evidently, as FB fits the noise, the variance of its estimations starts to rise whereas KR can almost resist the noise. KR's robustness is striking when looking at the bias part. Here, KR's bias is almost unchanged for different levels of noise. This robustness was also seen in Filipović et al. (2022) where KR was almost unaffected by artificial outliers. However, its robustness comes at the price of a non-zero bias even for zero noise, meaning KR estimations are consistently off. Contrary to KR, FB can excel for small levels of noise when looking at the bias part. Its ability

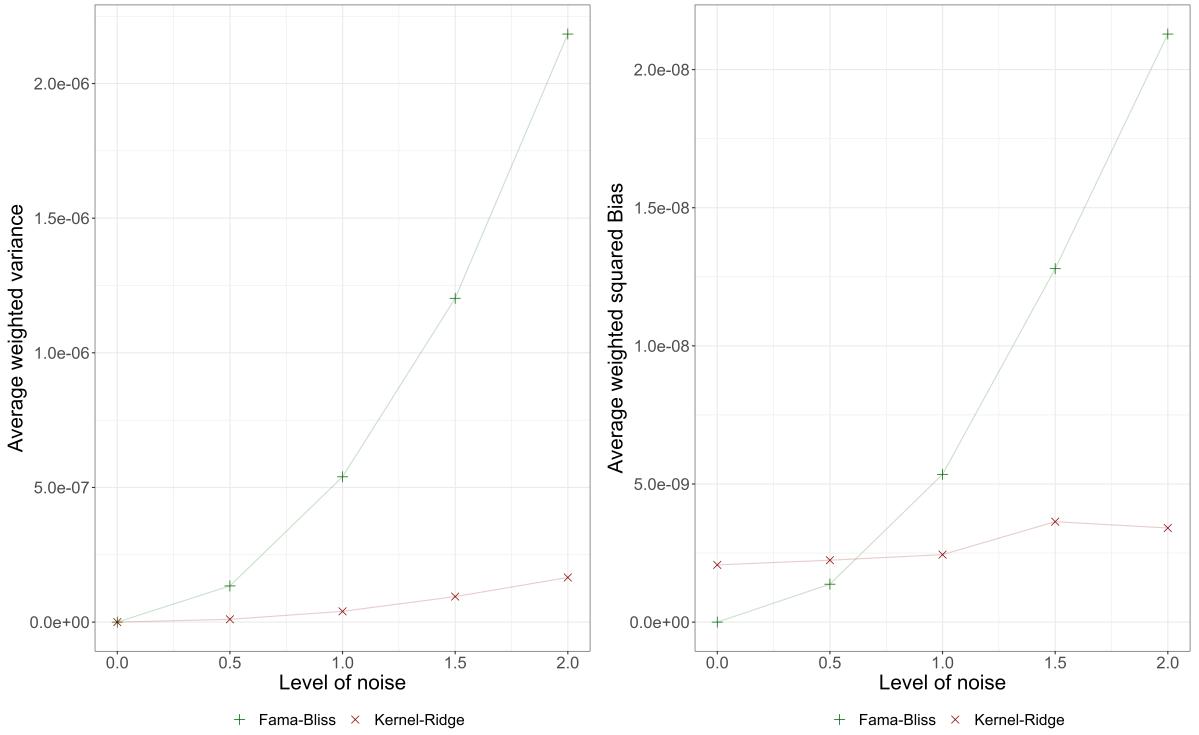


Figure 7. MSE decomposition: averaged and weighted variance (left) and squared bias (right) part following setup 3.3.7.

to price the observed bonds exactly gives it an edge when noise levels are low and observations do not diverge too much from the truth. As soon as noise levels reach 1 however, FB's overfitted estimations differ more from the truth than the robust but slightly biased estimations made by KR. Nevertheless, even for a noise value of 0.5, FB's combined performance that was measured by the true, in-sample RMSE is still worse than KR's (recall figure 5). Apparently, the variance of FB's estimations which is inherited from the noise outweighs predictions that are slightly closer to the truth.

4.4 Smooth FB

Unfortunately, KR seems to outperform FB in the wide array of scenarios that I have considered except when looking at the in-sample, observed RMSE. A natural explanation is FB's tendency to overfit individual bonds as illustrated in figure 4. Now, smooth FB is a direct attempt to avoid this overfitting while also offering a smooth method that is easier to implement than KR. Surprisingly, smooth FB seems at least equally capable compared to KR. While the in-sample results show no big gap between the estimators' abilities to fit the observed prices, smooth FB visibly beats KR in fitting the true prices for small and high noise values (left side of figure 8). This becomes more apparent when looking at the out-of-sample performances where smooth FB has an edge in finding the next day's true prices for larger noise values (right side of figure 8). Recall that smooth FB can be easily implemented by using large publicly available data sets consisting of estimations made by 'normal' FB and smoothing spline implementations in standard software. Therefore, my proposed smooth FB method might offer a more efficient way to implement a smooth and powerful yield curve estimator.

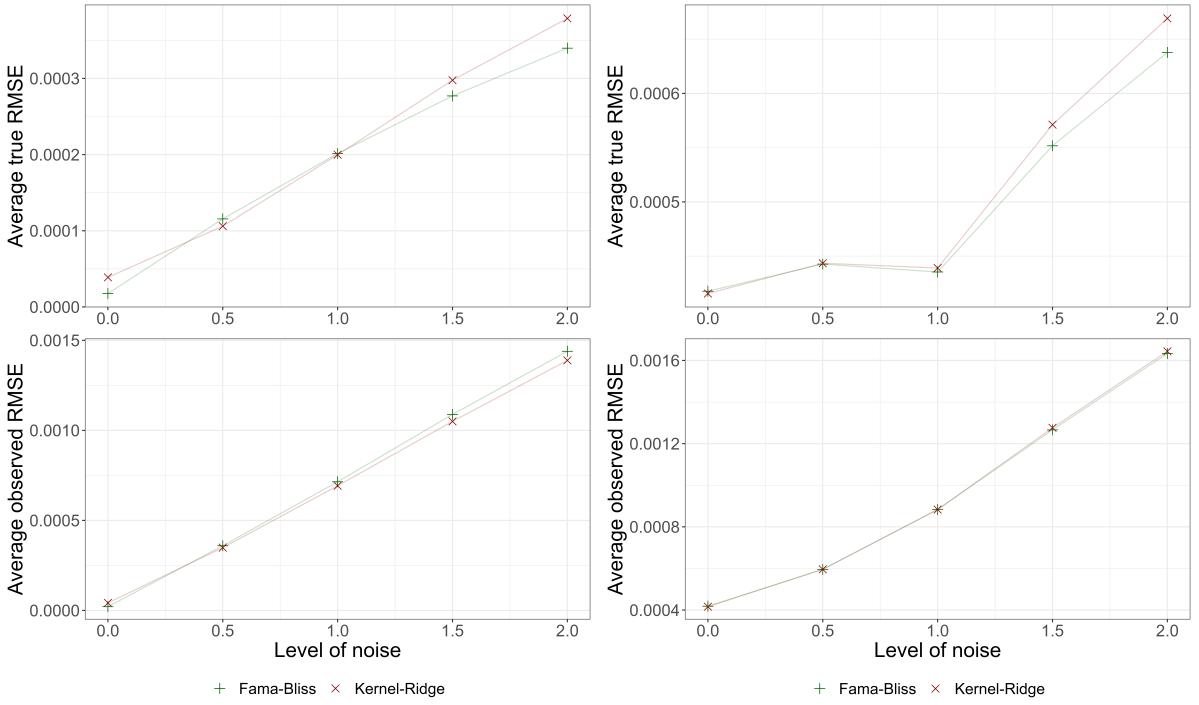


Figure 8. Average RMSE for the in-sample (left) and out-of-sample (right) evaluation with smooth FB (setup 3.3.8).

Nonetheless, it must be acknowledged that smooth FB got an unfair advantage by selecting the smoothness penalty with cross-validation. A simulation with a fine-tuned KR model would offer a more robust answer to the question of which method is truly superior. But, since cross-validation for the new KR method has to be self-implemented while cross-validation for smoothing splines is easily done with existing software, KR's implementation effort would rise even more.

5 Conclusion

5.1 Economic implications

Estimating the yield curve is of fundamental importance for the industry as well as research. Thus, the question of which method offers the best results is of concern for many different parties. But an answer to this question depends on the particular application. My simulations were focused on a pricing metric that is standard in the literature and can equally be interpreted as a yield error. In my simulations, both methods performed similarly well when markets were assumed to be perfect and as a result, no noise contaminated the bond's prices. If markets are assumed to function almost perfectly, then both methods seem to do the job. If this is not the case then the models' performances differed in some cases dramatically. The robustness of KR could be seen in every setup. Evidently, KR is much better than FB to find true prices and the general structure of my stochastic process. KR also clearly dominated in the out-of-sample evaluation. Contrary, FB can excel when the exact fitting of observed prices is the objective. The interpretation of the pricing error as the yield-to-maturity error could advise investing decisions

based on those yields. Applications that require extrapolation can not use 'normal' FB and have to rely on KR or smooth FB. My results can unfortunately not be compared to other studies except for KR's original paper due to KR's novelty. In Filipović et al. (2022), FB was equally susceptible to overfitting and KR dominated all other methods. In general, KR's global smoothness reward restricts the parameter estimation from overfitting individual bonds. This differs from other methods that make use of a kernel, like an estimator proposed by Liu and Wu (2021). Other studies like Jeffrey et al. (2006) or Bliss (1996) showed similar discrepancies between FB's in- and out-of-sample evaluation. Later also showed the robustness of a different version of smooth FB. I have shown that my version of smooth FB is a fast-to-implement estimator that is possibly at least as capable as KR. Here, future studies have to compare smooth FB and KR further to determine a clear superiority. Regardless of whether or not smooth FB is truly preferable, it is much faster to implement and can be seen as a 'quick and dirty' method to mimic KR's behavior and advantages.

5.2 Limitations and possible extensions

The limits of my simulations come primarily from the dgp as they are largely arbitrary. First, the yield curve sampling is not economically sound but just a stochastic process. A dgp that samples yield curves based on an interpretable economic model might be better suited for evaluating the yield curve estimators. Second, the noise inflicted on the true prices is assumed to be uncorrelated from one day to another and scaled by the bond's maturity. These assumptions are maybe not justified and so, other ways to inflict noise would make the simulations more elaborate. Third, while it seems plausible to have a maturity distribution that mimics that of current US treasuries, using different distributions could offer insights about the estimators' performances in the future where the distribution might change. The resemblance to the distribution of US treasuries also limits implications to the yield curve of US treasuries as other bond markets remained unconsidered. Although a pricing error encompasses a multitude of estimations for each bond, different evaluation metrics might unveil other results. Furthermore, questions regarding how well each estimator performs on different time intervals were largely unanswered. For example, readers that are only interested in estimating the yield curve for the first 12 months can not use my results to determine which estimator is preferred. Generally, the yield curve estimations were only used in the context of determining prices, other applications were left out. An investor that is interested in portfolio theory might propose simulations based on a portfolio consisting of various financial instruments that are directly or partially affected by the yield curve. The results coming from those simulations would shed a different light on the estimators. Lastly, I considered only two yield curve estimators and a version of smooth FB that got a head start in the evaluation. There are many other parametric and non-parametric methods proposed. Future simulation studies could implement those methods and come up with other settings like those, that I have outlined here and which were outside the scope of this thesis.

Appendix A Proofs and definitions

Definition 1 (Implied, unweighted, and squared pricing error). In order to solve for the forward rate in presence of S^i bonds with the same time-to-maturity at the i^{th} iteration of the FB algorithm (2.1.1) let the implied, unweighted, and squared pricing error be defined as

$$L(F^i) = \sum_{s=1}^{S^i} \left(P_{obs}^s - \sum_{j=1}^{m^s} b_j^s d(\tau_j^s) \right)^2. \quad (\text{A.1})$$

Here, I used the same notation as in the algorithm by Jeffrey et al. (2006), so $\{b_j^s\}_{j=1}^{m^s}$ are the payments of bond s occurring at $\{\tau_j^s\}_{j=1}^{m^s}$. Bond s is one of the S^i bonds with the same time-to-maturity. F^i influences the loss function by its relationship with $d(\tau)$ as already described in algorithm (2.1.1). The estimation of the forward rate for this interval is simply the minimizer of A.1.

Definition 2 (Duration). Let $\{c_j^i\}_{j=1}^{J^i}$ be the cash flows of bond i that are paid at $\{t_j^i\}_{j=1}^{J^i}$ where time is measured in years. Depending on the application, P^i and ytm^i can be either set to the true or the observed values resulting in the true or observed duration. The duration of bond i is given by

$$D^i = \frac{\sum_{j=1}^{J^i} (c_j^i \times t_j^i \times \exp(-t_j^i \times ytm^i))}{P^i}. \quad (\text{A.2})$$

Definition 3 (Log-normal density function). The density function of a log-normal distribution with mean μ and standard deviation σ is given by

$$f_{LN}(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right). \quad (\text{A.3})$$

Definition 4 (Kernel density estimator). A kernel density estimator for a sample of M iid. random variables $\{x_1, x_2, \dots, x_M\}$ is given by

$$\hat{f}_h(y) = \frac{1}{M\hat{h}} \sum_{i=1}^M K\left(\frac{y - x_i}{\hat{h}}\right). \quad (\text{A.4})$$

Where $K(u)$ is a kernel and \hat{h} an estimation of the bandwidth.

Definition 5 (Gaussian kernel). A Gaussian kernel is defined as

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right). \quad (\text{A.5})$$

Theorem 6. Given a sample $\{x_1, x_2, \dots, x_M\}$ and its kernel density estimator $\hat{f}_h(y)$ with Gaussian kernel $K(u)$. A random variable Z that samples from a normal distribution where μ is beforehand randomly set to one of the value of $\{x_1, x_2, \dots, x_M\}$ with $P(\mu = x_i) = \frac{1}{M}$ and $\sigma \equiv \hat{h} > 0$ has $\hat{f}_h(y)$ as its density function.

Proof. Let $f_Z(z)$ be the density function of the random variable Z as described in theorem 6 and $f(y | \mu, \sigma^2)$ be the normal density function with mean μ and standard deviation σ . Then

$$\begin{aligned} f_Z(z) &= \sum_{i=1}^M P(\mu = x_i) f(z | x_i, \hat{h}^2) \\ &= \frac{1}{M} \sum_{i=1}^M f(z | x_i, \hat{h}^2) \\ &= \frac{1}{M} \sum_{i=1}^M \frac{1}{\sqrt{2\pi}\hat{h}} \exp\left(-\frac{1}{2}\left(\frac{z - x_i}{\hat{h}}\right)^2\right) \\ &= \frac{1}{M\hat{h}} \sum_{i=1}^M \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{z - x_i}{\hat{h}}\right)^2\right) \\ &= \frac{1}{M\hat{h}} \sum_{i=1}^M K\left(\frac{z - x_i}{\hat{h}}\right) \\ &= \hat{f}_h(z). \end{aligned}$$

□

Definition 7 (Yield-to-maturity error). The yield-to-maturity error $RMSE^{YTM}$ is defined as

$$RMSE^{YTM} = \sqrt{\frac{1}{M} \sum_{i=1}^M (ytm^i - ytm_{imp}^i)^2}, \quad (\text{A.6})$$

where $\{ytm_{imp}^i\}_{i=1}^M$ are implied by the yield curve estimation. Depending on the application, one can calculate the true $RMSE_{true}^{YTM}$ and observed $RMSE_{obs}^{YTM}$ by either using $\{ytm_{true}^i\}_{i=1}^M$ or $\{ytm_{obs}^i\}_{i=1}^M$ as inputs for $\{ytm^i\}_{i=1}^M$ in a similar fashion to the RMSE in section 3.1.

Theorem 8 (MSE decomposition). The expected value of the MSE_{true} can be decomposed into a variance and bias part.

Proof.

$$\begin{aligned}
E[MSE_{true}] &= E \left[\sum_{i=1}^M \omega_{true}^i (P_{imp}^i - P_{true}^i)^2 \right] \\
&= E \left[\sum_{i=1}^M \omega_{true}^i \left(P_{imp}^i - E[P_{imp}^i] + E[P_{imp}^i] - P_{true}^i \right)^2 \right] \\
&= E \left[\sum_{i=1}^M \omega_{true}^i \left(\left(P_{imp}^i - E[P_{imp}^i] \right)^2 + 2 \left(P_{imp}^i - E[P_{imp}^i] \right) \left(E[P_{imp}^i] - P_{true}^i \right) + \left(E[P_{imp}^i] - P_{true}^i \right)^2 \right) \right] \\
&= E \left[\sum_{i=1}^M \omega_{true}^i \left(P_{imp}^i - E[P_{imp}^i] \right)^2 \right] + 2E \underbrace{\left[\sum_{i=1}^M \omega_{true}^i \left(P_{imp}^i - E[P_{imp}^i] \right) \left(E[P_{imp}^i] - P_{true}^i \right) \right]}_{const.} + E \underbrace{\left[\sum_{i=1}^M \omega_{true}^i \left(E[P_{imp}^i] - P_{true}^i \right)^2 \right]}_{const.} \\
&= \sum_{i=1}^M \omega_{true}^i E \left[\left(P_{imp}^i - E[P_{imp}^i] \right)^2 \right] + 2 \sum_{i=1}^M \omega_{true}^i \left(E[P_{imp}^i] - P_{true}^i \right) E \left[\left(P_{imp}^i - E[P_{imp}^i] \right) \right] + \sum_{i=1}^M \omega_{true}^i \left(E[P_{imp}^i] - P_{true}^i \right)^2 \\
&= \sum_{i=1}^M \omega_{true}^i \underbrace{E \left[\left(P_{imp}^i - E[P_{imp}^i] \right)^2 \right]}_{Var(P_{imp}^i)} + \sum_{i=1}^M \omega_{true}^i \underbrace{\left(E[P_{imp}^i] - P_{true}^i \right)^2}_{Bias(P_{imp}^i)^2} \\
&= \underbrace{\sum_{i=1}^M \omega_{true}^i Var(P_{imp}^i)}_{\text{Variance part}} + \underbrace{\sum_{i=1}^M \omega_{true}^i Bias(P_{imp}^i)^2}_{\text{Bias part}}
\end{aligned}$$

□

Appendix B Additional figures

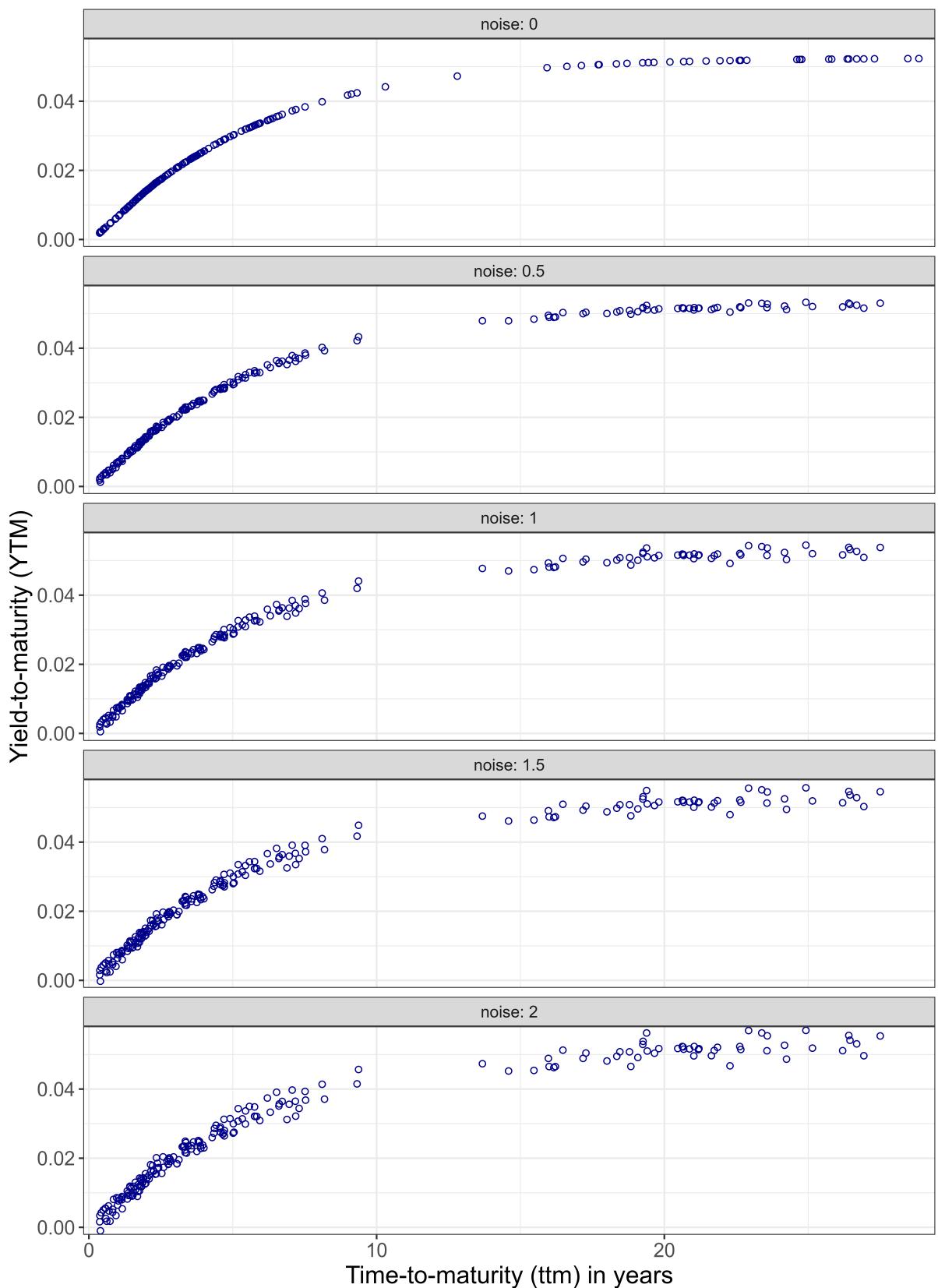


Figure B.1. Observed YTM of a portfolio of 200 zero-coupon bonds with different levels of noise.

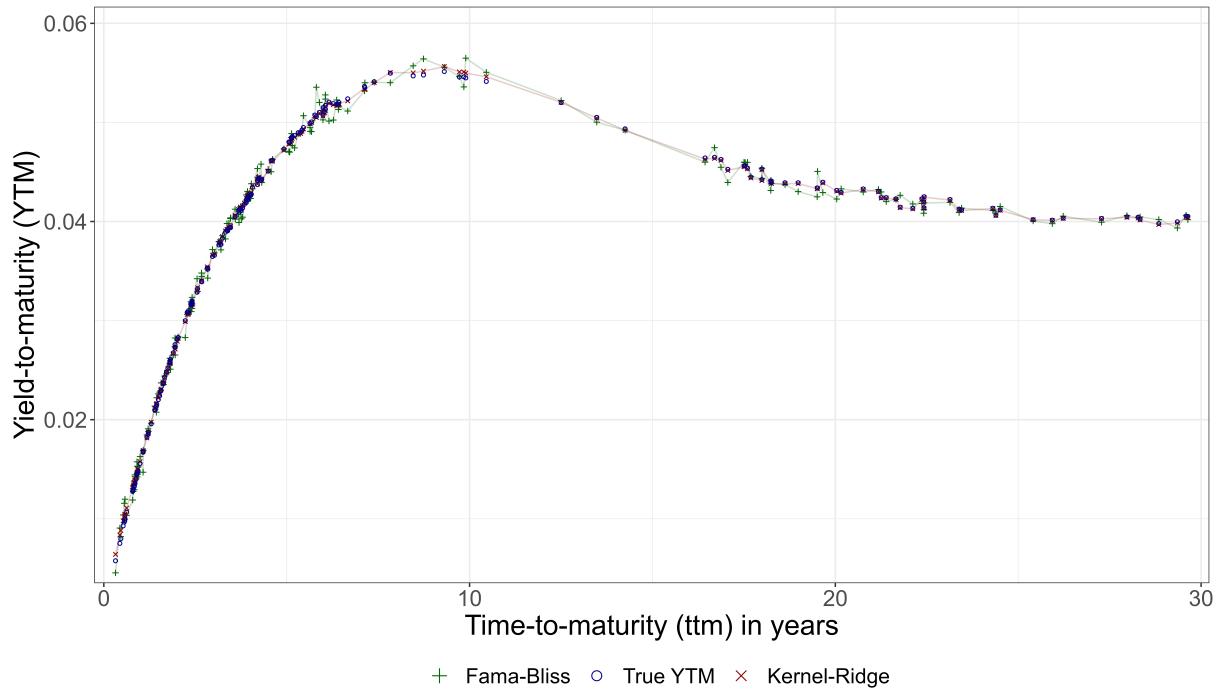


Figure B.2. True YTM and the estimations for a 200-bond portfolio with an underlying humped yield curve. The noise parameter was set to 1.

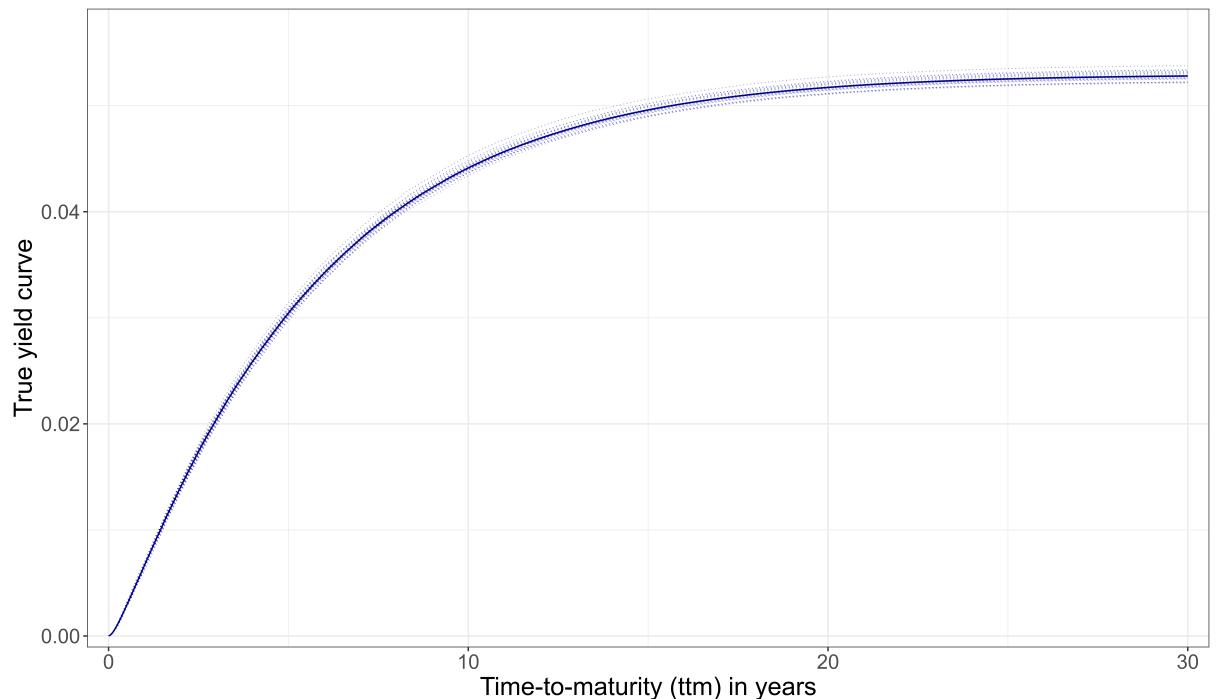


Figure B.3. A sample of 25 normal-shaped yield curves. The solid blue line is the average.

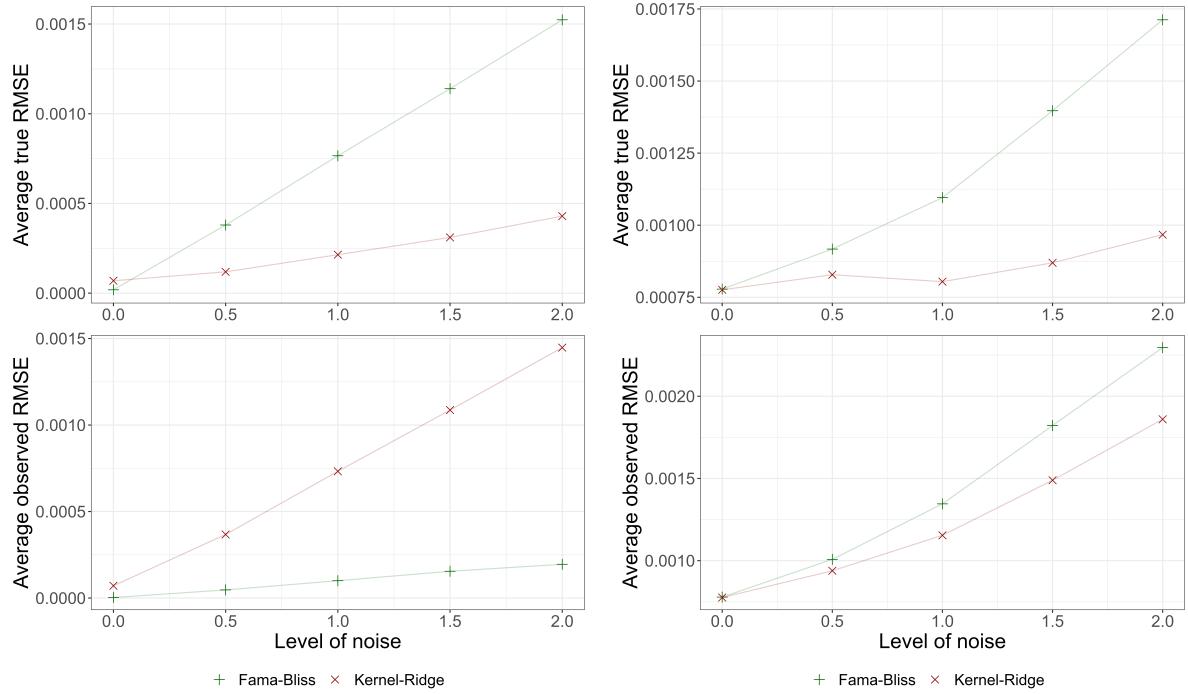


Figure B.4. Average RMSE for in-sample (left) and out-of-sample (right) evaluation with humped yield curves (setup 3.3.2).

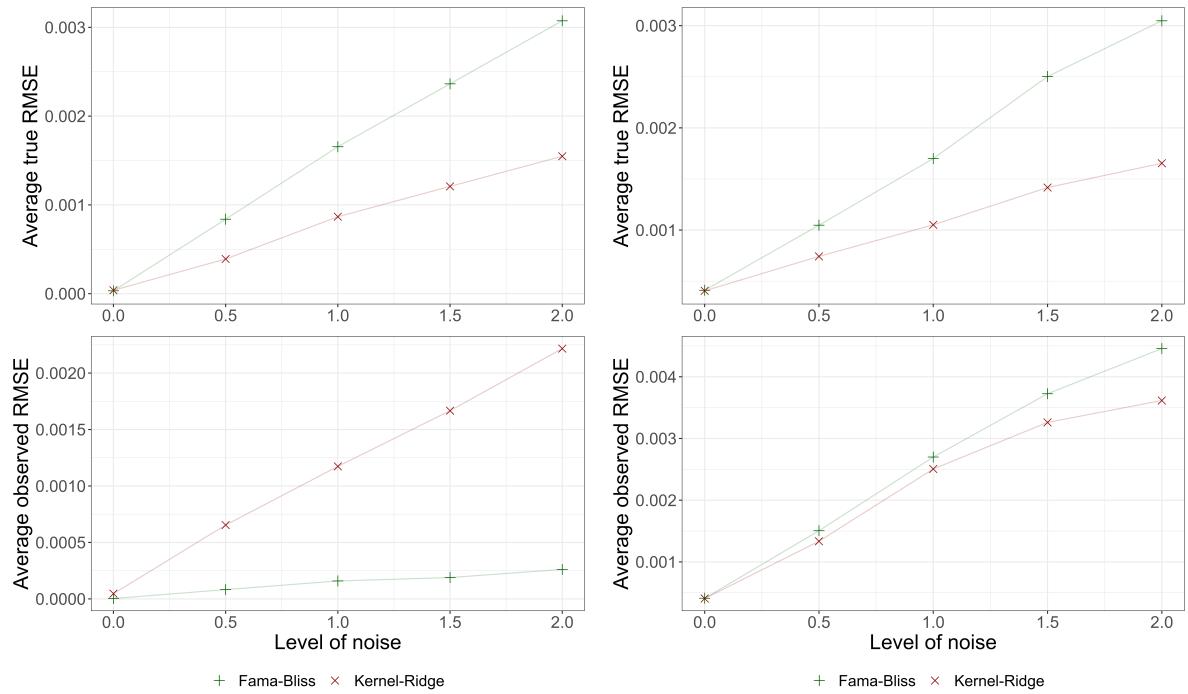


Figure B.5. Average RMSE for in-sample (left) and out-of-sample (right) evaluation without the 90-day filter (setup 3.3.3).

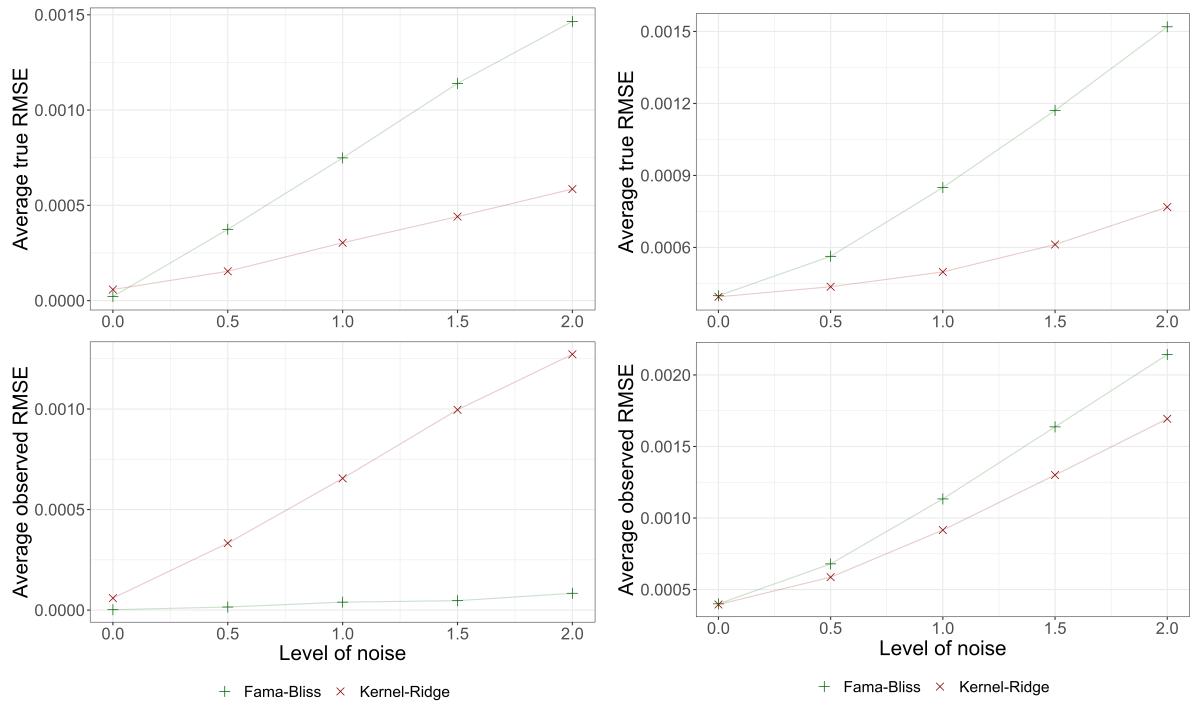


Figure B.6. Average RMSE for in-sample (left) and out-of-sample (right) evaluation for a 10-year time frame (setup 3.3.4).

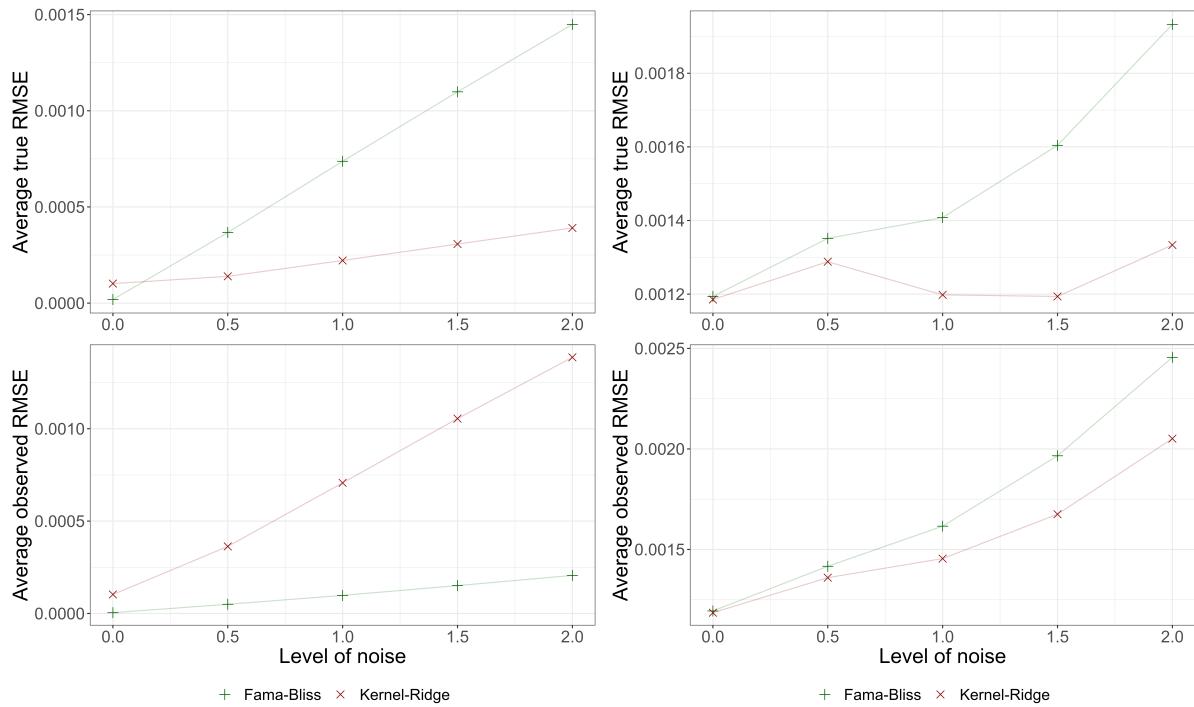


Figure B.7. Average RMSE for in-sample (left) and out-of-sample (right) evaluation for unsmooth yield curves (setup 3.3.5).

References

- Berk, Jonathan, and Peter DeMarzo.** 2019. *Corporate Finance Global Edition*. Chap. 3,6,12, 1184. Pearson Deutschland. <https://elibrary.pearson.de/book/99.150005/9781292304212>. [1, 2, 8]
- Bliss, Robert.** 1996. “Testing Term Structure Estimation Methods.” *Federal Reserve Bank of Atlanta, Working Paper* 9. [1, 3, 13, 20]
- Brealey, Richard A, Stewart C Myers, Franklin Allen, and Pitabas Mohanty.** 2012. *Principles of corporate finance*. Chap. 3. Tata McGraw-Hill Education. [4]
- Brennan, Michael J., and Eduardo S. Schwartz.** 1979. “A continuous time approach to the pricing of bonds.” *Journal of Banking and Finance* 3 (2): 133–55. [https://doi.org/https://doi.org/10.1016/0378-4266\(79\)90011-6](https://doi.org/https://doi.org/10.1016/0378-4266(79)90011-6). [10]
- Fama, Eugene F, and Robert R Bliss.** 1987. “The information in long-maturity forward rates.” *American Economic Review*, 680–92. [1–3, 12, 13]
- Fama, Eugene F, and Kenneth R French.** 2004. “The capital asset pricing model: Theory and evidence.” *Journal of economic perspectives* 18 (3): 25–46. [1]
- Filipović, Damir, Markus Pelger, and Ye Ye.** 2022. “Stripping the Discount Curve-a Robust Machine Learning Approach.” *Swiss Finance Institute Research Paper*, nos. 22-24. [1–5, 7–9, 17, 20]
- Fisher, Mark, Douglas W Nychka, and David Zervos.** 1995. “Fitting the term structure of interest rates with smoothing splines.” [14]
- Grolemund, Garrett, and Hadley Wickham.** 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>. [8]
- Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright.** 2007. “The U.S. Treasury yield curve: 1961 to the present.” *Journal of Monetary Economics* 54 (8): 2291–304. <https://doi.org/https://doi.org/10.1016/j.jmoneco.2007.06.029>. [7, 12]
- Hasselman, Berend.** 2022. *nleqslv: Solve Systems of Nonlinear Equations*. R package version 3.3.3, <https://CRAN.R-project.org/package=nleqslv>. [8]
- Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman.** 2009. *The elements of statistical learning: data mining, inference, and prediction*. Chap. 5, vol. 2. Springer. [13, 14]
- Hull, John C.** 2003. *Options futures and other derivatives*. Pearson Education India. [1]
- Jeffrey, Andrew, Oliver Linton, and Thong Nguyen.** 2006. “Flexible term structure estimation: which method is preferred?” *Metrika* 63 (1): 99–122. [1, 3, 20, 21]
- Liu, Yan, and Jing Cynthia Wu.** 2021. “Reconstructing the yield curve.” *Journal of Financial Economics* 142 (3): 1395–425. [1, 7, 9, 12, 20]
- Mohamed, Theron.** 2023. “‘Big Short’ investor Michael Burry predicts a US recession in 2023 and another inflation spike.” Markets Insider. Accessed February 19,

2023. https://markets.businessinsider.com/news/stocks/big-short-michael-burry-recession-inflation-economy-fed-rate-cuts-2023-1?utm_medium=ingest&utm_source=markets. [12]
- Nag, Ashok K., and Sudip K. Ghose.** 2000. “Yield Curve Analysis for Government Securities in India.” *Economic and Political Weekly* 35 (5): 339–48. Accessed January 15, 2023. <http://www.jstor.org/stable/4408880>. [12]
- Nelson, Charles R, and Andrew F Siegel.** 1987. “Parsimonious modeling of yield curves.” *Journal of business*, 473–89. [2]
- Posit team.** 2022. *RStudio: Integrated Development Environment for R*. Boston, MA: Posit Software, PBC. <http://www.posit.co/>. [8]
- R Core Team.** 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>. [8]
- Sheather, Simon J, and Michael C Jones.** 1991. “A reliable data-based bandwidth selection method for kernel density estimation.” *Journal of the Royal Statistical Society: Series B (Methodological)* 53 (3): 683–90. [7]
- Svensson, Lars E. O.** 1994. “Estimating and interpreting forward interest rates.” [2]
- The Wall Street Journal.** 2023. “U.S. Treasury Quotes.” Dow Jones and Company. Accessed January 3, 2023. <https://www.wsj.com/market-data/bonds/treasuries>. [7]
- U.S. Department of the Treasury.** 2023a. “About Treasury Marketable Securities.” The United States Government. Accessed January 4, 2023. <https://treasurydirect.gov/marketable-securities/>. [7]
- U.S. Department of the Treasury.** 2023b. “Timeline of U.S. Treasury Bonds.” TreasuryDirect. Accessed January 16, 2023. <https://www.treasurydirect.gov/research-center/timeline/bonds/>. [12]
- Vasicek, Oldrich.** 1977. “An equilibrium characterization of the term structure.” *Journal of Financial Economics* 5 (2): 177–88. [https://doi.org/https://doi.org/10.1016/0304-405X\(77\)90016-2](https://doi.org/https://doi.org/10.1016/0304-405X(77)90016-2). [10]
- Wahba, Grace, and Yuedong Wang.** 2019. “Representer theorem.” *Wiley StatsRef: Statistics Reference Online*, 1–11. [4]
- Wickham, Hadley.** 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>. [8]
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al.** 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>. [8]

Selbstständigkeitserklärung

Ich versichere hiermit, dass ich die vorstehende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass die vorgelegte Arbeit noch an keiner anderen Hochschule zur Prüfung vorgelegt wurde und dass sie weder ganz noch in Teilen bereits veröffentlicht wurde. Wörtliche Zitate und Stellen, die anderen Werken dem Sinn nach entnommen sind, habe ich in jedem einzelnen Fall kenntlich gemacht.

1. März 2023

Marvin Lob