Reconstructing the Yield Curve[*]

Yan Liu

Texas A&M University, College Station, TX 77843 USA

Jing Cynthia Wu

University of Notre Dame, Notre Dame, IN 46556 USA

National Bureau of Economic Research, Cambridge, MA 02138 USA

First version: November 18, 2018
Current version: June 29, 2019

Abstract

The constant-maturity zero-coupon Treasury yield curve is one of the most studied datasets. We construct a new dataset with a non-parametric method. Our curve is globally smooth while still capturing important local variation. We show our dataset preserves information in the raw data and has much smaller pricing errors than existing benchmarks. Our dataset is available online.[1] We also provide how much information is in the raw data to complement our dataset.

1 Introduction

The constant-maturity zero-coupon Treasury yield curve is one of the most studied datasets. Researchers use it to study the term structure of interest rates (e.g., Ang and Piazzesi (2003), Duffee (2002), Hamilton and Wu (2012), Diebold and Rudebusch (2013), Wu and Xia (2016)), run return forecasting regressions (e.g., Fama (1984) and Cochrane and Piazzesi (2005)), analyze monetary policy (e.g., Rudebusch (2002) and Bernanke and Reinhart (2004)), and price other assets and derivatives (e.g., Hull and White (1990) and Jarrow and Yildirim (2003)). We construct a new dataset that represents the information in the raw data, and make it available to researchers.

The most popular zero-coupon Treasury yield curve datasets are Fama and Bliss (1987) and Gürkaynak et al. (2007) (GSW hereafter). However, both of them have their own limits. Fama and Bliss (1987) only have maturities of 1, 2, $\cdots$, 5 years. For researchers who are interested in return forecasting regressions with holding periods in months or who need maturities longer than five years, the only option is Gürkaynak et al. (2007).

GSW use a parametric method to interpolate a smooth yield curve. However, to obtain a smooth yield curve that resembles the intermediate range, they discard all securities with less than three months to maturity and all Treasury bills. Therefore, by construction, the short end of their yield curve has large pricing errors, which the authors acknowledge in their original paper.

However, the problematic short end of the yield curve propagates to other maturities because some coupons of longer-term bonds are discounted with short-term discount rates. On the other hand, the long end of their yield curve (e.g., maturities longer than five years) are also subject to extrapolation due to the nature of their parametric smoothing method. In spite of these shortcomings of the GSW dataset, researchers still rely on its short and long end, because GSW is the only option with a wide range of maturities.

To address these issues, we construct a new zero-coupon yield curve dataset with a kernel-smoothing method. This non-parametric method allows us to generate a globally

smooth yield curve while still capturing important local variation. In contrast to parametric methods that have a fixed degree of freedom throughout maturities, our non-parametric method adaptively chooses the local degree of freedom based on the amount of information available at a given maturity. As a result, we do not need to discard securities at the short end or Treasury bills, which we find contain important information in disciplining the overall behavior of the yield curve. Consequently, our dataset represents information in the raw data, not only for the medium run, but more importantly, for the short run and the long run.

We propose a novel approach to selecting kernel weights, which characterize the local degrees of freedom for the kernel-smoothing method, and determine how much local versus global information we pool to estimate the yield curve at a given maturity. We choose weights adaptively, and they are maturity specific and data driven. Specifically, we define the weight around a maturity as a normal distribution with the standard deviation given by the bandwidth. The bandwidth is inversely related to the amount of local information. For a region with abundant observations, the bandwidth is smaller. Consequently, weights are more concentrated locally.

We present our zero-coupon yield curve with monthly maturities from 1 month to 360 months. As a complement, we also provide how much information is in the raw data by calculating the bandwidth for each maturity at each time period. In general, the short term is associated with the smallest bandwidth, implying ample observations. By contrast, the bandwidth at maturities longer than 10 years is often large, due to intermittent issuance. Albeit a popular choice in the literature, the 30-year yield sometimes pools information from bonds with maturities that are 10 years away, even for the post-1990 sample. We recommend researchers use the bandwidth as additional information to assess the quality of the zero-coupon yield curve and the availability of the raw data.

We construct our dataset by applying our method to the CRSP Treasuries Time Series from June 1961 to December 2018. We compare our dataset with GSW in terms of how they

fit the raw data. First, we find that for maturities less than one year, GSW generate large and sometimes extreme pricing errors (e.g., in the magnitude of 10% in annual percentage points). Our method is able to reduce the average pricing error by as much as 53%. GSW have large pricing errors in the short end because they ignore the raw data in that section, and subsequently extrapolate the short end from longer-term bonds.

Second, although our dataset performs similarly to GSW over the medium term (i.e., maturities between one year and five years), it significantly outperforms GSW for maturities longer than five years, with a reduction of 45% in the average pricing error. This result highlights the constraint the parametric model used in GSW faces in fitting the long end of the yield curve. Overall, our model consistently outperforms GSW, both across the entire maturity spectrum and across time.

Our paper is organized as follows. In Section 2, we describe the non-parametric kernel-smoothing method. In Section 3, we discuss bandwidth selection. In Section 4, we apply the method to the US Treasuries data and illustrate the performance of our method. We offer concluding remarks in the final section.

2 Kernel-Smoothing Method

The goal is to extract a zero-coupon yield curve $y(n)$ for any maturity $n \in \mathcal{N}$ from observed Treasury bills, notes, and bonds, many of which have coupon payments. For theory, this section uses the support $\mathcal{N} = (-\infty, +\infty)$. In our application, we make it $\mathcal{N} = \{1, 2, ..., 360\}$ months.[2]

Estimation of the yield curve amounts to minimizing a weighted average of the distance between the fitted price and the observed price across all available bonds. To obtain the entire yield curve, the number of yields of interest often exceeds the number of observations, which creates non-uniqueness. As a result, existing methods impose additional constraints

---

[2]For earlier years when relatively long-maturity bonds are not available, the support is $\mathcal{N} = \{1, 2, ..., n\}$, where $n < 360$ is the maturity limit that we will specify later on.

on the minimization problem to obtain a unique and smooth solution. For example, Nelson and Siegel (1987), Svensson (1994), and GSW assume a parametric functional form of the yield curve and extract it from the data.

As an alternative to parametric approaches, we rely on a non-parametric method. The main advantage of a non-parametric framework is that, in contrast to parametric methods, the yield curve does not need to assume the same functional form across all maturities. For example, the short end of the yield curve typically has more local patterns, whereas longer term yields are smoother. A typical parametric method would struggle in capturing both features and need to compromise, whereas non-parametric methods can be designed to adapt to both features.

Our framework builds on the work of Linton et al. (2001), who introduced the non-parametric kernel-smoothing approach in estimating the yield curve. In particular, Linton et al. (2001) provide a theoretical justification for non-parametric methods by establishing the asymptotic distribution of the yield curve estimates when the yield function is assumed to be locally linear.

Different from their paper, we focus on the empirical performance based on a finite sample of bonds. Specifically, our goal is to construct a smooth zero-coupon yield curve that best describes the raw data. We make the following methodological contributions. First, we propose a new method for bandwidth selection in Section 3 targeting the unique features of the Treasuries. Second, we provide yield estimates over a denser set of maturities compared to the literature, namely, $\mathcal{N} = \{1, 2, ..., 360\}$. Third, we derive analytical derivatives for the first-order conditions of the objective function to facilitate computation. Fourth, our objective function is weighted by durations of bonds, which follows the literature on fitting the yield curve parametrically and is new to the non-parametric literature.

2.1 Pricing Error for a Security

At a given point in time, suppose we focus on a generic bond.[3] It is characterized by its observed price p, its sequence of cash flows $\{c_j\}_{j=1}^{J}$ including its principal, and the corresponding maturities $\{\nu_j\}_{j=1}^{J}$. Note that we use $n \in \mathcal{N}$ to denote a generic maturity of the zero-coupon yield curve, and ν denotes known maturities of cash flows. Note, ν across all cash flows and bonds do not cover the entire support $\mathcal{N}$.

Given $y(\nu_j)$, the implied bond price $\widehat{p}$ is

$$\widehat{p} = \sum_{j=1}^{J} c_j \exp\left(-y(\nu_j)\nu_j\right). \tag{2.1}$$

The goal is to extract the entire zero-coupon yield curve $y(n)$ for all maturities $n \in \mathcal{N}$ from observed bond prices. However, ν across all cash flows does not cover the entire support $\mathcal{N}$. Therefore, we cannot obtain $y(n)$ by simply inverting (2.1).

We instead connect a given ν_j with an arbitrary $n_j \in \mathcal{N}$ by approximating $y(\nu_j)$ with $y(n_j)$ using a first-order Taylor expansion:

$$y(\nu_j) \approx y(n_j) + (\nu_j - n_j)y'(n_j), \tag{2.2}$$

where $y'(n_j)$ is the first derivative of the yield curve evaluated at n_j. Now, we can approximate the bond price in (2.1) using (2.2)

$$\widehat{p}(n_1, n_2, ...n_J) \approx \sum_{j=1}^{J} c_j \exp\left[-\left(y(n_j) + (\nu_j - n_j)y'(n_j)\right)\nu_j\right], \tag{2.3}$$

where each $y(\nu_j)$ for the cash flow ν_j is approximated by an arbitrary point on the zero coupon yield curve $y(n_j)$. Note, $n_1, n_2, ..., n_J$ could be different.

In general, n_j could be any maturity in $\mathcal{N}$. However, the closer n_j is from ν_j, the more

---

[3]For brevity, we omit indicators for both time and bond for now.

information the j-th coupon payment provides on $y(n_j)$. To capture this idea, we use a normal kernel-weighting function:

$$
\begin{aligned}
K(n_j, \nu_j) &= K_{h(\nu_j)}(n_j - \nu_j), \\
&= \frac{1}{\sqrt{2\pi h(\nu_j)^2}} \exp\left[-\frac{(n_j - \nu_j)^2}{2h(\nu_j)^2} \right],
\end{aligned}
\tag{2.4}
$$

where $h(\nu_j)$ is the bandwidth parameter or the standard deviation of the normal distribution. The weighting function has two features. First, given the bandwidth, the weight is higher when n_j is closer to ν_j. Second, the bandwidth $h(\nu_j)$ is a function of ν_j. This is essential for our application and allows us to pool information more locally around one maturity and more globally around another.

When $h(\nu_j)$ goes to zero, the cash flow c_j only provides information for $y(\nu_j)$, but does not provide any information for $y(n_j)$ when $n_j \neq \nu_j$. Therefore, a narrow bandwidth overweights information locally and tends to generate a non-smooth yield curve. On the other hand, when $h(\nu_j)$ approaches infinity, all maturities are weighted equally. Hence, a wide bandwidth pools information more globally, but may generate yield curves that are overly smooth and lack local variation. Details of how we select the bandwidth are in Section 3.

Given the kernel weights, the kernel-weighted squared pricing error is

$$
\mathcal{E} = \int \ldots \int (p - \widehat{p}(n_1, n_2, \ldots n_J))^2 \prod_{j=1}^{J} K(n_j, \nu_j) dn_j,
\tag{2.5}
$$

where $\widehat{p}(n_1, n_2, \ldots n_J)$ is defined in (2.3). Note we have $\int \ldots \int \prod_{j=1}^{J} K(n_j, \nu_j) dn_j = 1$, which makes it an appropriate weight function.[4]

---

[4]The other condition for being a weight function is that it is positive everywhere. This is trivially satisfied for the normal kernel function.

2.2 Summarizing Information across Bonds

We have thus far constructed the kernel-weighted squared pricing error for a generic bond. To combine information from all available bonds at a given point in time, we need to add up the squared pricing errors across bonds. Suppose I bonds are available. Let the kernel-weighted squared pricing error for bond i be $\mathcal{E}^i$ for $i = 1, \ldots, I$, where $\mathcal{E}^i$ is defined in (2.5).

The same discrepancy between the actual price and the fitted price has different implications for two bonds that have different maturity structures. For example, a \$1 pricing error is more pronounced for a short-term Treasury bill as opposed to a 10-year Treasury note. This difference can be captured by weighting $\mathcal{E}^i$ with $1/D_i^2$, where D_i is bond i' duration, defined as

$$D = \sum_{j=1}^{J} \frac{\nu_j c_j \exp(-\nu_j \bar{y})}{p},$$

and the yield to maturity (YTM) $\bar{y}$ is the constant discount rate that equates the present value of the bond's cash flows with its price:

$$p = \sum_{j=1}^{J} c_j \exp(-\nu_j \bar{y}). \tag{2.6}$$

The duration-weighted pricing error can be interpreted as the equally weighted error in the yield space. Therefore, our objective function is

$$\mathcal{S}(y(\cdot), y'(\cdot)) \;=\; \sum_{i=1}^{I} \frac{1}{D_i^2} \cdot \mathcal{E}^i, \tag{2.7}$$

where $y(\cdot)$ is the yield function and $y'(\cdot)$ is its first derivative. Our goal is to minimize this objective function to obtain $y(n)$ and $y'(n)$ for all $n \in \mathcal{N}$.

To our knowledge, we are the first to use this weighting scheme in a non-parametric framework. We consider using durations to weight bond prices to be important, because

doing so allows us to put more weight on fitting the shorter end of the yield curve, which affects coupon payments of bonds at all maturities. The same weighting scheme has been applied by several papers that estimate the yield curve parametrically (e.g., Nelson and Siegel (1987); GSW).

Minimizing the objective function (2.7) with respect to $y(\cdot)$ and $y'(\cdot)$ is a non-trivial optimization problem. The main issue is that the integral in (2.5) does not have a closed-form expression and needs to be approximated. Therefore, we need to choose a discrete support to facilitate computation. We choose $\mathcal{N} = \{1, 2, ..., 360\}$ months, which is denser than Jeffrey et al. (2006), for example. Our choice of a dense support in $\mathcal{N}$ requires the estimation of a large number of parameters. We derive analytical derivatives of the first-order conditions for (2.7) to alleviate the numerical burden, and provide efficient and accurate estimates of the yield curve; see Appendix A. Once we have estimates of the yield curve over this discrete support, our framework permits a kernel-weighted interpolation scheme to provide estimates for maturities that are not in the support. Our choice of a dense support ensures the estimated yield curve is smooth over the entire maturity range.[5] We detail the steps to minimize the objective function as well as the interpolation scheme in Appendix A.

2.3 Model-Implied Bond Price

The model-implied bond price is

$$\widehat{p} = \int \ldots \int \widehat{p}(n_1, n_2, ...n_J) \prod_{j=1}^{J} K(n_j, \nu_j) dn_j. \tag{2.8}$$

---

[5]Note that interpolation guarantees the estimated yield curve is always continuous. However, a less dense support may lead to kinks in the estimated yield curve, which make the yield curve less smooth.

Once we have the estimated $y(\cdot)$ and $y'(\cdot)$ over $\mathcal{N} = \{1, 2, ..., 360\}$, we approximate this object with

$$\widehat{p} = \sum_{j=1}^{J} c_j \left(\frac{\sum_{n=1}^{360} K(n, \nu_j) \exp \left[- \Big(y(n) + (\nu_j - n)y'(n) \Big) \nu_j \right]}{\sum_{n=1}^{360} K(n, \nu_j)} \right). \qquad (2.9)$$

In Section 4, we use $\widehat{p}$ as well as the associated yield to maturity for model comparison.

3 Bandwidth

One main methodological contribution of our paper is to propose a bandwidth selection method for the yield curve. The choice of bandwidth determines the smoothness of the estimated yield curve, which is crucial to generate a globally smooth yield curve while not missing important local variation. Section 3.1 proposes our adaptive bandwidth selection procedure for the yield curve, and section 3.2 leverages the notion of bandwidth to summarize information content in the raw data.

3.1 Adaptive Bandwidth Selection Procedure

We propose a data-driven approach for choosing bandwidths. We follow the basic idea of adaptive bandwidth selection proposed in the literature on non-parametric estimators (see, e.g., Park and Marron (1990), Fan and Gijbels (1995), and Ruppert et al. (1995)). Our specific choices are new to the literature, making us the first to apply adaptive bandwidth selection to the estimation of the yield curve.

For each ν that corresponds to a cash flow, we choose $h(\nu)$ such that N_0 bonds mature within the two-bandwidth interval around ν (i.e., $[\nu - 2h(\nu), \nu + 2h(\nu)]$). In our main analysis, we set N_0 at 10. For a maturity region with lots of observations, the bandwidth $h(\nu)$ is small, and vice versa. To price this cash flow at ν, the relevant region in the zero-coupon yield curve is $n \in [\nu - 2h(\nu), \nu + 2h(\nu)]$, which covers 95% of probability.

If observations were equally spaced in a region of the yield curve, our choice of the bandwidth is equivalent to any maturity n on the zero-coupon yield curve in this region falling within the two-bandwidth intervals around these N_0 observations, which allows us to have a sample of N_0 bonds to estimate the yield curve at maturity n. For this reason, we call N_0 the effective number of local observations. Note these N_0 bonds are not weighted equally in estimating the yield curve at maturity n, because our kernel function assigns higher weights to bonds whose maturities are closer to n.

In practice, observations are not equally spaced, and they are asymmetric around ν. For these reasons, we adapt our bandwidth selection procedure as follows. Let $N([\nu_a, \nu_b])$ denote the number of bonds whose maturities fall into the interval $[\nu_a, \nu_b]$. We first define the left-hand-side bandwidth at maturity ν (i.e., $h^l(\nu)$) as

$$h^l(\nu) = \frac{1}{2} \min b$$
$$s.t. \quad N([\nu - b, \nu)) \geq N_0/2. \tag{3.1}$$

If no value of b satisfies $N([\nu - b, \nu)) \geq N_0/2$, we set $h^l(\nu)$ at $\nu/2$.

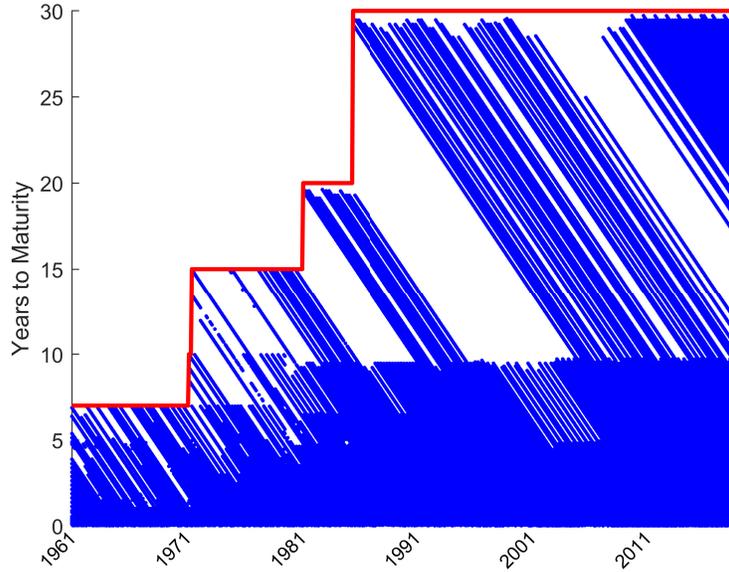Similarly, define the right-hand-side bandwidth at maturity ν (i.e., $h^r(\nu)$) as

$$h^r(\nu) = \frac{1}{2} \min b$$
$$s.t. \quad N((\nu, \nu + b]) \geq N_0/2, \tag{3.2}$$

with $h^r(\nu) = \frac{1}{2}(n_{max} - \nu)$ if $N((\nu, \nu + b]) \geq N_0/2$ cannot be satisfied for any b, where $n_{max} = 360$ months.

Because the normal kernel is symmetric around ν, we consolidate $h^l(\nu)$ and $h^r(\nu)$ into one bandwidth:

$$h(\nu) \quad = \quad \min\{\max\{3, h^l(\nu), h^r(\nu)\}, 60\}, \tag{3.3}$$

Figure 1: **Outstanding Treasury Securities**

Maturity distribution of outstanding securities, 1961–2018.

where three months is the minimum and 60 (i.e., five years) is the maximum bandwidth we set for any maturity. We next discuss the procedure.

Discussion   Calculating $h^l(\nu)$ and $h^r(\nu)$ separately guarantees that we take information from both the left side and the right side of ν. This is important because the maturity distribution of outstanding Treasury securities on a given day often contains gaps, leading to asymmetry between $h^l(\nu)$ and $h^r(\nu)$; see Figure 1. For example, suppose a 10-year gap is present in the maturity space: no bonds exist with maturities between $\nu_a = 120$ (i.e., 10 years) and $\nu_b = 240$ (i.e., 20 years). Also suppose a large number of bonds exist with maturities that fall just below $\nu_a = 120$, implying $h^l(\nu_a)$ is small (in particular, $h^l(\nu_a) \ll 60 = \frac{1}{2} \times (240 - 120)$). Now consider the bandwidth choice at ν_a. If we set the bandwidth $h(\nu_a)$ at $h^l(\nu_a)$, the bond price at maturity ν_a only provides information for the yield curve

up to maturity $\nu_a + 2h^l(\nu_a)$,[6] leaving the majority of the yield curve between ν_a and ν_b undetermined. Moreover, the implied curve between ν_a and ν_b is likely to be non-smooth. Our solution is to set the bandwidth at $h^r(\nu_a)$, which is the larger one between $h^l(\nu_a)$ and $h^r(\nu_a)$.

For shorter maturities, many observations contain potential micro structure noise and liquidity issues. For a fixed $N_0 = 10$, the bandwidth of $\max\{h^l(\nu), h^r(\nu)\}$ tends to be small. For example, $\max\{h^l(\nu), h^r(\nu)\}$ is on average around 0.5 months at the maturity of three months. Such a small bandwidth tends to generate substantial local variation in the estimated yield curve, which may not reflect the underlying true yield curve. Therefore, our choice of a minimum bandwidth of three months allows us to pool information from maturities that are within half a year of ν to smooth out the estimated yield curve.

On the other hand, too large a bandwidth may bias the yield curve estimate because the Taylor expansion in (2.2) can be inaccurate. We therefore set the maximum bandwidth at 60 months, allowing us to pull information within a 10-year radius. This maximum bandwidth only applies to long maturities where the data are sparse and have gaps in the maturity distribution.

Fixing the number of local observations at N_0 allows us to pool roughly the same amount of local information to estimate the yield curve at each maturity. Another benefit is that it automatically adjusts for the total number of Treasury securities available at each date. When more bonds exist (as in the later part of our sample), bandwidths in general shrink, which allows us to better capture the local variation in the yield curve.

3.2 Information Content in the Raw Data

In this section, we leverage the notion of bandwidth to summarize the information content in the raw data. Different from (3.3), which is the bandwidth for each cash flow ν, we are

---

[6]This is only approximately true, because the normal kernel assigns a non-zero weight to any maturity. However, it assigns relatively large weights to observations that are within two bandwidths.

now interested in the information contained at each maturity n on the zero-coupon yield curve. We propose to use

$$h(n) \;=\; \min\{h^l(n), h^r(n)\},$$

where $h^l(n)$ and $h^r(n)$ are calculated by (3.1) and (3.2). Note, if b does not exist for (3.1), we set $h^l(n)$ at ∞. The same applies to $h^r(n)$.

Why do we take the minimum instead of the maximum? We use the previous example with a 10-year gap in the maturity space between 10 years and 20 years to illustrate. For the bond with $\nu = 120$, $h^l(\nu) \ll h^r(\nu)$. But it needs to provide information to maturities within the gap $n \in (180, 240)$. This explains the maximum in (3.3). For $n = 120$ on the estimated zero-coupon yield curve, we still have $h^l(n) \ll h^r(n)$. However, the information we use to estimate the yield at $n = 120$ primarily comes from bonds on the left side, and $h^l(n)$ is small. Hence, we need to take the minimum instead.

Figure 2 provides two examples. The left panel is May 1990, and the right panel is May 2005. We plot bandwidths at the top. Yield to maturity is at the bottom and each dot corresponds to one outstanding security.

For both dates, the bandwidth increases with maturity in general, indicating observations are more concentrated on the short end. For May 1990, no outstanding securities have maturities between 180 and 296 months, which results in the spike in bandwidth. In May 2005, the longest maturity is 300 months, and we see the bandwidth increases sharply after that.

Figure 3 shows the time series of bandwidth for various maturities, with the red vertical bar indicating the beginning of 1990. Data on the short end are abundant, and the bandwidths for one, three, and six months are generally below 0.5 months.

The Treasury does not always issue notes and bonds with longer maturities. For example, it started issuing the 10-year notes in September 1971, 15-year bonds in December 1971, 20-

Figure 2: **Bandwidth on Selected Dates**
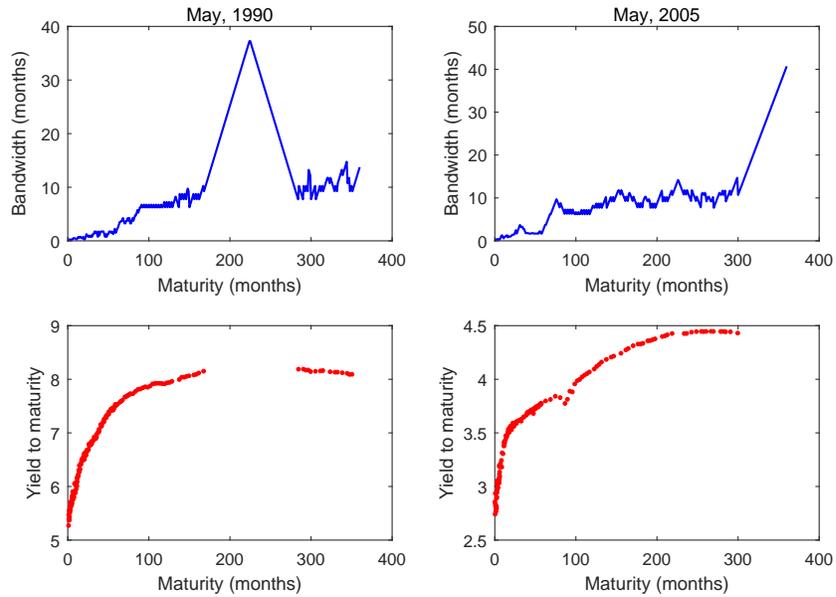
Bandwidth on two selected dates. We plot the cross section of bandwidth for May 1990 and May 2005. The top panels plot the bandwidth. The bottom panels plot the yield to maturity.

year bonds in July 1981, and 30-year bonds in November 1985. Even after these starting dates, they still issue them intermittently. This is consistent with Figure 1.

In general, the bandwidths become smaller after 1990 for maturities longer than one year. But they remain large for maturities longer than 10 years. The 30-year maturity is popular for the study of the behavior of long-term yields in the literature. However, due to the intermittent issuance, even the post-1990 sample's bandwidth can get as large as 60 months, implying a lack of observations and hence pulling information from maturities that are 10 years away.

Figure 3: **Time Series of Bandwidth**
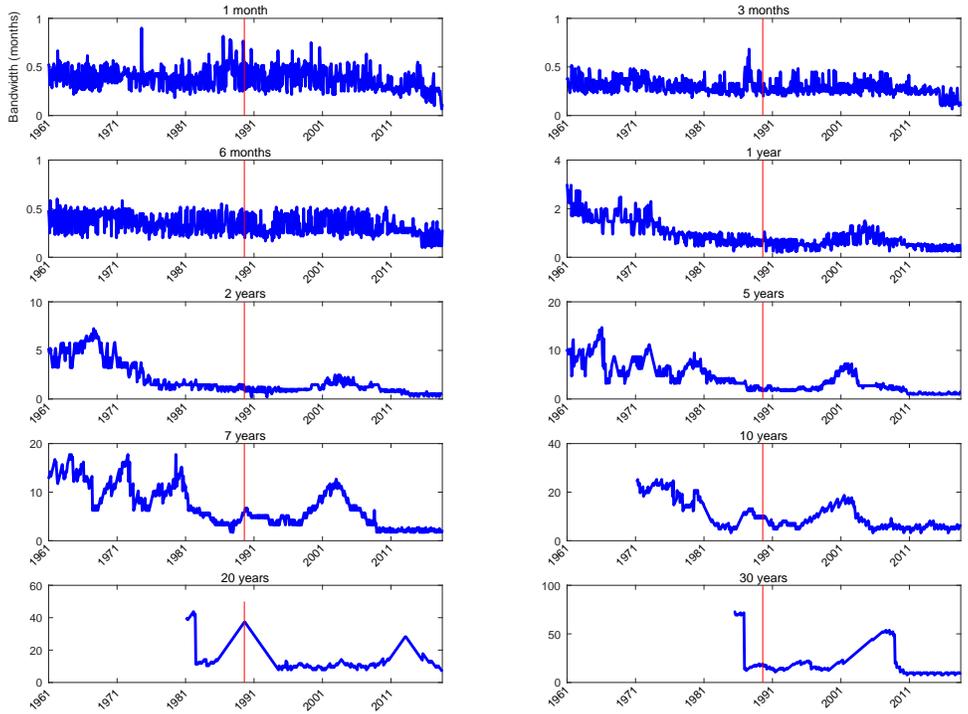
Time series of bandwidth for different maturities. We plot the time series of bandwidth over the entire sample for 1961 to 2018, and the red vertical bar marks the beginning of 1990.

4   The Performance of the New Yield Curve

Section 4.1 describes the data and the filtering procedures. Sections 4.2 - 4.4 illustrate the performance of our new yield curve. Section 4.2 focuses on selected dates to provide intuition. Sections 4.3 and 4.4 take a more systematic approach to evaluate the goodness of fit, with the former examining summary statistics and the latter assessing the time series of pricing errors.

4.1   Data

The raw CUSIP-level data come from the CRSP Treasuries Time Series. For each bond, we observe the end-of-day bid and ask (and average) prices, maturities, coupon payments, and schedules as well as other characteristics. The sample is from June 1961 to December 2018

in monthly frequency.

To construct the zero-coupon yield curve, we focus on a set of securities whose prices are only determined by their coupon payments and maturities, and not by other bond characteristics. To do so, we apply several filters to the raw CRSP Treasury data, similar to the literature. In particular,

1. *Only include fully taxable, non-callable, and non-flower bond issues (i.e., CRSP ITYPE equals 1, 2, 3, or 4).*

   This step ensures bonds with tax benefits and option-like features are not included in our sample. The same filter is applied by Fama and Bliss (1987).

2. *Exclude the two most recently issued securities with maturities of 2, 3, 4, 5, 7, 10, 20, and 30 years for securities issued in 1980 or later.*

   This procedure follows GSW and aims to delete on-the-run (or "first off-the-run") issues that often trade at a premium compared to other issues due to their liquidity and specialness.

3. *Apply an outlier-detection algorithm detailed in Appendix B.*

   Our approach is algorithmic in nature as opposed to the ad-hoc procedures in the literature.

Our procedures are similar to Fama and Bliss (1987) and GSW. The main difference between our filtering procedures and GSW is that we do not discard securities with shorter maturities or Treasury bills. We argued in Subsection 3.2 that these securities contain important information in the raw data for estimating the zero-coupon yield curve, and we subsequently show the benefit of keeping them subsequently.

Figure 1 summarizes the maturity structure for all outstanding Treasury securities over the entire sample period. The Treasury started issuing the 10-year notes in September 1971,

15-year bonds in December 1971, 20-year bonds in July 1981, and 30-year bonds in November 1985. We set the maximum maturity n_{max} accordingly, which is marked in red in Figure 1.
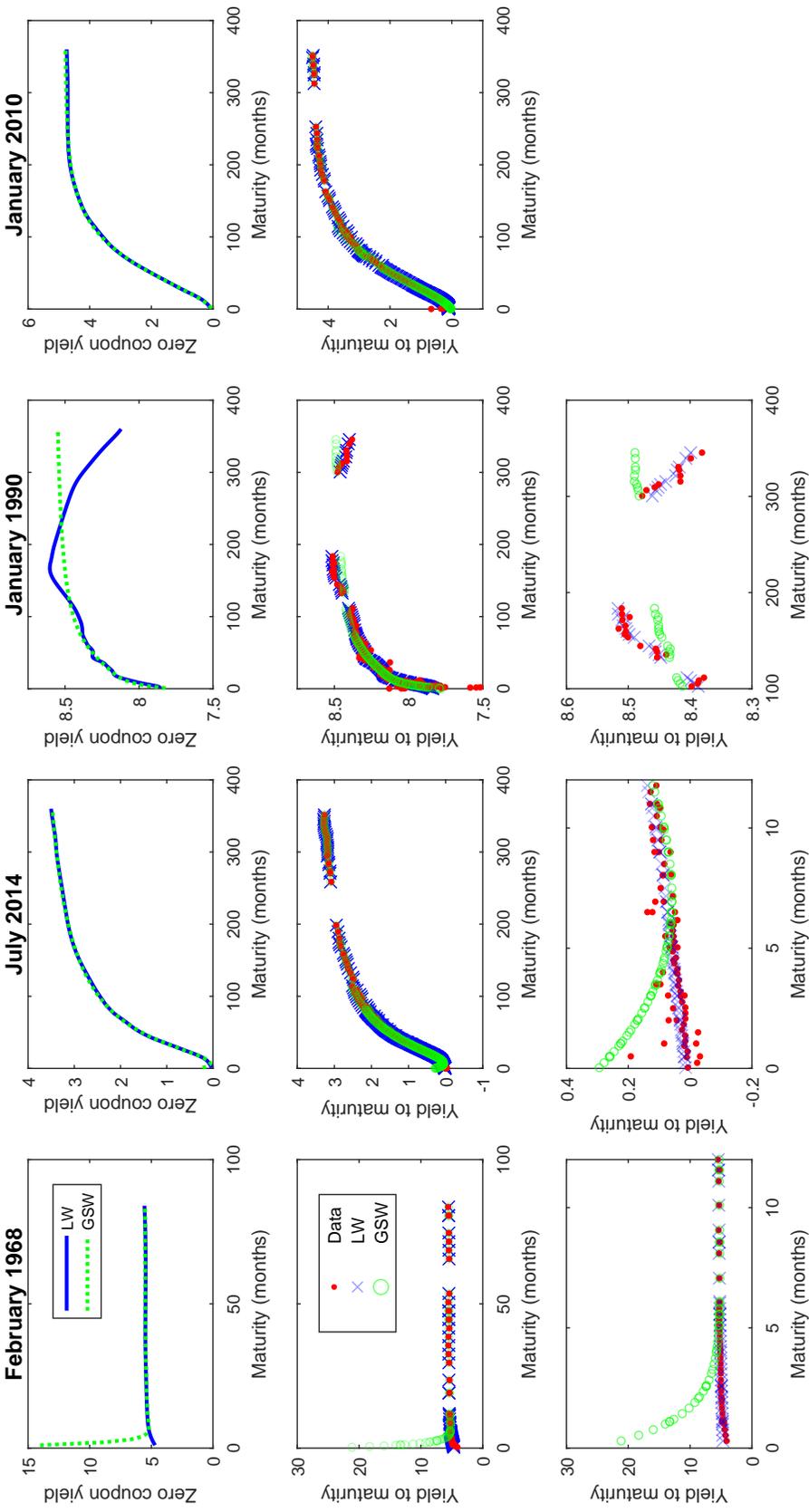
4.2   Yield Curves on Selected Dates

To gain some insights into the performance of our method, we use the yield to maturity (YTM) to compare our newly constructed yield curve with the raw data as well as the yield curve implied by GSW. For the raw data, the YTM is computed by (2.6). The model-implied YTM is the solution to (2.6), except the left-hand side is replaced with the model-implied price $\widehat{p}$, which is defined in (2.9) for our model.

Figure 4 illustrates the comparison for four dates: February 1968 (first column), July 2014 (second column), January 1990 (third column), and January 2010 (last column). The top row shows the zero-coupon yield curve, the next two rows are the YTM. Red indicates observations, blue uses our method, and green is GSW. We reestimate GSW's parameters based on our raw data; see details in Appendix C. The results are similar if we use the published parameters for GSW.

For February 1968, the main difference between our curve and GSW's is the short end. Whereas our zero-coupon yield approaches around 5% at a maturity of zero, it reaches a level of 15% for GSW's estimate. Importantly, as shown in panel (2,1) and more clearly in panel (3,1), the data do not support GSW's large estimate of the yield at the short end, leading to pricing errors that are in the magnitude of 20%.

The main difference between our model and GSW's in July 2014 is again the short end, which can be better seen in the (3,1) panel. This date is associated with the historical zero lower bound (ZLB) period for the United States, when the nominal interest rate is constrained by its lower bound at zero. Our curve captures the pattern in the raw data, and in consistent with the ZLB: The short end converges to zero when maturity approaches zero. By contrast, GSW has a U shape at the short end, and the difference in YTM between GSW and the raw data is about 0.3%.

Figure 4: **Yield Curves on Selected Dates**

Yield curve estimates on selected dates. We plot the yield curve estimates, as well as the implied yield to maturity (YTM)'s for for four particular dates: February 1968, July 2014, January 1990, and January 2010. For each date, the top panels plot the two zero-coupon yield curve estimates, one for our method ('LW') and one for GSW ('GSW'). The middle panels plot the YTM of the data ('Data') as well as the implied YTM for our method ('LW') and GSW ('GSW') across all maturities. The bottom panels zoom into a certain maturity range to highlight the difference in YTM.

The above examples on the short end are driven by two reasons. First, GSW drop observations in the short end, including all observations with less than three months to maturity and all Treasury bills. Second, the shape of their yield curve is dictated by their parametric model. The former may exacerbate the latter, because parametric models fitted only to securities on the relatively long end may generate unstable and poorly identified estimates on the short end that are inconsistent with the data.[7] By contrast, our non-parametric method allows us to include all the raw observations in our estimation, and can fit the short end without sacrificing the fit in other maturity segments.

For January 1990, the main difference is for longer maturities; see the (3,3) panel. The raw data contain a humped shape, whereas GSW's estimate is monotonically increasing in maturity. Therefore, GSW generate pricing errors that are systematically positive or negative: They are positive for maturities between 100 and 120 months and between 300 and 350 months, and negative for maturities between 120 and 200 months. Note that although the difference in YTM of 0.1% might seem small, it translates into a 0.5% difference in the zero-coupon yield. By contrast, our estimate fits the raw data well across all maturities.

What drives the performance of GSW is a large gap in the maturity distribution combined with limited observations in the long end, which accounts for a large fraction of our data based on Figure 1. This feature, combined with parametric methods with a limited degree of freedom, tends to underfit the yield curve around the gap, leading to systematic pricing errors for bonds on both sides of the gap. By contrast, our framework allows us to flexibly capture the local variation of the yield curve for this maturity range, resulting in a substantial reduction in pricing errors.

Finally, the last date (January 2010) presents a time when our yield curve estimate agrees with GSW's. In the meantime, the implied YTMs from both methods also closely match the data, showing that when a simple parametric function is sufficient to explain the variations in the raw data, our non-parametric method also produces a smooth yield curve and does

---

[7]See Section 5 of GSW for a related discussion on the instability of their estimates.

not overfit.

4.3 Summary Statistics

Section 4.2 illustrates that our method can capture various shapes of the yield curve and performs better than GSW on selected dates. We next systematically evaluate the performance of our dataset.

Let the actual bond price and the model-implied bond price be p_i and $\widehat{p}_i$ for $i = 1, 2, ..., I$. We first define two measures of pricing error that are directly related to our objective function. The first is the root-mean-squared pricing error (RMSPE) that calculates the square root of the mean-squared distance between p_i and $\widehat{p}_i$, that is, $\sqrt{\frac{1}{I} \sum_{i=1}^{I} (p_i - \widehat{p}_i)^2}$. The second is the duration-weighted root-mean-squared pricing error (WRMSPE) defined as $\sqrt{\frac{1}{I} \sum_{i=1}^{I} w_i^2 (p_i - \widehat{p}_i)^2}$, where $w_i = \frac{D_i^{-1}}{\sum_{i=1}^{I} D_i^{-1}}$ is the weight for bond i. Note WRMSPE is equivalent to our objective function that also weights pricing errors by bond durations.

We next define two metrics related to absolute pricing errors. They are the mean absolute pricing error, $\text{MAPE} = \frac{1}{I} \sum_{i=1}^{I} |p_i - \widehat{p}_i|$, and the duration-weighted mean absolute pricing error, $\text{WMAPE} = \sum_{i=1}^{I} w_i |p_i - \widehat{p}_i|$.

Bliss (1996) argues the bid-ask spread needs to be taken into account when calculating the pricing error. We follow Bliss (1996) to define the bid-ask-spread-adjusted pricing error as:

$$\varepsilon_i = \begin{cases} \widehat{p}_i - p_i^a & \text{if } \widehat{p}_i > p_i^a, \\ p_i^b - \widehat{p}_i & \text{if } \widehat{p}_i < p_i^b, \\ 0 & \text{otherwise,} \end{cases}$$

where p_i^a and p_i^b are the ask and bid quotes for the bond. The corresponding mean absolute pricing error (denoted as MAPE (Bliss)) and duration-weighted absolute pricing error (denoted as WMAPE (Bliss)) are defined as $\frac{1}{I} \sum_{i=1}^{I} \varepsilon_i$ and $\sum_{i=1}^{I} w_i \varepsilon_i$.

Next, rather than calculating the error between the actual and the fitted price, we define the mean absolute yield error (MAYE) as the average absolute error between the observed

and the fitted yield to maturity.

Lastly, we follow Bliss (1996) to define the hit rate (HR (Bliss)) as $\frac{1}{I}\sum_{i=1}^{I} \mathbb{1}_{\{p_i^b \leq \widehat{p}_i \leq p_i^a\}}$, where $\mathbb{1}_{\{p_i^b \leq \widehat{p}_i \leq p_i^a\}}$ is the indicator function that equals 1 if $\widehat{p}_i$ falls within $[p_i^b, p_i^a]$. The hit rate calculates the frequency of times the fitted price falls within the bid-ask spread.

To summarize, the eight metrics we consider are RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), WMAPE (Bliss), MAYE, and HR (Bliss). For the first seven, a smaller error indicates a better model, whereas a larger hit rate is associated with better performance.

Table 1 reports the performance comparison between our method and GSW for nine maturity buckets together with an overall comparison. Bold highlights the better performer. Panel A evaluates our method. In Panel B, we estimate GSW's curve with our raw data.[8]

Our method performs better than GSW across all metrics and maturity buckets. The improvement is substantial, and the reduction in pricing errors across all bonds (last column) ranges between 33% and 63%, with the largest reduction occurs in WMAPE (Bliss).

Across maturity buckets, our model performs significantly better at the short end and the longer end. For maturities less than three months, the percentage reduction in the pricing error of our model relative to GSW ranges from 65% to 84%, with WMAPE (Bliss) implying the largest reduction.

For maturities above five years, our model again presents a substantial improvement over GSW. The percentage reduction in pricing error ranges from 24% to 33% over the maturity range between 5 years and 10 years, from 45% to 54% between 10 years and 20 years, and 74% to 84% between 20 years and 30 years.

As a robustness check, we also report in Panel C using GSW's published parameters. Overall, Panels B and C have a similar performance, with Panel B being slightly better because it reoptimizes based on our raw data.

---

[8]Note that we follow GSW and drop all securities with maturities less than three months, as well as Treasury bills.

Table 1: **In-Sample Performance Summary**

	Maturity Bucket									All
	[0,3mth)	[3mth, 1yr)	[1yr, 2yr)	[2yr,5r)	[5yr, 7yr)	[7yr, 10yr)	[10yr, 15yr)	[15yr, 20yr)	[20yr, 30yr]	
Panel A: LW										
RMSPE	**0.014**	**0.039**	**0.078**	**0.151**	**0.264**	**0.489**	**0.451**	**0.197**	**0.117**	**0.185**
WRMSPE	**0.012**	**0.034**	**0.076**	**0.141**	**0.262**	**0.479**	**0.448**	**0.198**	**0.118**	**0.060**
MAPE	**0.011**	**0.030**	**0.060**	**0.111**	**0.213**	**0.402**	**0.391**	**0.162**	**0.092**	**0.092**
WMAPE	**0.009**	**0.026**	**0.059**	**0.103**	**0.211**	**0.392**	**0.387**	**0.163**	**0.092**	**0.023**
MAPE (Bliss)	**0.005**	**0.014**	**0.030**	**0.065**	**0.152**	**0.318**	**0.289**	**0.115**	**0.059**	**0.059**
WMAPE (Bliss)	**0.004**	**0.012**	**0.029**	**0.059**	**0.151**	**0.308**	**0.286**	**0.116**	**0.059**	**0.012**
MAYE	**0.086**	**0.057**	**0.044**	**0.037**	**0.044**	**0.064**	**0.049**	**0.013**	**0.006**	**0.050**
HR (Bliss)	**0.419**	**0.427**	**0.518**	**0.384**	**0.291**	**0.202**	**0.271**	**0.310**	**0.353**	**0.411**
Panel B: GSW, Re-estimated										
RMSPE	0.040	0.053	0.090	0.189	0.350	0.648	0.696	0.449	0.464	0.277
WRMSPE	0.035	0.049	0.088	0.177	0.347	0.634	0.688	0.451	0.461	0.094
MAPE	0.036	0.042	0.071	0.141	0.295	0.554	0.603	0.405	0.405	0.145
WMAPE	0.031	0.040	0.070	0.131	0.293	0.540	0.595	0.407	0.403	0.045
MAPE (Bliss)	0.029	0.026	0.039	0.092	0.230	0.463	0.495	0.350	0.367	0.109
WMAPE (Bliss)	0.025	0.025	0.038	0.084	0.228	0.449	0.487	0.352	0.364	0.032
MAYE	0.315	0.089	0.052	0.046	0.060	0.087	0.075	0.032	0.028	0.107
HR (Bliss)	0.236	0.349	0.407	0.290	0.174	0.114	0.135	0.104	0.128	0.284
Panel C: GSW, Original Model Parameterization										
RMSPE	0.056	0.068	0.103	0.209	0.347	0.715	0.747	0.458	0.513	0.305
WRMSPE	0.047	0.065	0.101	0.197	0.344	0.699	0.739	0.460	0.508	0.107
MAPE	0.052	0.056	0.083	0.156	0.282	0.609	0.646	0.411	0.451	0.160
WMAPE	0.042	0.055	0.081	0.147	0.279	0.592	0.637	0.412	0.448	0.057
MAPE (Bliss)	0.043	0.038	0.048	0.106	0.217	0.516	0.535	0.356	0.412	0.123
WMAPE (Bliss)	0.035	0.039	0.047	0.098	0.215	0.500	0.527	0.358	0.408	0.043
MAYE	0.425	0.123	0.061	0.052	0.058	0.096	0.081	0.033	0.031	0.138
HR (Bliss)	0.166	0.268	0.356	0.259	0.198	0.103	0.127	0.113	0.096	0.240

We report the summary statistics of pricing error for different models across different maturity buckets. We present results for three models: LW (our model), GSW re-estimated (GSW estimated based on our data), and GSW, original parameterization (using the original parameter values in GSW). For each maturity bucket (or across all bonds) and for each date, we calculate eight measures of pricing error: root-mean-squared pricing error (RMSPE), duration-weighted root-mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted absolute pricing error (WMAPE), mean absolute pricing error adjusted for bid-ask spread (MAPE (Bliss)), duration-weighted absolute pricing error adjusted for bid-ask spread (WMAPE (Bliss)), mean absolute yield error (MAYE), and the hit rate (HR (Bliss)). RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), and WMAPE (Bliss) are based on a face value of $100. MAYE is based on annualized percentage yield. We report the averaged pricing errors over the full sample from June 1961 to December 2018.

Figure 5: **Time Series of Mean Absolute Error in YTM: The Short End**
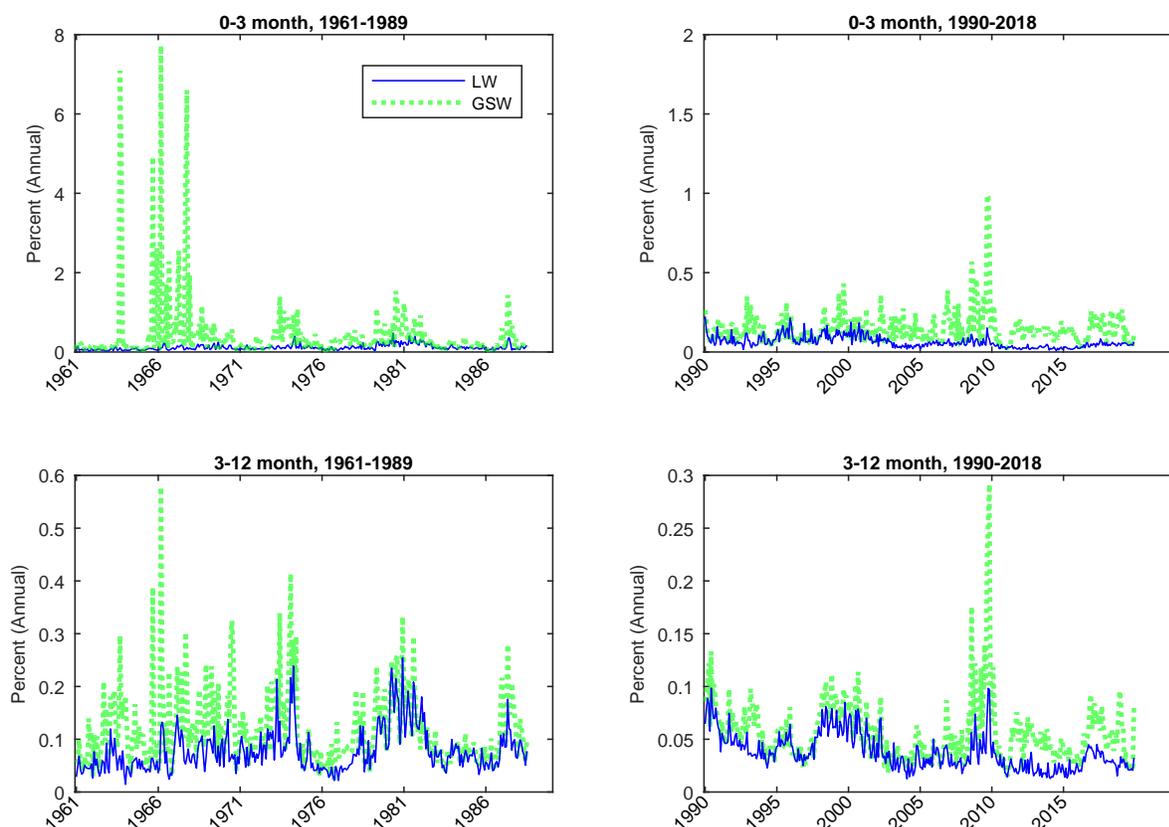
Time series of pricing errors for maturities less than one year. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW over the entire sample period (i.e., 1961–2018). The top panel examines bonds with maturities less than three months; the bottom panel examines bonds with maturities between three months and one year.

4.4   Time Series Evidence

This section examines the time series of pricing errors to provide more insights into the performance of our method.

The short term   We first examine the short end of the yield curve, that is, maturities that are less than one year. Figure 5 shows our model performs consistently better than GSW across different time periods and maturities. The top panel examines maturities less

than three months, and the bottom panel examines maturities between three months and one year. We split the full sample into the 1961–1989 (left panels) and the 1990–2018 (right panels) sub samples given the general decline in pricing error over time.[9]

For maturities less than three months (top panel), we observe that GSW occasionally generate large pricing errors of around 7%. The left panel of Figure 4 illustrates one such example in February 1968. Our method is able to reduce these pricing errors significantly.

For maturities between three months and one year (bottom panel), our model continues to outperform GSW. In particular, our model does better than GSW for the 1961-1975 period and the more recent post-2009 period. The post-2009 period is associated with the zero lower bound and subsequent low interest rate environment in the United States. As in Figure 4 (July 2014), we have illustrated that our method fits the short end of the yield curve better for this special period in history.

The large pricing errors of GSW at the short end come from the fact that they exclude all securities with less than three months to maturity as well as all Treasury bills. Consequently, GSW extrapolate the short end of the yield curve from securities with longer maturities, which leads to imprecise and sometimes extreme estimates of the short end of the yield curve.

Moreover, the issue of the short end of the yield curve of GSW is unlikely to be solved by simply including securities with short maturities in their estimation. The challenge is that the parametric model used in GSW has a limited degree of freedom and cannot simultaneously capture short-term, medium-term, and long-term yields.

By contrast, our non-parametric framework with adaptive bandwidth presents a natural solution to this challenge, because it adjusts the amount of local information to construct the yield curve.

---

[9]To ensure the reoptimization of GSW's model based on our data does not cause the large pricing errors on the short end for their model, we plot the minimum pricing error between their original parameterization and our re-estimated version in Figure 5.
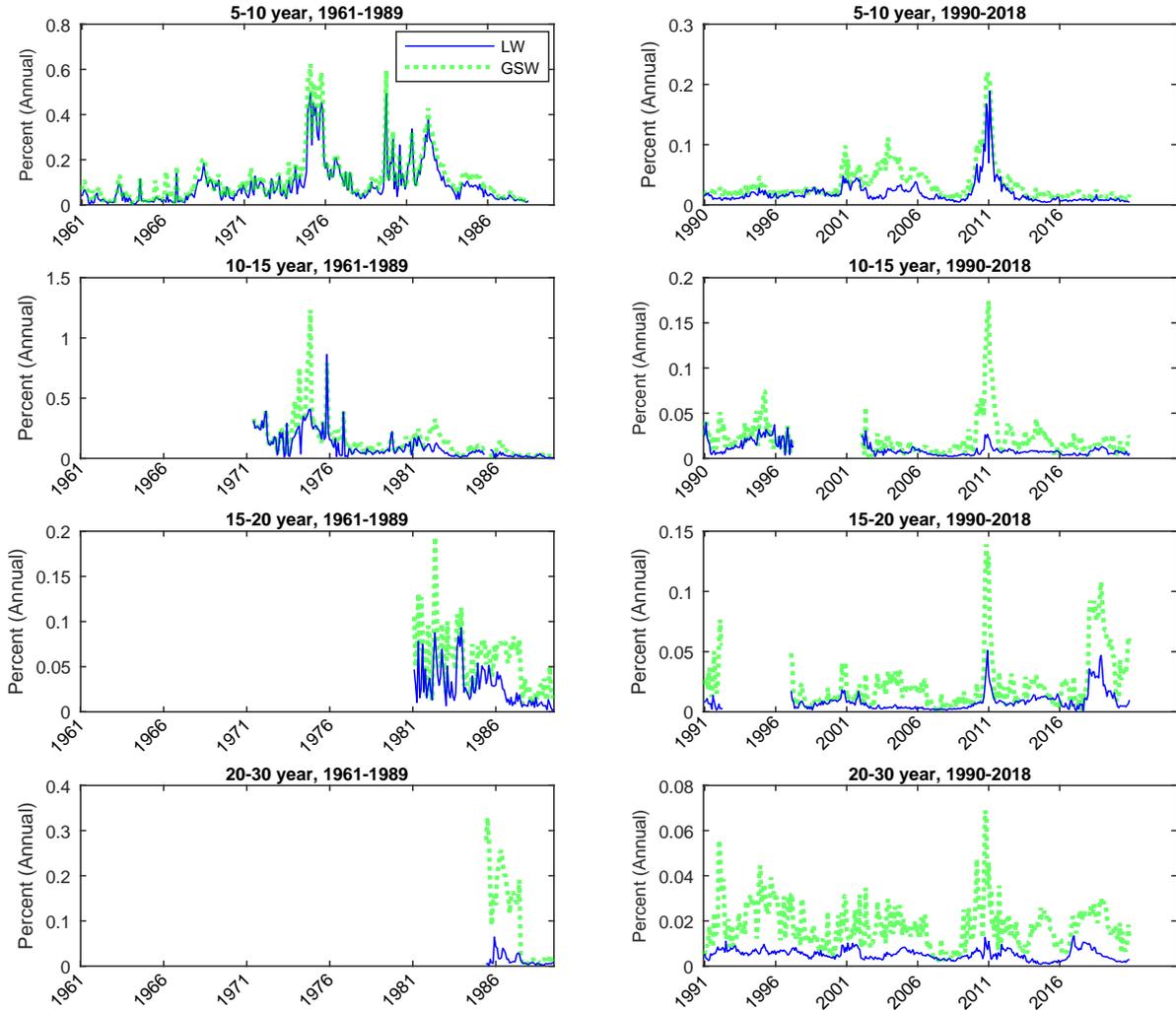
Figure 6: **Time Series of Mean Absolute Error in YTM: The Medium End**
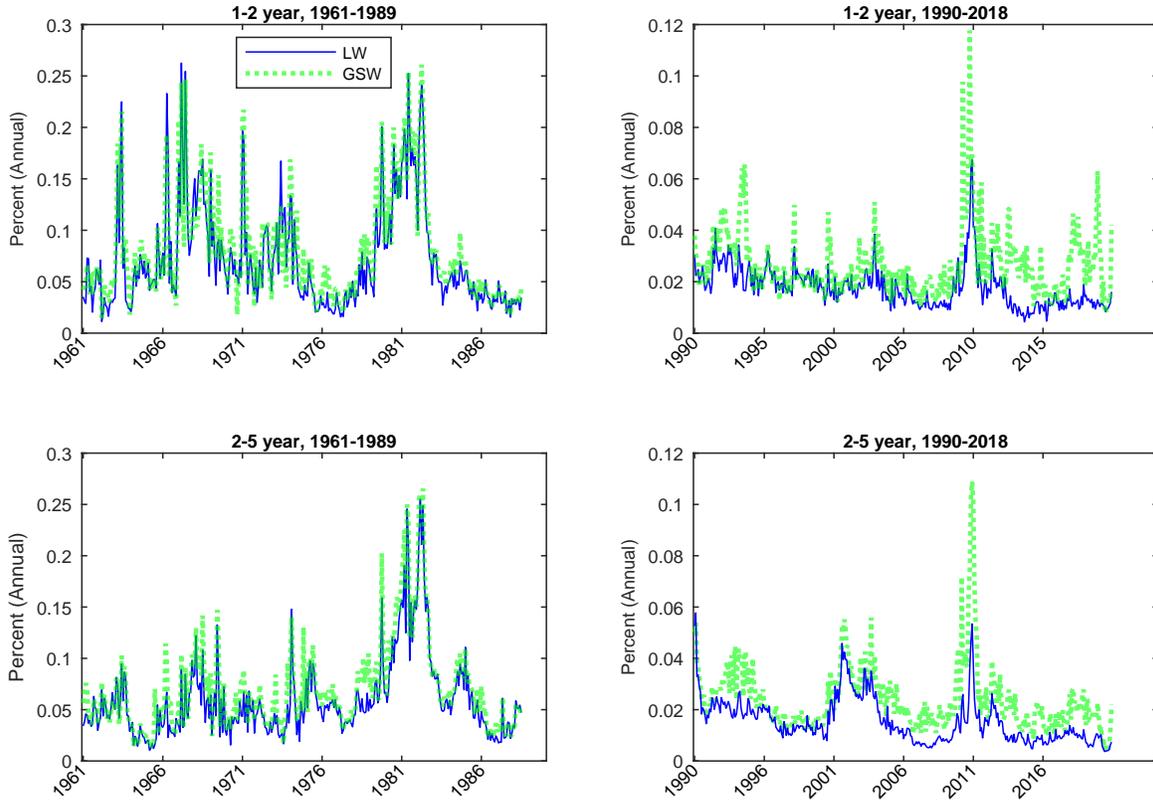
Time series of pricing errors for maturities between one year and five years. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW over 1961–1989 (left panels) and 1990–2018 (right panels). We group bonds into two maturity buckets: 1–2 year (top row) and 2–5 year (bottom row).

The Medium Term Figure 6 focuses on maturities between one and five years. Our model performs similarly to GSW, with the exception of the post-2009 sample. The zero lower bound lasts from 2009 to 2015, and interest rates remain low after that time. During this period, our model significantly outperforms GSW's. The first column of Figure 4 illustrates the intuition. This similarity in performance is consistent with the observation that abundant data are available over this maturity range, causing parametric models such as GSW's to use most of its degree of freedom to fit this part of the data.

Figure 7: **Time Series of Mean Absolute Error in YTM: The Long End**

Time series of pricing errors for maturities between five years and 30 years. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW over 1961–1989 (left panels) and 1990–2018 (right panels). We group bonds into four maturity buckets: 5–10 year (top row), 10–15 year (second row), 15–20 year (third row), and 20–30 year (bottom row).

The Long Term   For maturities above five years (Figure 7), we see substantial improvement of our model compared to GSW. Between five years and 10 years, we start to see the improvement of our method. For example, between 2000 and 2006, we are able to reduce the MAYE from 0.08% in GSW to around 0.02%.

Maturities longer than 10 years contain large spikes in pricing error generated by GSW.

For example, between 1986 and 1990, GSW's pricing error for maturities between 20 years and 30 years reaches 0.3%. By contrast, the pricing error from our method always stays under 0.05%. Moreover, our improvement applies not only to the pre-1990 sample for which a limited number of long-term securities are outstanding, but also to the post-1990 sample, including the most recent sample when abundant data on the long end are available.

5   Conclusion

The zero-coupon yield curve provides important information about financial markets and the macroeconomy. It is widely used by researchers and practitioners. Our paper develops a new dataset by a non-parametric kernel-smoothing method. Our dataset is available upon request. Our proposed non-parametric method allows us to generate a smoothed yield curve while preserving the information in the raw data. We show our yield curve estimate provides a more accurate description of the data than estimates in the existing literature.

References

Ang, Andrew and Monika Piazzesi, "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, 2003, *50*, 745–787.

Bernanke, Ben S and Vincent R Reinhart, "Conducting monetary policy at very low short-term interest rates," *American Economic Review*, 2004, *94* (2), 85–90.

Bliss, Robert R, "Testing term structure estimation methods," Technical Report, Working Paper, Federal Reserve Bank of Atlanta 1996.

Cochrane, John H and Monika Piazzesi, "Bond risk premia," *American Economic Review*, 2005, *95* (1), 138–160.

Diebold, Francis X. and Glenn D. Rudebusch, *Yield Curve Modeling and Forecasting.*, Princeton, NJ: Princeton University Press, 2013.

Duffee, Gregory R., "Term premia and interest rate forecasts in affine models," *The Journal of Finance*, 2002, *57* (1), 405–443.

Fama, Eugene and Robert R Bliss, "The Information in Long-Maturity Forward Rates," *American Economic Review*, 1987, *77* (4), 680–92.

Fama, Eugene F, "Term premiums in bond returns," *Journal of Financial economics*, 1984, *13* (4), 529–546.

Fan, Jianqing and Irene Gijbels, "Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, pp. 371–394.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright, "The U.S. Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics*, November 2007, *54* (8), 2291–2304.

Hamilton, James D. and Jing Cynthia Wu, "The effectiveness of alternative monetary policy tools in a zero lower bound environment," *Journal of Money, Credit, and Banking*, 2012, *44 (s1)*, 3–46.

Hull, John and Alan White, "Pricing interest-rate-derivative securities," *The Review of Financial Studies*, 1990, *3* (4), 573–592.

Jarrow, Robert and Yildiray Yildirim, "Pricing treasury inflation protected securities and related derivatives using an HJM model," *Journal of Financial and Quantitative Analysis*, 2003, *38* (2), 337–358.

Jeffrey, Andrew, Oliver Linton, and Thong Nguyen, "Flexible Term Structure Estimation: Which Method is Preferred?," *Metrika*, Feb 2006, *63* (1), 99–122.

Linton, Oliver, Enno Mammen, Jans Perch Nielsen, and Carsten Tanggaard, "Yield curve estimation by kernel smoothing methods," *Journal of Econometrics*, 2001, *105* (1), 185–223.

Nelson, Charles R. and Andrew Siegel, "Parsimonious modelling of yield curves," *Journal of Business*, 1987, *60* (4), 473–489.

Park, Byeong U and James S Marron, "Comparison of data-driven bandwidth selectors," *Journal of the American Statistical Association*, 1990, *85* (409), 66–72.

Rudebusch, Glenn D, "Term structure evidence on interest rate smoothing and monetary policy inertia," *Journal of monetary economics*, 2002, *49* (6), 1161–1187.

Ruppert, David, Simon J Sheather, and Matthew P Wand, "An effective bandwidth selector for local least squares regression," *Journal of the American Statistical Association*, 1995, *90* (432), 1257–1270.

Svensson, Lars E.O., "Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994," Working Paper 4871, National Bureau of Economic Research September 1994.

Tukey, John W, *Exploratory data analysis*, Vol. 2, Reading, Mass., 1977.

Wu, Jing Cynthia and Fan Dora Xia, "Measuring the macroeconomic impact of monetary policy at the zero lower bound.," *Journal of Money, Credit, and Banking*, 2016, *48* (2-3), 253–291.

A   Details on Estimation

The first-order conditions for the minimization problem in (2.7) are

$$\sum_{i=1}^{I}\sum_{j=1}^{J_i} \Phi_j^i(n; y, y') \cdot 1/D_i^2 \;=\; 0, \tag{A.1}$$

$$\sum_{i=1}^{I}\sum_{j=1}^{J_i} \Phi_j^i(n; y, y')(n - \nu_j^i) \cdot 1/D_i^2 \;=\; 0, \tag{A.2}$$

where $\Phi_j^i(n; y, y')$ is given by

$$\Phi_j^i(n; y, y') \;=\; \left(K_{h(\nu_j^i)}(n - \nu_j^i) c_j^i \nu_j^i d_j^i(n) \right)$$

$$\times \left(p^i - c_j^i d_j^i(n) - \sum_{\substack{k=1 \\ k \neq j}}^{J_i} \left(\int K_{h(\nu_k^i)}(n - \nu_k^i) c_k^i d_k^i(n) dn \right) \right), \tag{A.3}$$

$$d_k^i(n) \;=\; \exp\left[- \left(y(n) + (\nu_k^i - n)y'(n) \right) \nu_k^i \right]. \tag{A.4}$$

Note that equation (A.3) (and therefore equations (A.1) and (A.2)) contains integrals. Although, in principle, solving equations (A.1) and (A.2) numerically is possible,[10] we follow Jeffrey et al. (2006) and approximate the integrals with interpolations that are functions of $y(\cdot)$ and $y'(\cdot)$.[11]

In particular, suppose the support of $y(\cdot)$ and $y'(\cdot)$ is $\mathcal{N} = \{1, 2, ..., 360\}$. We approximate the integrals in equation (A.3) as

$$\int K_{h(\nu_k^i)}(n - \nu_k^i) d_k^i(n) dn$$

$$\approx \frac{\sum_{n=1}^{360} K_{h(\nu_k^i)}(n - \nu_k^i) \exp\left[- \left(y(n) + (\nu_k^i - n)y'(n) \right) \nu_k^i \right]}{\sum_{n=1}^{360} K_{h(\nu_k^i)}(n - \nu_k^i)}. \tag{A.5}$$

---

[10]See Linton et al. (2001) for the iterative algorithms they propose to solve system of equations that are similar to equations (A.1) and (A.2).

[11]We implemented both Linton et al. (2001) (in particular, the log-linear specification) and Jeffrey et al. (2006) for our model. Our experience is that Jeffrey et al. (2006) indeed offer a more stable and computationally efficient solution than Linton et al. (2001).

On the other hand, viewing $d_k^i(n)$ as the discount rate at ν_k^i approximated by the yield curve at n, $\int K_{h(\nu_k^i)}(n - \nu_k^i)d_k^i(n)dn$ can be interpreted as the kernel-smoothed discount rate at ν_k^i. Let the corresponding zero-coupon yield be $\widehat{y}(\nu_k^i)$, which is defined through $\int K_{h(\nu_k^i)}(n - \nu_k^i)d_k^i(n)dn = \exp[-\nu_k^i \times \widehat{y}(\nu_k^i)]$, we obtain $\widehat{y}(\nu_k^i)$ as[12]

$$\widehat{y}(\nu_k^i) = -\frac{1}{\nu_k^i} \log \left(\frac{\sum_{n=1}^{360} K_{h(\nu_k^i)}(n - \nu_k^i) \exp\left[- \left(y(n) + (\nu_k^i - n)y'(n) \right)\nu_k^i \right]}{\sum_{n=1}^{360} K_{h(\nu_k^i)}(n - \nu_k^i)} \right). \qquad \text{(A.6)}$$

Replacing ν_k^i by an arbitrary maturity ν in (A.6), we arrive at the formula that we use to interpolate the yield curve at any maturity ν.

In sum, we seek to solve equations (A.1) and (A.2) with respect to $y(n)$ and $y'(n)$ for $n \in \mathcal{N} = \{1, 2, ..., 360\}$, where $\Phi_j^i(n; y, y')$ is given by equations (A.3) and (A.4), but with the integrals in equation (A.3) replaced by equations (A.5) and (A.6). In essence, we are solving a system of non-linear equations. By construction, all of these equations involve functions that are infinitely differentiable. We provide closed-form gradients for these equations,[13] which allows us to solve these non-linear equations efficiently.

B   Outlier Detection

Our outlier-detection algorithm follows several steps.

First, we drop observations whose yield to maturity is higher than 40% (annualized). In the data, sometimes bond price appears to be too low (equivalently, yield to maturity

---

[12] The above interpolation can be interpreted as the solution to an optimization problem that is similar to (2.7) for a pure discount bond with a maturity of ν_k^i. More specifically, for a given estimated $\widetilde{y}(\cdot)$ and $\widetilde{y}'(\cdot)$, the solution to the minimization problem $\min_{y(\nu_k^i)} \int \left(\exp\left[- y(\nu_k^i) \times \nu_k^i \right] - \exp\left[- \left(\widetilde{y}(n) + (\nu_k^i - n)\widetilde{y}'(n) \right)n \right] \right)^2 K_{h(\nu_k^i)}(n - \nu_k^i)dn$ is given by $y(\nu_k^i) = -\frac{1}{\nu_k^i} \log \left(\int \left(\exp\left[-\left(\widetilde{y}(n) + (\nu_k^i - n)\widetilde{y}'(n) \right)n \right] \right) K_{h(\nu_k^i)}(n - \nu_k^i)dn \right)$. Because we only have solutions for $\widetilde{y}(\cdot)$ and $\widetilde{y}'(\cdot)$ over $\mathcal{N} = \{1, 2, ..., 360\}$, the interpolated version of this solution is given by equation (A.6).

[13] This is another benefit of replacing the integrals in equation (A.3) with interpolated yields as in equation (A.6).

appears to be too high). Across time, bond prices in general reach their lowest during the early 1980s recession, approaching a level of 20% in yield to maturity. We therefore set $40\% = 2 \times 20\%$ as a lenient threshold in yield to maturity to drop low-price observations. Note potential outliers that have a high yield to maturity but still below 40%, which are not dropped after this step, are likely to be dropped after the following steps.

After the initial step, we follow the $1.5\times$ IQR (interquartile range) rule in statistics to drop outliers (see, e.g., Tukey (1977)). For our application, however, one complication in applying this rule is that spatial variation exists in bond yields across the maturity spectrum. We therefore adapt this rule by taking into account the potential systematic variation in yields across maturities.

In particular, we first specify two bandwidth parameters ($\theta_1 = 10\%$ and $\theta_2 = 20\%$) that are defined to capture a certain fraction of the data (θ is different from N_0, which is the effective number of observations for our non-parametric model). For a given bandwidth θ and a bond observation, we apply the $1.5\times$ IQR rule to the $100\theta\%$ of observations that are the closest to the given bond in terms of maturity. We drop the observation if it fails the $1.5\times$ IQR rule for yield to maturity for both θ_1 and θ_2. Our use of both θ_2 (which is more global) and θ_1 (which is more local) ensures we do not drop observations that may appear to be outliers locally (i.e., when evaluated with θ_1) but stop being so globally (i.e., when evaluated with θ_2), and vice versa.

After establishing our outlier-detection algorithm, we examine selected dates to ensure our algorithm strikes a reasonable balance in dropping extreme observations and keeping information. On average, around 1.5% of observations are dropped at each date after applying our filtering procedure.

C   Details on Our Implementation of GSW

We obtain GSW's parameters from the Federal Reserve Board's webpage, and use them as starting values and re-estimate their model based on our raw bond data. Besides applying our filters described in Subsection 4.1, we also drop securities with less than three months to maturity and all Treasury bills, following GSW. In addition, we follow GSW by using the Nelson-Siegel four-parameter specification for the period before 1980 and GSW's six-parameter specification for the post 1980 period.

For most months, the re-estimated GSW curve is very similar to their original curve computed using their published parameters. This confirms the similarity in the underlying raw data we use. For a few months, the two versions have a substantial difference in the short end, where observations are omitted in estimation following GSW. This instability is consistent with what GSW find; see Section 5 of their paper. Given parameter instability of GSW, we compare our method with both the re-estimated GSW and their reported parameters.