Supplementary Appendix for
"A term structure framework for green bond spreads and portfolio strategies."

Contents

Appendix A. Feature definition

This appendix outlines the definitions of the green bond attributes obtained from the Bloomberg terminal. Green municipal bonds are classified using Bloomberg's indicator function, distinguished by a unique CUSIP number. This CUSIP is used to match the yield datasets with the attributes datasets.

Table A.1: Description of attributes

Type	**Attributes**	**Description**
Categorical	ID_CUSIP	Security identification number for the U.S. and Canada. The Committee on Uniform Security Identification Procedures (CUSIP) number consists of nine alphanumeric characters. The first six characters identify the issuer, the following two identify the issue, and the final character is a check digit.
	MUNI_TAX_PROV	Describes the United States (U.S.) federal and state income tax status of the bond. Will designate if a bond is subject to Federal Tax or Alternative Minimum Tax (AMT), or if the bond is Bank Qualified. Additionally, field may return whether the bond is subject to state income tax.
	CPN_TYP	Specifies how the periodic payment is structured. Under specific scenarios, payment structure details are replaced with information detailing the state of the instrument (Defaulted, Exchanged, Flat Trading, Funged, Prelim, When Issued).
	CRNCY	Currency in which the security was issued.
	ISSUER_BULK	The full name of the issuing entity of the security.
	SELF_REPTD_GREEN_INSTR_INDCTR	Indicates that the issuer has self-reported that the net proceeds of the fixed income instrument will be applied toward green projects or activities that promote climate change mitigation or adaptation, or other environmental sustainability purposes.
	BB_COMPOSITE	Blend of a security's Moody's, S&P, Fitch, and DBRS ratings.
	MUNI_LONG_INDUSTRY_TYP	Full name of industry sector bonds.
	CALLABLE	Indicates whether the security has a call provision.
	MARKET_ISSUE	Market in which the bond was issued.
	MUNI_PURPOSE	Describes how the proceeds of the bonds are being used.
	FINANCING_TYPE	Identifies whether the municipal bond's proceeds were used as money for new projects ('New Money'), refunding/refinancing of prior debt, or a combination of such types.
	State Code	State code associated with the issuer of the security.
Numerical	CPN_FREQ	Number of times per year interest is paid.
	CPN	Current interest rate of the security.
	ISSUE_DT	Date the security is issued.
	MATURITY	Date the principal of a security is due and payable.
	SPREAD_AT_ISSUANCE_TO_WORST	Spread for tax-exempt bonds is calculated from Bloomberg Valuation Service (BVAL) AAA (I493) curve for deals brought to market before 9/25/17; and from Bloomberg Valuation Service (BVAL) AAA Callable (BS1211) curve thereafter. Spread is calculated to appropriate interpolated point on the curve.
	DUR_ADJ_MID	Modified Duration for Non-Mortgages based on the Mid Price (PR003, PX_MID) of the security is returned.
	YIELD_ON_ISSUE_DATE	The yield the bond offered on the issue date. This field is not available for variable rate demand obligations (VRDOs).
	ISSUE_PX	Price of the security at issue.
	AMT_ISSUED	Cumulative amount issued from the original security pricing date through to the current date for debt securities.
	MUNI_ISSUE_SIZE	Returns the aggregate of the entire deal size (in USD), comprising all the series of bonds brought to market by the same lead underwriter in the deal to which the security belongs as detailed in the official statement.
	MTY-YEARS	The number of years until the principal amount becomes due.

Appendix A.1. UOP definition

BBG does not offer specific definitions for municipal UOPs. To ensure clarity and accuracy, we relied on multiple external references to compile comprehensive and precise definitions for each UOP category included in our dataset. We collected the definitions for UOPs in our dataset from a variety of reputable online sources.

Table A.2: Terms, Definitions, and Sources

Term	Definition	Source
CURRENT REFUNDING	Refunding of a bond within 90 days of the call date, typically done to take advantage of lower rates.	http://www.msrb.org/Glossary/Definition/REFUNDING.aspx
WATER UTILITY IMPROVEMENTS	Investments in infrastructure related to water supply, treatment, and distribution systems.	https://www.epa.gov/dwsrf/water-infrastructure-improvements-water-systems
ADVANCE REFUNDING	Issuing new bonds to replace old ones before the call date, using the proceeds to retire the original bonds.	http://www.msrb.org/Glossary/Definition/ADVANCE-REFUNDING.aspx
TRANSIT IMPROVEMENTS	Enhancements to public transportation infrastructure—buses, subways, and rail systems.	https://www.transit.dot.gov/grants
PUBLIC IMPROVEMENTS	Improvements to public facilities—roads, parks, and community centers.	https://www.gfoa.org/capital-improvement-plan
SCHOOL IMPROVEMENTS	Upgrades and repairs to educational facilities, building or renovating schools.	https://www.ed.gov/school-improvement
PUBLIC FACILITIES	Construction and maintenance of government buildings and community centers.	https://www.nlc.org/program-initiative/community-economic-development
REFUNDING NOTES	Short-term notes issued to temporarily finance the refunding of longer-term debt.	http://www.msrb.org/Glossary/Definition/NOTES.aspx
SEWER IMPROVEMENTS	Upgrades and expansion of sewage and wastewater treatment infrastructure.	https://www.epa.gov/npdes/municipal-wastewater
ELECTRIC LIGHT & POWER IMPS.	Investments in electric-utility infrastructure: generation, transmission, distribution.	https://www.eia.gov
RECREATIONAL FACILITY IMPS.	Investments in parks, sports complexes, and other recreational facilities.	https://www.nrpa.org/

Appendix B. Data Handling

Appendix B.1. Outlier handling

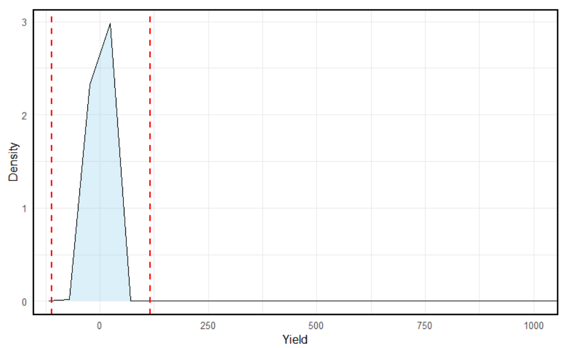
To address potential outliers in the 'yield' variable, a two-step iterative process based on Z-scores is employed. Initially, Z-scores are calculated for each yield observation. Data points with absolute Z-scores exceeding a threshold of 3 are identified as potential outliers and systematically removed from the dataset. Following this initial removal, Z-scores are recalculated for the remaining 'yield' values. This approach allowed for a step-wise identification and removal of potential outliers, providing a systematic means to address extreme values in the 'yield' variable. The iterative nature of the process enabled a controlled adjustment of the dataset while considering the impact on statistical properties. Moreover, implementing this method will result in the removal of approximately 0.05% of the data, suggesting that it is a suitable approach for preserving data integrity.

The table below presents a statistical summary of yield observations in California following the application of each step.

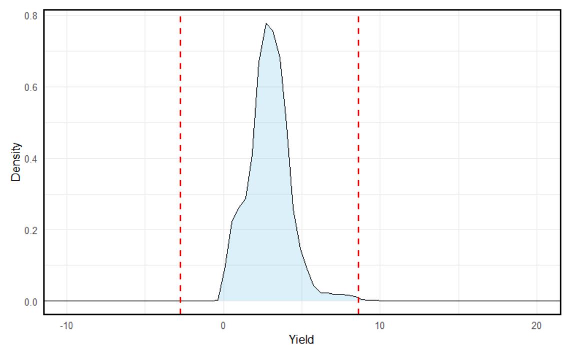
Table B.1: Summary statistics for yield at each step of outlier handling

Steps	Descriptive Statistics					
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Initial Dataset	-117.062	2.104	2.893	3.123	3.684	23858.704
First Step	-106.321	2.104	2.892	2.950	3.683	116.527
Second Step	-1.900	2.099	2.885	2.892	3.671	8.613

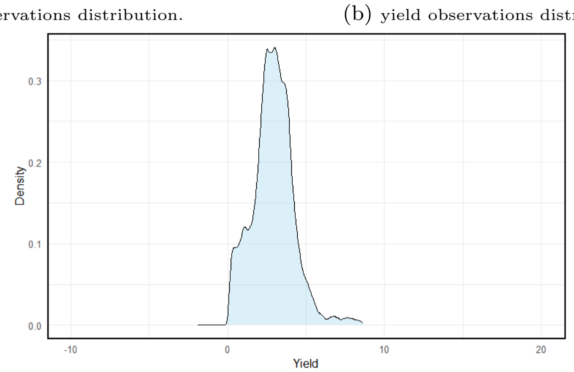
The following plots show the distribution of observed yield in each step.



(a) Initial yield observations distribution.



(b) yield observations distribution after the first step.



(c) yield observations distribution after the second step.

Figure B.1: **Distribution of yields**
This figure shows the distribution of observed yield in California through outliers handling steps. Red horizontal dashed lines represent ± 3 Z-score thresholds.

Appendix C. Bond partitioning for bootstrapping

This appendix describes the bond partitioning scheme used to support the bootstrap construction of green bond term structures and the subsequent portfolio analysis. To ensure that the bootstrapped yield curves are based on economically meaningful and statistically balanced sub-samples, we partition the universe of California municipal green bonds along three key structural dimensions: callability, coupon rates, and tax status. These attributes are known to influence bond pricing, yield dynamics, and investor demand, and they form the basis for the attribute-specific groups employed in the ARL analysis and portfolio construction.[1]

The figures and tables in this appendix provide illustrative evidence on how these characteristics shape the cross-sectional distribution of green bond yields and motivate the partition boundaries adopted in the empirical analysis.

Callability:

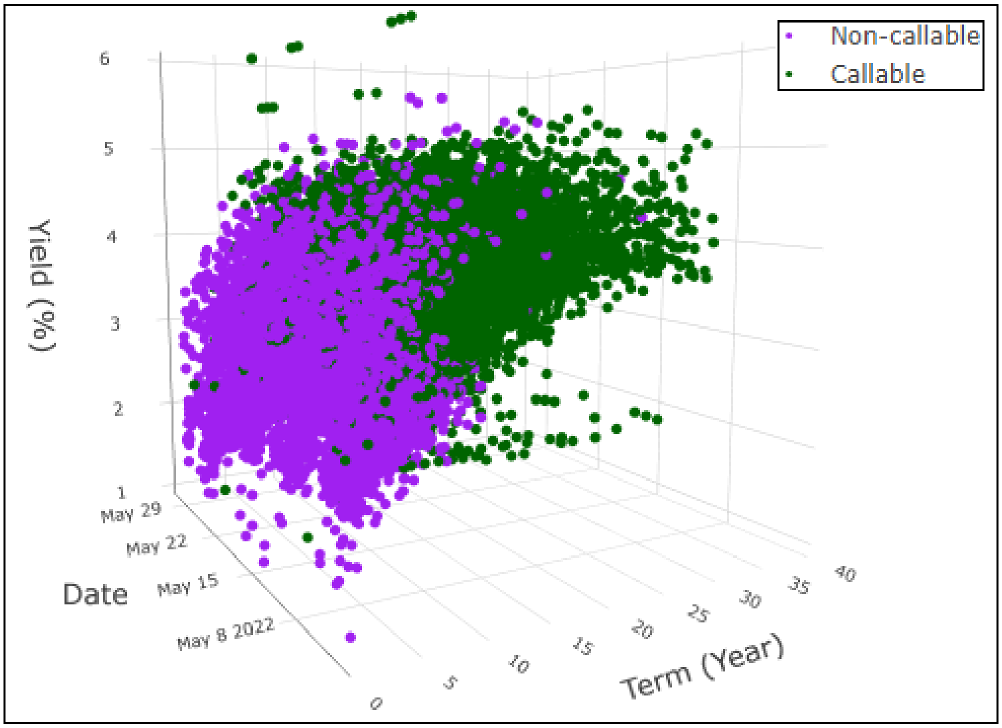


Figure C.1: **Callable and non-callable bond yield variation**. The figure shows the callable (in green) and non-callable (in purple) green bond yield in May 2022.

As an illustrative example, Fig.C.1 shows callable (in green) and non-callable (in purple) green bond yields across the term structure in May 2022.[2] It is apparent that callable bonds

[1]The partitioning scheme is designed to group bonds by salient structural features rather than by observed pricing outcomes; nevertheless, an inherent limitation is that any partitioning may inadvertently align with spread variation. The resulting curve-based spreads should therefore be interpreted as conditional on the chosen structural dimensions rather than as fully orthogonal to all sources of heterogeneity. In particular, residual variation related to issuer credit quality, insurance status, sector, issue size, or liquidity may continue to influence curve differences and is absorbed into the relative pricing measure rather than explicitly removed.

[2]Callable bonds give the issuer the right to redeem the bonds before maturity, particularly when interest rates decline. If rates have fallen since the bond's issuance, the issuer may choose to call the bond and

typically exhibit extended maturities and comparatively higher yields, which is a common feature of the spreads in our sample period see next subsection). Callable bonds would often offer a higher yield than non-callable bonds to compensate investors for the additional risk associated with potential early redemption (Chen et al., 2010). Callable bonds also feature longer maturities because of the flexibility they provide to issuers. The callable feature allows issuers to redeem the bonds before their scheduled maturity, presenting advantages such as interest rate management and flexibility in adjusting debt portfolios. Issuers may issue callable bonds with longer maturities to take advantage of favorable changes in interest rates, and this flexibility can attract investors seeking longer-term commitments.

Coupon rates: The coupon rate is a key determinant of a bond's yield, impacting both its fixed income component and its attractiveness relative to prevailing interest rates. Higher coupon rates generally lead to higher yields, making a bond more appealing to investors in certain market conditions.[3]

---

issue new bonds at a lower interest rate, resulting in early redemption and impacting the investor's yield. Callable bonds also introduce reinvestment risk for investors. Investors must reinvest the proceeds at the prevailing market interest rates if a callable bond is called. If rates have decreased, this reinvestment may occur at lower yields, reducing the overall return. Moreover, price volatility is a characteristic of callable bonds, which can be more pronounced than non-callable bonds, especially during interest rate volatility. A decline in interest rates increases the likelihood of the bond being called, potentially leading to capital losses for investors.

[3] The coupon rate represents the fixed annual interest payment as a percentage of the bond's face value. This rate determines the fixed income component of the bond, contributing to a higher yield. Generally, bonds with higher coupon rates have higher YTMs, assuming the bond is trading at par. The coupon rate affects a bond's attractiveness to investors in a changing interest rate environment. If prevailing interest rates are lower than the bond's coupon rate, the bond becomes more attractive, potentially leading to increased demand and a higher price. Bonds with higher coupon rates are generally less sensitive to interest rate changes, providing a cushion against potential price declines.

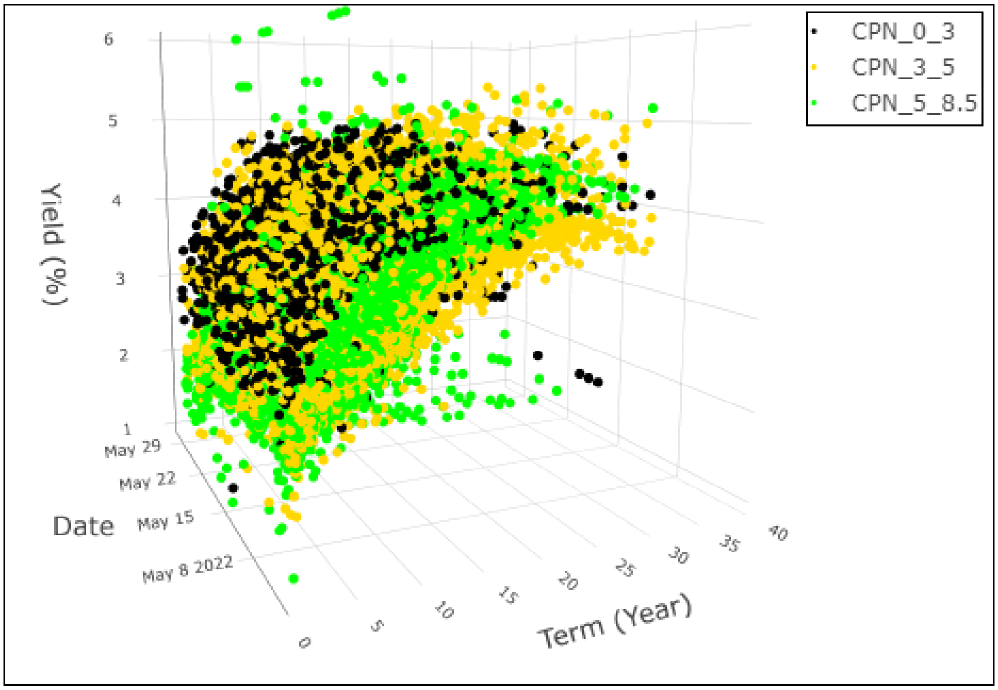
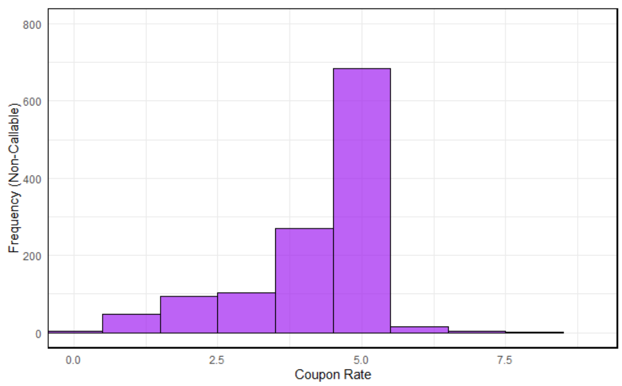
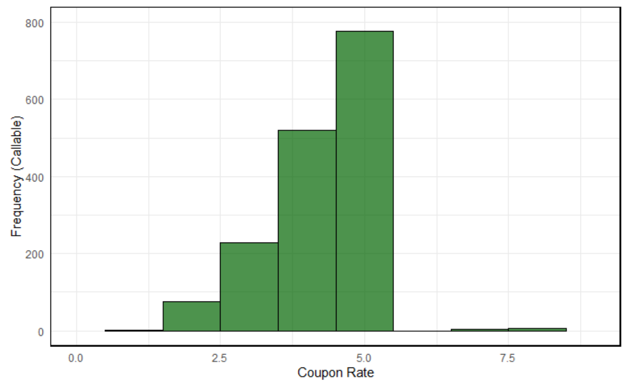


Figure C.2: **Green bond yield in May 2022 color-coded by coupon.**
This figure depicts green bond yields in California in May 2022 color-coded by coupon; 0-3% in black, 3-5% in yellow and 5-8.5% in green.

We use Fig. C.2 as an illustration for the coupon rates of green bond yields in California in May 2022. Clusters of coupon rates are evident for coupons rates between 0-3%, 3-5% and 5-8.5%. For a more comprehensive assessment of the characteristics of these clusters, we inspect the distribution of coupon rates for callable and non-callable bonds. Fig. C.3(a) and (b) shows the distribution of coupon rates for callable and non-callable bonds in California, respectively and Table C.1 provides a statistical overview of the coupon rates for callable and non-callable green bonds in our sample.



(a) Non-callable bonds coupon histogram.



(b) Callable bonds coupon histogram.

Figure C.3: **Distribution of coupon rates for callable and non-callable bonds.** This figure depicts the distribution of callable and non-callable green bond yields in California from 2020–2024.

Table C.1: Statistical summary of coupon rates of green bonds in California

Type	No. of Bonds	Descriptive Summary of Coupon						
		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev.
Non-Callable	1225	0.200	4.000	5.000	4.140	5.000	7.690	1.170
Callable	1615	1.000	3.720	4.000	4.150	5.000	8.500	0.890

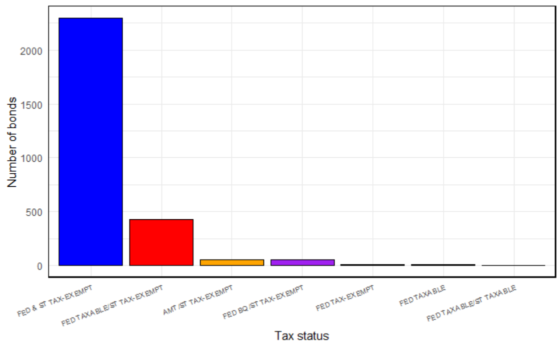
To ensure balanced sample sizes across partitions, it is important to find an optimal trade-off between the number of bonds and the coupon rate intervals. An efficient approach involves considering the quartiles of the coupon intervals for both callable and non-callable bonds. We accordingly segment the coupon range into three distinct partitions: minimum to the 1st quartile, 1st quartile to the 3rd quartile, and 3rd quartile to the maximum for each bond partition. To ensure well-defined and more precise intervals with a relatively similar range between callable and non-callable bonds, we round down the 1st quartiles of non-callable and callable bonds to 3.0% and 3.5%, respectively, and the 3rd quartiles of non-callable and callable bonds to 7.0% and 8.5%, respectively. Table C.2 presents the statistical summary of the resultant partitions in our sample. As shown in the table, each partition includes a sufficient number of bonds to be used in the bootstrapping application and display suitable statistical properties.

Table C.2: Statistical summary of non-callable and callable bonds for partitions of coupon rates

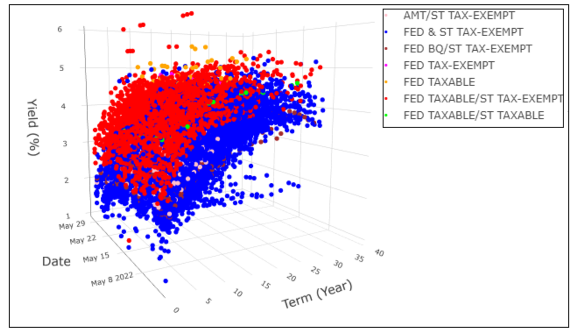
	Non-Callable								Callable							
Year	Coupon Rates (%)	No. of Bonds	Min	Max	Mean	Median	3rd Qu.	St.dev.	Coupon Rates (%)	No. of Bonds	Min	Max Max	Mean	Median	3rd Qu.	St.dev.
2020	[0.2-3.0)	85	0.593	2.945	2.326	2.440	2.622	0.448	[1.0-3.5)	204	1.701	3.500	3.024	3.000	3.153	0.284
	[3.0-5.0)	159	2.991	4.932	3.585	3.630	4.000	0.458	[3.5-5.0)	276	3.502	4.875	3.988	4.000	4.000	0.185
	[5.0-7.69]	363	5.000	7.690	5.098	5.000	5.000	0.354	[5.0-8.5]	537	5.000	8.500	5.059	5.000	5.000	0.400
2021	[0.2-3.0)	158	0.245	2.976	1.971	2.115	2.500	0.681	[1.0-3.5)	291	1.000	3.500	2.905	3.000	3.098	0.373
	[3.0-5.0)	294	2.991	4.857	3.752	4.000	4.000	0.407	[3.5-5.0)	439	3.502	4.842	3.992	4.000	4.000	0.152
	[5.0-7.69]	438	5.000	7.690	5.081	5.000	5.000	0.339	[5.0-8.5]	548	5.000	8.350	5.053	5.000	5.000	0.385
2022	[0.2-3.0)	159	0.199	2.976	1.851	1.984	2.445	0.682	[1.0-3.5)	302	1.000	3.500	2.823	3.000	3.000	0.419
	[3.0-5.0)	309	2.991	4.928	3.797	4.000	4.000	0.378	[3.5-5.0)	521	3.502	4.875	3.996	4.000	4.000	0.136
	[5.0-7.69]	620	5.000	7.690	5.038	5.000	5.000	0.218	[5.0-8.5]	712	5.000	8.350	5.029	5.000	5.000	0.276
2023	[0.2-3.0)	148	0.317	2.945	1.913	1.984	2.451	0.608	[1.0-3.5)	306	1.000	3.500	2.810	3.000	3.000	0.432
	[3.0-5.0)	303	2.991	4.928	3.829	4.000	4.000	0.361	[3.5-5.0)	526	3.502	4.875	4.001	4.000	4.000	0.147
	[5.0-7.69]	606	5.000	7.690	5.049	5.000	5.000	0.205	[5.0-8.5]	724	5.000	8.250	5.021	5.000	5.000	0.130
2024	[0.2-3.0)	113	0.514	2.945	1.899	1.965	2.338	0.549	[1.0-3.5)	296	1.000	3.500	2.803	3.000	3.000	0.405
	[3.0-5.0)	249	3.000	4.700	3.850	4.000	4.000	0.332	[3.5-5.0)	527	3.502	4.875	4.001	4.000	4.000	0.152
	[5.0-7.69]	527	5.000	6.950	5.062	5.000	5.000	0.228	[5.0-8.5]	714	5.000	5.500	5.016	5.000	5.000	0.080

Tax status: The tax status of bonds can influence the yields of bonds. Since investors receive tax benefits, tax-exempted bonds yields could be lower than taxable ones with similar risk profiles.[4] Fig. C.4(a) shows the tax status frequency in California and Fig. C.4(b) shows the green bonds yields for different tax statuses in May 2022. The tax status of green bonds in California is concentrated in two categories: the *Federal and State Tax Exempt* (represented by blue spots) and *Federal Taxable and State Tax Exempt* status (represented by red spots). As Fig. C.4 reveals, most of the green bonds in California are *Federal and State Tax Exempt* and typically, the *Federal and State Tax Exempt* bonds have comparatively lower yields relative to the *Federal Taxable and State Tax Exempt* bonds.

[4]Taxes affect municipal bond yields because the interest income from these bonds is typically exempt from federal income tax, and sometimes state and local taxes as well. This tax advantage makes municipal bonds more attractive to investors in higher tax brackets, allowing issuers to offer lower yields compared to taxable bonds. In contrast, bonds without such tax exemptions need to offer higher yields to compensate investors for the tax burden. Therefore, the tax status of a bond directly influences its yield (Cestau et al., 2019; Perlovsky and DeMarco, 2018).



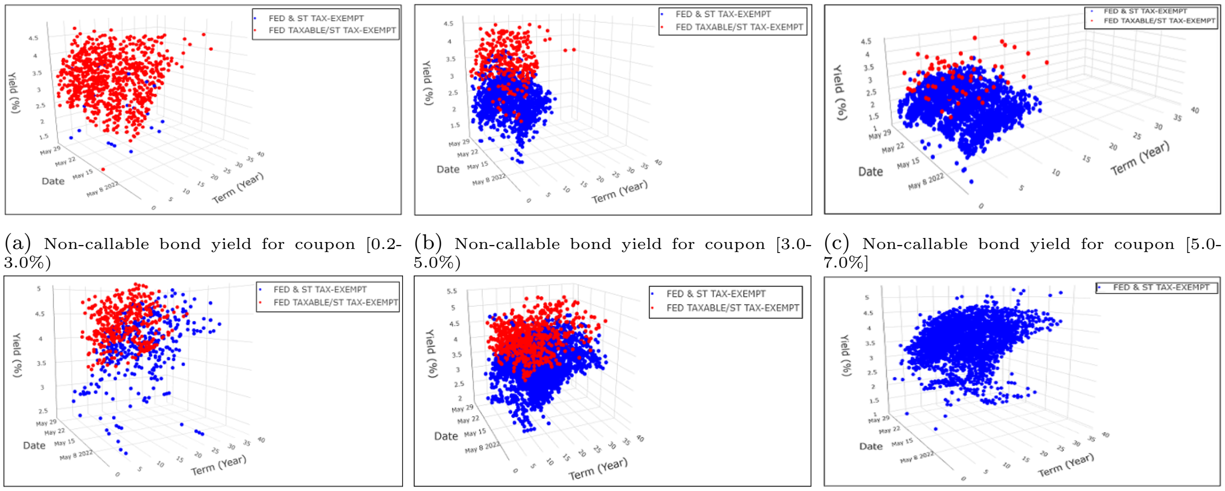
(a) Green bonds tax status in California.



(b) Bond yields in May 2022

Figure C.4: **Green bonds tax statuses and yield in May 2022 color-coded by tax status.** This figure depicts the number of green bond according to tax status and green bond yields in California color-coded by tax status. Panel (a) depicts the number of bonds within each tax province in our data set and panel (b) shows the bond yield in May 2022 color-coded by tax status – blue for *Federal and State Tax Exempt*, red for *Federal Taxable and State Tax Exempt* and other colors for the remaining statuses.

After we partition the sample of green bonds according to the coupon rate and callability, and examine the concentration of the tax status of the bonds within each of these partitions. Figure C.5 provides a visual representation of the concentration of tax statuses across the coupon rate and callability partitions, where interesting clustering properties emerge in a representative sample in May 2022. Note that California municipal green bonds exhibit similar characteristics during the sample period (2020-2024) as depicted in the next subsection.



(a) Non-callable bond yield for coupon [0.2-3.0%)
(b) Non-callable bond yield for coupon [3.0-5.0%)
(c) Non-callable bond yield for coupon [5.0-7.0%]
(d) Callable bond yield for coupon [1.0-3.5%)
(e) Callable bond yield for coupon [3.5-5.0%)
(f) Callable bond yield for coupon [5.0-8.5%)

Figure C.5: **Green bond yield in May 2022 color-coded by tax status for partitions of the coupon rates.** This figure depicts non-callable (top panels) and callable (bottom panels) green bond yields in California color-coded with tax status for partitions of coupon rates in May 2022. The *Federal Taxable and State Tax Exempt* (FTSE) and *Federal and State Tax Exempt* (TE) are represented in red and blue spots, accordingly.

Accordingly, we make the partitions of the green bonds in California to ensure representative sufficiently balanced sub-samples in each partition based on their callability, coupon rates, and tax status. We use the abbreviation FTSE for *Federal Taxable and State Tax Exempt* bonds,[5] and TE for all tax-exempt bonds,[6] with *Federal and State Tax Exempt* bonds comprising the majority of this category.

- Group 1 (G1): This partition includes non-callable bonds with coupon rates ranging from 0.20 - 3%, see Fig. C.5(a). This partition prominently features bonds categorized as FTSE (represented by red spots). Consequently, no further separation is considered necessary for this specific partition.

- Groups 2 and 3 (G2 and G3): This partition covers non-callable green bond yields with coupon rates from 3 - 5%, see Fig. C.5(b). These bonds have clear clustering on two

[5]To maintain a homogeneous representation of tax implications, we exclude other taxable bonds, including *FED TAXABLE* and *FED TAXABLE/ST TAXABLE*, which have the lowest counts of 6 and 5 green bonds, respectively (see Fig. C.4(a)) ensuring that our taxable bonds represent the *Federal Taxable and State Tax Exempt* status.

[6]Other tax-exempt bonds include *AMT/ST TAX-EXEMPT*, *FED BQ/ST TAX-EXEMPT* and *FED TAX-EXEMPT*.

tax statuses, FTSE and TE. Thus, we distinguish between these two groups of bonds and label the FTSE and TE partitions as Groups 2 and 3, respectively.

- Group 4 (G4): This partition covers non-callable bond yields with coupon rates from 5 - 7.69%, see C.5(c). There is a predominant concentration of bonds with the TE and the FTSE tax status. The challenge is that the number of FT bonds is insufficient for proper segmentation. Thus, no additional partitioning was considered in this coupon partition of non-callable bonds, and we anticipate relatively noisier patterns from this group of bonds.

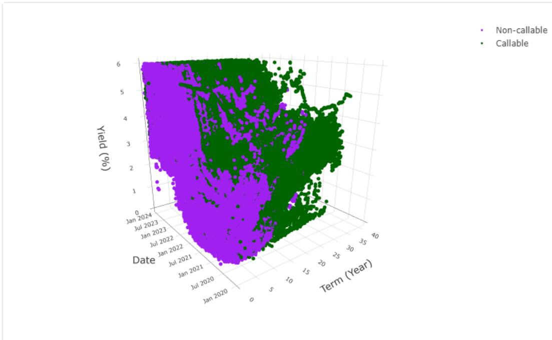
- Groups 5 and 6 (G5 and G6): This partition considers callable green bonds with coupon rates ranging from 1 - 3.5%, see Fig. C.5(d). For these bonds, there is a clear distinction between FTSE and TE bonds labelled as Groups 5 and 6, respectively.

- Groups 7 and 8 (G7 and G8): This partition includes callable green bonds with coupon rates ranging from 3.5 - 5%, see Fig. C.5(e). We further separate the FTSE and TE bonds within this partition and obtain the groups Groups 7 and 8, respectively.

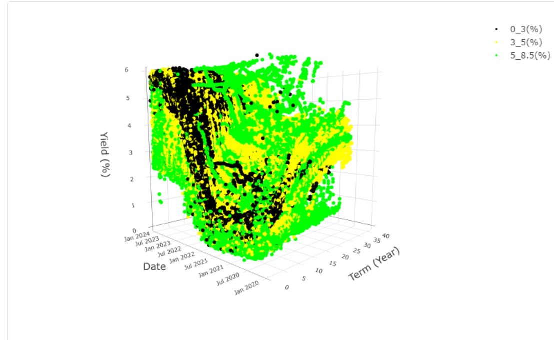
- Groups 9 (G9): This partition considers callable green bonds with coupon rates ranging from 5 - 8.50%, see Fig. C.5(f). Most of these green bonds belong to the TE tax status. There is a group with bonds from other tax status, predominantly FTSE tax status, however, the quantity of these bonds is not sufficient for a robust segmentation. We keep this as one group labelled as Group 9 and we anticipate noisier bootstrap curves on some days from this group.

Appendix C.1. YTM color-coded by structuring properties for the entire sample period.

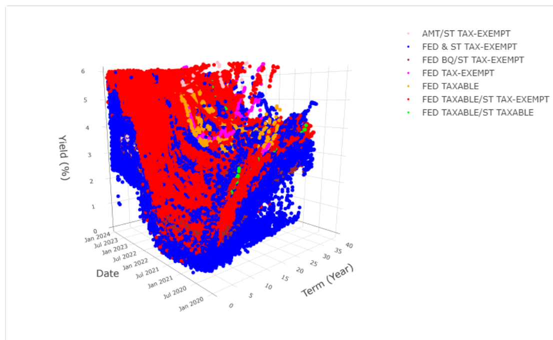
The following plot presents a series of California green bond YTM 3d-plots, color-coded based on callability, coupon rate and tax-status.



(a) Green bonds YTM in 2020-2023, color-coded by callability.



(b) Green bonds YTM in 2020-2023, color-coded by Coupon rates.



(c) Green bonds YTM in 2020-2023, color-coded by Tax status.

Figure C.6: **Green bond YTM, color-coded by callability, coupon range and tax status**. This figure shows the YTM of California green bonds from 2020 to 2023, color-coded by different structuring attributes. Subplot (a) presents a color-coded 3D plot based on callability, while subplots (b) and (c) display yields color-coded by coupon rates and tax status, respectively.

Appendix C.2. Calculation of remaining time to maturity

As outlined in subsection 2.1.1 of the main paper, our methodology adopts a dynamic approach with a daily frequency. To capture the evolving trend of bonds' yields, it is necessary to update both τ_n and τ_N. These terms undergo continuous updates throughout the dataset's entire time span, following the Actual/Actual day count convention. Actual/Actual or ACT/ACT calculates the daily interest using the actual number of days in the year and then multiplies that by the actual number of days in each time period. The U.S. Treasury bonds and municipal green and conventional bonds use this convention in calculations of rates day count convention.

Appendix D. Construction of green bond yield spreads

Firstly, we use the observed daily YTM of green bonds and a risk-free reference yield, such as the U.S. Treasury par yield, to discount the respective cash flows of the bonds to compute synthetic green and conventional bond prices. We then use these synthetic bond prices to compute the green and reference equivalent Zero-Coupon Bond Yields to Maturity (ZCBYTM).

More specifically, we compute the value of a green bond i by using the YTM of this green bond to discount its cash flows. Thus, all the cash flows of a green bond are discounted by the same fixed rate – its YTM. The value of this bond is then used to compute the equivalent green ZCBYTM. The corresponding reference rate from a risk-free yield curve is selected, so it matches the tenor of the green bond i. We use this reference rate as the fixed rate to discount the (same) cash flows to determine the equivalent risk-free ZCBYTM. The difference between green ZCBYTM and risk-free ZCBYTM determines the spread of green bond i.

- **Green Equivalent ZCBYTMs**

  We consider green bond i at time t with coupon amount C_i, face value (par amount) FV_i, coupon frequency m_i, and annualized $YTM^G_{i,t}$, which has payments in $\hat{N}$ periods from first issuance over the entire term of the bond and in N periods according to the remaining time to maturity. As we propose a dynamic approach on a daily basis, we define $\tau_{i,n}$ representing the remaining year fraction to the n^{th} payment at time t for bond i as follows:

$$\tau_{i,n} := \max\left\{n - \frac{t}{(365/m_i)}, 0\right\}. \tag{D.1}$$

  We also define the indicator function as follows:

$$\mathbb{I}(\tau_{i,n}) := \begin{cases} 1, & \text{if } \tau_{i,n} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{D.2}$$

  When $\tau_{i,n}$ reaches zero, it signifies that we have either reached or exceeded the coupon payment date for coupon i. At this point, the indicator function will be employed to exclude previously paid coupons from the calculation.

  Thus, we calculate equivalent green ZCBYTMs at time t for bond i, denoted as $YTM^{(GZCB)}_{i,t}$ according to the following steps:

1. Calculate the present value of bond i at time t denoted as $P_{i,t}$ as follows:

$$P_{i,t} = \sum_{n=1}^{N} \frac{C_i \mathbb{I}(\tau_{i,n})}{(1+YTM_{i,t}^{(G)})^{\tau_{i,n}}} + \frac{FV_i}{(1+YTM_{i,t}^{(G)})^{\tau_{i,N}}}. \tag{D.3}$$

2. Define an equivalent zero-coupon bond, characterized by a face value denoted as $\widetilde{FV}$. The face value of the equivalent zero coupon bond i is defined as follows:

$$\widetilde{FV_i} := \max\left\{FV_i, FV_i + C_i\right\}. \tag{D.4}$$

   $\widetilde{FV}$ is adjusted for bonds that pay both a coupon and par amount on the maturity date. This adjustment is important for bonds where the final coupon is paid at maturity, as it ensures that the equivalent zero-coupon bond properly accounts for the entire cash flow structure. It is defined as the maximum of the original bond's FV or $FV + C$ if there's a coupon on the last day. Subsequently the present value of the equivalent zero coupon bond i at time t, $P_{i,t}^{ZCB}$ is defined as:

$$P_{i,t}^{ZCB} = \frac{\widetilde{FV_i}}{(1+YTM_{i,t}^{(GZCB)})^{\tau_{i,N}}}. \tag{D.5}$$

3. Equate $P_{i,t}$ to $P_{i,t}^{ZCB}$ and rearranging, we can obtain $YTM_{i,t}^{(GZCB)}$ as follows :

$$\frac{\widetilde{FV_i}}{(1+YTM_{i,t}^{(GZCB)})^{\tau_{i,N}}} = \sum_{n=1}^{N} \frac{C_i \mathbb{I}(\tau_{i,n})}{(1+YTM_{i,t}^{(G)})^{\tau_{i,n}}} + \frac{FV_i}{(1+YTM_{i,t}^{(G)})^{\tau_{i,N}}}, \tag{D.6}$$

$$YTM_{i,t}^{(GZCB)} = \left[\frac{1}{\widetilde{FV_i}}\left(\sum_{n=1}^{N} \frac{C_i \mathbb{I}(\tau_{i,n})}{(1+YTM_{i,t}^{(G)})^{\tau_{i,n}}} + \frac{FV_i}{(1+YTM_{i,t}^{(G)})^{\tau_{i,N}}}\right)\right]^{-\frac{1}{\tau_{i,N}}} - 1. \tag{D.7}$$

- **Reference equivalent ZCBYTM**

  We compute a reference equivalent ZCBYTM rate, denoted as $YTM_{i,t}^{(RZCB)}$, using a reference risk-free rate. Discounting rates are derived from the fitted spline curves of annual reference rates at time t, based on the same tenor used in the green ZCBYTM calculation. These rates are applied similarly to the previous stage, discounting the associated cash flows as follows:

$$YTM_{i,t}^{(RZCB)} = \left[\frac{1}{\widetilde{FV_i}} \left(\sum_{n=1}^{N} \frac{C_i \mathbb{I}(\tau_{i,n})}{(1 + r_t^{(Tr)}(\tau_{i,N}))^{\tau_{i,n}}} + \frac{FV_i}{(1 + r_t^{(Tr)}(\tau_{i,N}))^{\tau_{i,N}}} \right) \right]^{-\frac{1}{\tau_{i,N}}} - 1, \tag{D.8}$$

where $r_t^{(Tr)}(\tau_{i,N})$ is the corresponding reference rate for $\tau_{i,N}$ year(s) maturity bond at time t and $YTM_{i,t}^{(RZCB)}$ is the reference equivalent ZCBYTM rate of the bond i at time t corresponding to the discount factor $r_t^{(Tr)}(\tau_{i,N})$.

- **Spread Calculation**

  Once the $YTM_{i,t}^{(GZCB)}$ and $YTM_{i,t}^{(RZCB)}$ have been determined, we can determine the daily spread based on YTM. We define the non-parametric $S_{i,t}$ as follows: [7]

$$S_{i,t} := YTM_{i,t}^{(GZCB)} - YTM_{i,t}^{(RZCB)}. \tag{D.9}$$

Essentially, the tenor-specific approach uses a single matching tenor point on the green and reference yield curves and compares a time series of green ZCBYTM with a time series of risk-free ZCBYTM at the same tenor point to obtain the spread of the green bond i. We repeat this method for all green bonds on day t, and consequently compute the spread for all available green bonds on this day.

[7] We opt to define a YTMZCB corresponding to the reference rate $r_t^{(Tr)}(\tau_{i,N})$, as opposed to a direct comparison of this rate ($r_t^{(Tr)}(\tau_{i,N})$) with the $YTM_{i,t}^{(RZCB)}$. This approach allows us to capture specific bond characteristics, including coupon rate, coupon frequency, and maturity, through defining a reference rate. From a mathematical standpoint, the rationale lies in dealing with a non-linear (hyperbolic) transformation of rates. It is important to note that the input rate ($r_t^{(Tr)}(\tau_{i,N})$) undergoes a non-linear transformation, and this discrepancy with $YTM_{i,t}^{(RZCB)}$ becomes more pronounced, especially as the maturity period extends.

Appendix D.1. U.S. par yield curve features

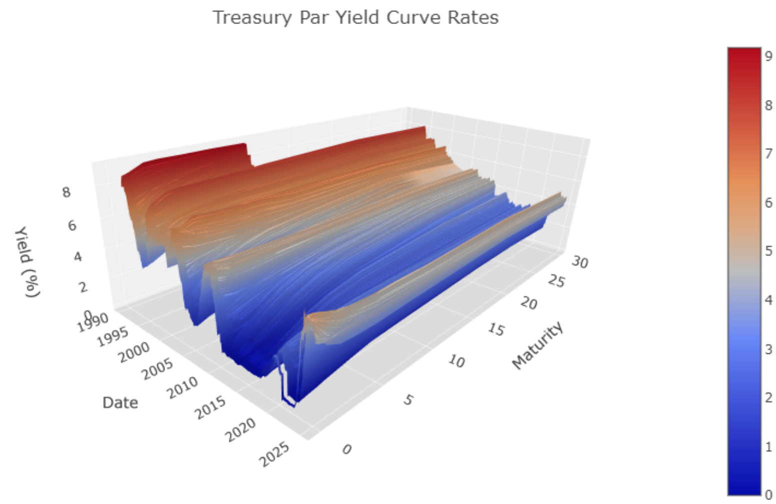
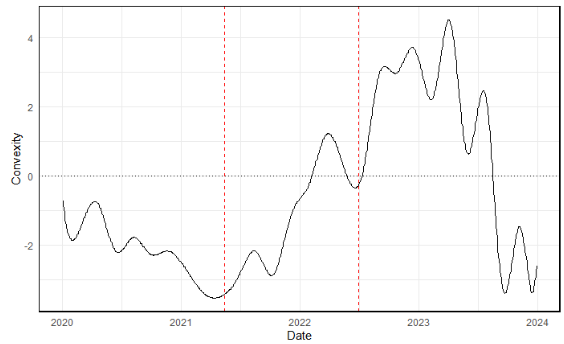


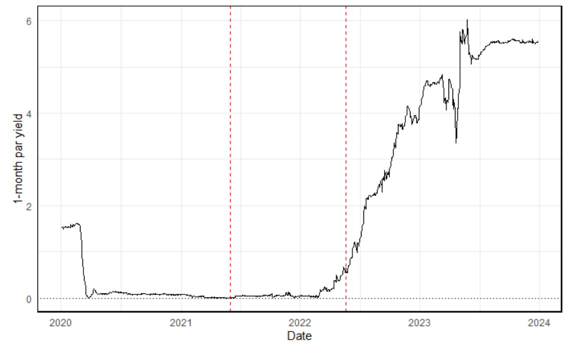
Figure D.1: **U.S treasury par yields**
This figure depicts the dynamic evolution of U.S treasury par yields between 1990-2023.

Fig. D.1 presents the dynamic evolution of U.S Par yields over time.

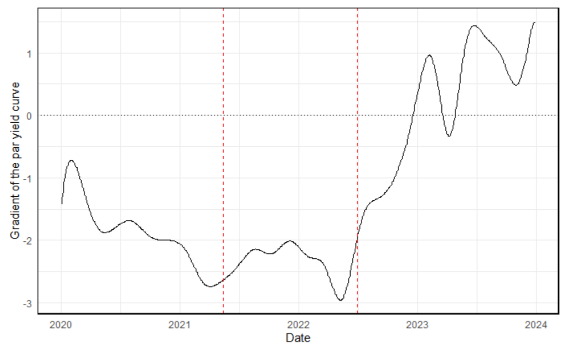
To enhance our understanding of the behavior of green bond spreads, we investigate features of the U.S. par yield curve, including convexity and gradient.



(a) U.S. par yield curve convexity (β_2)



(b) Interest rate trend



(c) U.S. par yield curve gradient(slope)

Figure D.2: **U.S. par yield convexity and interest rate.**
This figure shows the convexity and slope of U.S. par yield curve and 1-month yield trend within 2020-2023. Two vertical dashed lines mark the dates 2021-05-13 and 2022-09-06.

Thus, we apply the Nelson-Siegel model to the daily rates of different tenors of the U.S. par yield curve, which serve as the reference rates in the spreads calculations, specifically,

$$Y(\tau) = \beta_0 + \beta_1 \left(1 - e^{-\frac{\tau}{\lambda}}\right) + \beta_2 \left(\frac{\tau}{\lambda} - e^{-\frac{\tau}{\lambda}}\right),$$

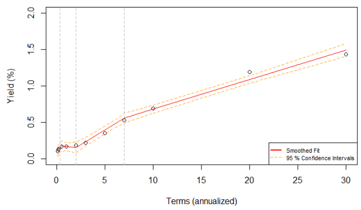
where β_0, β_1, and β_2 represent the level, slope, and curvature of the yield curve, respectively. Note that, β_0 captures long-term average interest rates, β_1 reflects the speed of short-term rates converging to the long-term average, and β_2 denotes the convexity of the yield curve, indicating deviations from linearity in the relationship between maturity and yield (Diebold and Li, 2006).

We use 30 cubic B-splines to smooth these coefficients. We then plot smoothed β_2 and β_1, representing the yield curve's convexity and gradient. We include the 1-month U.S. Treasury par yields to reflect market expectations of interest rates. Figure D.2 displays the convexity and slope (gradient) of the U.S. par yield curve alongside the prevailing interest rate regime within the specified interval.

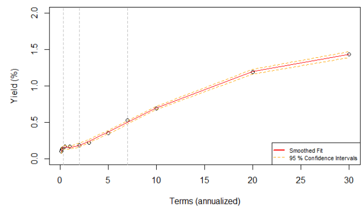
Appendix D.2. B-Spline curves fitted to U.S. Treasury yield par curve

B-Spline curves are a mathematical representation commonly used for estimating yield curves. The term "spline" refers to a flexible strip used to manually draw smooth curves, and "Basis" indicates that the curve is constructed as a linear combination of basis functions. B-Spline curves are particularly useful for capturing complex movements and irregularities in yield curves over different maturities (Yallup, 2012). More specifically, B-spline curves provide flexibility in representing the shape of the yield curve, allowing for a smooth and continuous representation. The degree of smoothness is controlled by the order of the B-Spline curve, with higher orders producing more flexibility but also requiring more control points. Control points or knots act as anchor points that influence the shape of the curve, and the basis functions dictate how each control point contributes to the overall curve. Cubic splines are optimal in minimizing Mean Squared Error (MSE), see (Hastie et al., 2009).

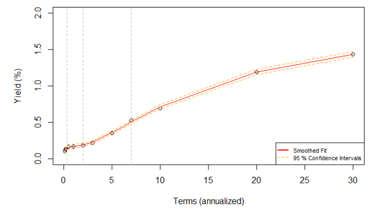
For demonstration purposes, Fig. D.3 depicts the fitted spline curves (order 3) to the U.S. Treasury yield par curve on four typical dates within each year of our study . Each row presents different degrees of splines. The fit is satisfactory and the splines capture the distinct patterns that prevail during the sample period. Order 3 splines are used for the analysis, as they offer the best fit.



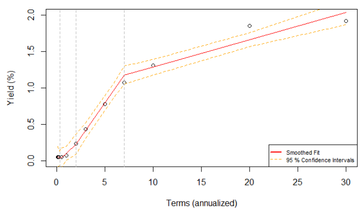
(a) U.S. Treasury Par yield (B1 Spline) on 2020-05-26



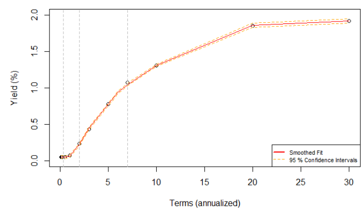
(b) U.S. Treasury Par yield (B2 Spline) on 2020-05-26



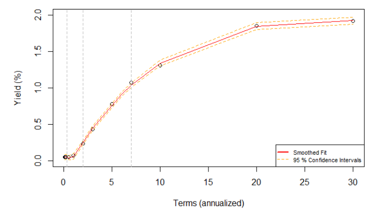
(c) U.S. Treasury Par yield (B3 Spline) on 2020-05-26



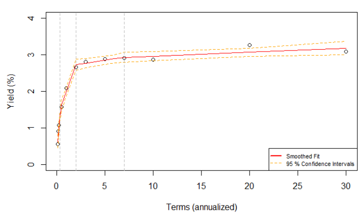
(d) U.S. Treasury Par yield (B1 Spline) on 2021-07-15



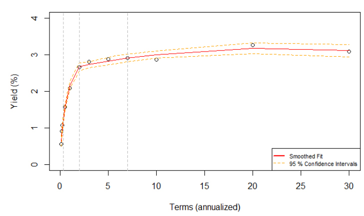
(e) U.S. Treasury Par yield (B2 Spline) on 2021-07-15



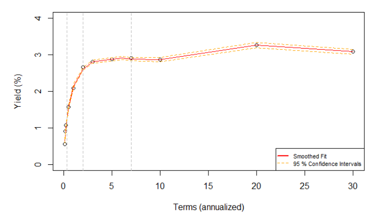
(f) U.S. Treasury Par yield (B3 Spline) on 2021-07-15



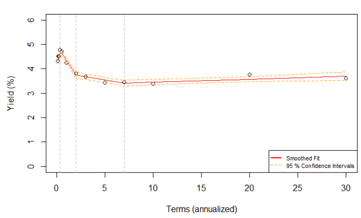
(g) U.S. Treasury Par yield (B1 Spline) on 2022-05-23



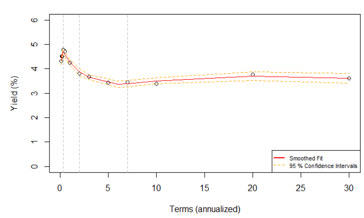
(h) U.S. Treasury Par yield (B2 Spline) on 2022-05-23



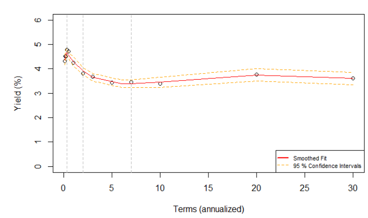
(i) U.S. Treasury Par yield (B3 Spline) on 2022-05-23



(j) U.S. Treasury Par yield (B2 Spline) on 2023-03-17



(k) U.S. Treasury Par yield (B3 Spline) on 2023-03-17



(l) U.S. Treasury Par yield (B3 Spline) on 2023-03-17

Figure D.3: **Fitted curves to the U.S. Treasury Par yield**
The figure displays B-spline curves of varying degrees fitted to the U.S. Treasury Par yield on different days across the four-year dataset interval. The red line represents the fitted B3-Spline bootstrapped curve, while the dotted orange curve depicts the 95% confidence intervals for the predicted rates. The grey vertical lines represent the control points or knots.

Appendix E. Association Rule Mining Methods: Apriori and Bayesian Formulations

Our experimental framework applies **Association Rule Learning (ARL)** Agrawal et al. (1993)—a machine learning technique that, when viewed through a Bayesian network structure, provides a probabilistic interpretation—to analyze green bond spreads. This approach addresses key questions central to investor screening, such as identifying bonds with spreads that are either positive or negative relative to risk-free benchmarks (e.g., U.S. Treasuries).

Effective screening requires an analysis of both single-tenor spreads and term structures, capturing factors like bond maturity and yield curve shapes. Yield curve shapes—such as inverted, flat, or steep—reflect broader economic conditions that influence spreads based on a bond's tenor and attributes.

Additionally, screening practices may target specific bond characteristics, such as the **Use of Proceeds (UOP)** category, which affects green bonds' potential to meet environmental goals. For example, the UOP category (e.g., green energy vs. transportation) affects the impact of green labels, index listings, and environmental outcomes. It can also shape default and repayment risks as characteristics like tenor, maturity, callability, coupon rate, and issuer sector interact with UOP to directly influence spreads.

In the following subsections, we outline our approach, starting with a description of ARL and the Apriori algorithm. We then transition to a **Bayesian Network Interpretation** for a probabilistic understanding of ARL, leading into **Bayesian Model Selection** for interpreting results rigorously.

Appendix E.1. Frequentist Approach: Association Rule Learning and the Apriori Algorithm

Association Rule Learning (ARL) uncovers relationships between variables in a dataset, operating based on observed co-occurrence frequencies. The Apriori algorithm identifies association rules that meet predefined support and confidence thresholds Agrawal et al. (1993). In our green bond context, let $\{\mathbf{x}_{i,t}\}_{i=1,t=1}^{M,T}$ be our set of observations for M bonds over T time stamps, giving N = M × T total bond-time observations. Each observation represents a bond-time transaction:

$$\mathbf{x}_{i,t} = [x_{a_1}, \ldots, x_{a_D}, s_{i,t}] \tag{E.1}$$

where:

- $\{x_{a_1}, \ldots, x_{a_D}\}$ represents the one-hot encoded categorical attributes, where D is the total number of categories across all attributes. For example:
  - Rating levels: [AAA, AA, A] $\rightarrow$ [1,0,0]
  - Tax status: [Exempt, Taxable] $\rightarrow$ [1,0]
  - Coupon range: [0-3%, 3-5%, 5-8.5%] $\rightarrow$ [0,1,0]
- $s_{i,t} \in \{0, 1\}$ indicates whether bond i at time t belongs to the specific spread category being analyzed (S+, S-, or S0)

Let $\mathcal{T} = [\mathbf{X_A}, \mathbf{s}] \in \{0, 1\}^{N \times (D+1)}$ be our transaction database where:

- Matrix $\mathbf{X_A} \in \{0, 1\}^{N \times D}$ represents all N bond-time observations across their one-hot encoded categorical attributes, where each row contains the concatenated one-hot vectors of all attributes. For a bond at time t with:
  - Rating = AAA: [1,0,0]
  - Tax status = Exempt: [1,0]
  - Coupon = 3-5%: [0,1,0]

  The corresponding row in $\mathbf{X_A}$ would be: [1,0,0,1,0,0,1,0]
- Vector $\mathbf{s} \in \{0, 1\}^N$ indicates for each observation whether it belongs ($s_{i,t} = 1$) or not ($s_{i,t} = 0$) to the specific spread category being analyzed

An **association rule** is defined as a mapping from a set of attribute category indicators to the specific spread category. The effectiveness of these rules is evaluated using three primary metrics: **Support**, **Confidence**, and **Lift**, defined as follows:

1. The *Support* measures the proportion of bond-time observations where the attribute category combination occurs together with the analyzed spread category:

   $$\text{Supp}(\mathbf{X_A} \rightarrow \mathbf{s}) = \frac{\text{count}(\mathbf{X_A} \cap \mathbf{s})}{N}, \tag{E.2}$$

   where $\text{count}(\mathbf{X_A} \cap \mathbf{s})$ counts observations containing both the specified attribute category indicators and the analyzed spread category ($s_{i,t} = 1$).

2. The *Confidence* indicates the reliability of the rule by measuring the proportion of transactions containing the attribute category indicators that also belong to the analyzed spread category:

$$\mathrm{Conf}(\mathbf{X_A} \rightarrow \mathbf{s}) = \frac{\mathrm{Supp}(\mathbf{X_A} \cap \mathbf{s})}{\mathrm{Supp}(\mathbf{X_A})} = \frac{\mathrm{count}(\mathbf{X_A} \cap \mathbf{s})}{\mathrm{count}(\mathbf{X_A})}. \tag{E.3}$$

3. The *Lift* assesses the degree of dependency between attribute category indicators and the analyzed spread category:

$$\mathrm{Lift}(\mathbf{X_A} \rightarrow \mathbf{s}) = \frac{\mathrm{Conf}(\mathbf{X_A} \rightarrow \mathbf{s})}{\mathrm{Supp}(\mathbf{s})} = \frac{\mathrm{Supp}(\mathbf{X_A} \cap \mathbf{s})}{\mathrm{Supp}(\mathbf{X_A}) \cdot \mathrm{Supp}(\mathbf{s})}. \tag{E.4}$$

A Lift > 1 indicates a positive association, Lift $= 1$ signifies independence, and Lift < 1 suggests a negative association between the attribute category combination and the analyzed spread category.

Appendix E.2. Probabilistic View of ARL

Bayesian methods are advantageous for handling uncertainty, small datasets, or sparse data, as they avoid overfitting by incorporating prior distributions. They also allow for a more nuanced interpretation of rules, as posterior probabilities provide a measure of the certainty of discovered associations.

The Bayesian framework in ARL provides a probabilistic interpretation of relationships between categorical bond attributes (represented in our transaction database $\mathcal{T} = [\mathbf{X_A}, \mathbf{s}]$) and spread outcomes. While this framework can accommodate multiple outcomes, we will show how it specializes effectively to our specific case where we analyse one spread category at a time over our N = M $\times$ T bond-time observations.

By framing association rules within a Bayesian bipartite network structure, we can represent dependencies between encoded categorical indicators and spread outcomes as directed probabilistic relationships. In the general case, these relationships connect the one-hot encoded indicators of m categorical attributes to spread outcomes at different time points, while in our specific application, we analyze one spread category (S+, S-, or S0) at a time.

In this Bayesian bipartite network, one-hot encoded categorical indicators form one set of nodes, while spread outcomes form the other set. Directed edges from encoded indicators to spread outcomes represent probabilistic dependencies. This setup allows for a clear and structured representation of conditional relationships, as illustrated in Fig. E.1, where the rule connecting encoded attributes to outcomes is represented as directed edges from indicator nodes to spread nodes.

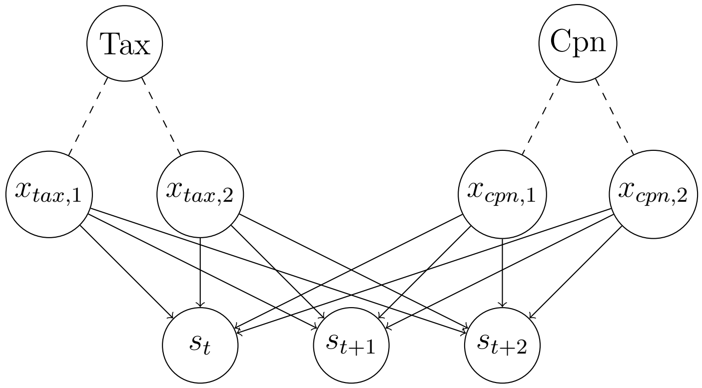


Figure E.1: **Bayesian Network Structure**. Dashed lines represent one-hot encoding relationships, while arrows represent probabilistic dependencies. Spread outcomes s_t evolves across time.

In the general framework, for each categorical attribute a (e.g., rating, tax status), category-level priors $P(x_{a,k})$ represent the initial probabilities of observing each category k independently. For instance, for coupon categories $k \in \{1, 2, 3\}$ corresponding to ranges $\{0\text{-}3\%, 3\text{-}5\%, 5\text{-}8.5\%\}$, we assign priors $P(x_{\text{cpn},k})$ that reflect our initial beliefs about their occurrence and importance for spread behaviour.

Each node in our network corresponds to one category of a categorical attribute, where directed links indicate parent-child relationships. The Bayesian network represents a unique joint probability distribution:

$$P(x_{1,k_1}, \ldots, x_{m,k_m}, s) = \prod_{i=1}^{n} P(x_{i,k_i} \mid Pa(x_{i,k_i})) \tag{E.5}$$

where x_{i,k_i} represents category k_i of attribute i, and s is the spread outcome.

With category priors, the general Bayesian probabilistic view can be represented as:

$$\begin{aligned} P(x_{1,k_1}, \ldots, x_{m,k_m}, s) &= \prod_{i=1}^{m} P(x_{i,k_i}) P(s \mid x_{1,k_1}, \ldots, x_{m,k_m}) \\ &= \prod_{i=1}^{m} \text{Supp}(x_{i,k_i}) \frac{P(\{s, x_{1,k_1}, \ldots, x_{m,k_m}\})}{P(x_{1,k_1}, \ldots, x_{m,k_m})} \end{aligned} \tag{E.6}$$

In our specific application to green bond spreads, since we analyze one spread category at a time, this reduces to:

$$\begin{aligned}
P(x_{1,k_1}, \ldots, x_{m,k_m}, s) &= \prod_{i=1}^{m} P(x_{i,k_i}) P(s \mid x_{1,k_1}, \ldots, x_{m,k_m}) \\
&= \prod_{i=1}^{m} \text{Supp}(x_{i,k_i}) \frac{P(\{s, x_{1,k_1}, \ldots, x_{m,k_m}\})}{P(x_{1,k_1}, \ldots, x_{m,k_m})} \\
&= \prod_{i=1}^{m} \text{Supp}(x_{i,k_i}) \frac{\text{Supp}(\{s, x_{1,k_1}, \ldots, x_{m,k_m}\})}{\text{Supp}(x_{1,k_1}, \ldots, x_{m,k_m})}.
\end{aligned} \tag{E.7}$$

From this specialized framework, we can derive the Bayesian Confidence (BC) of a rule. For a rule involving specific categories of attributes leading to a spread outcome:

$$\begin{aligned}
\text{BC}(\{x_{1,k_1}, \ldots, x_{m,k_m}\} \to s) &= P(s \mid x_{1,k_1}, \ldots, x_{m,k_m}) \\
&= \frac{P(x_{1,k_1}, \ldots, x_{m,k_m}, s)}{P(x_{1,k_1}, \ldots, x_{m,k_m})}
\end{aligned} \tag{E.8}$$

We can incorporate a penalty for the rule's length (total number of categories involved, denoted as L):

$$\text{BC}_L(\{x_{1,k_1}, \ldots, x_{m,k_m}\} \to s) = \left(\frac{P(x_{1,k_1}, \ldots, x_{m,k_m}, s)}{P(x_{1,k_1}, \ldots, x_{m,k_m})} \right)^{|L|} \tag{E.9}$$

To illustrate this framework, consider an association rule mapping from $\mathbf{X_A}$ to $\mathbf{s}$, where $\mathbf{X_A}$ contains specific one-hot encoded categories:

$$\begin{aligned}
\mathbf{X_A} : \{&x_{tax,1} = 1 \text{ (tax-exempt)},\ x_{price,2} = 1 \text{ (at-par)}, \\
&x_{call,1} = 1 \text{ (callable)},\ x_{cpn,2} = 1 \text{ (3-5\%)}\} \to \mathbf{s} = 1
\end{aligned}$$

The BC for this specific rule with length penalty is:

$$\text{BC}_L(\mathbf{X_A} \to \mathbf{s}) = \left(\frac{P(\mathbf{X_A}, \mathbf{s})}{\text{Supp}(\mathbf{X_A})} \right)^{|L|} \tag{E.10}$$

where $\mathbf{X_A}$ represents our specific combination of one-hot encoded categories.

The Bayesian Lift (BL) extends this framework by comparing conditional and marginal probabilities:

$$\begin{aligned}
\text{BL}(\mathbf{X_A} \to \mathbf{s}) &= \frac{\text{BC}(\mathbf{X_A} \to \mathbf{s})}{P(\mathbf{X_A})P(\mathbf{s})} = \frac{P(\mathbf{X_A}, \mathbf{s})}{P(\mathbf{X_A})P(\mathbf{s})} \\
&= \frac{P(\{x_{tax,1}, x_{price,2}, x_{call,1}, x_{cpn,2}, \mathbf{s}\})}{\prod P(x_{i,k_i})P(\mathbf{s})} \in (0, \infty)
\end{aligned} \tag{E.11}$$

The interpretation of BL provides crucial insights into the strength and nature of associations between one-hot encoded categories in $\mathbf{X_A}$ and spread outcomes $\mathbf{s}$:

- When BL $= 1$, we have $\frac{P(\mathbf{s}|\mathbf{X_A})}{P(\mathbf{s})} = 1$, equivalent to $\frac{P(\mathbf{X_A},\mathbf{s})}{P(\mathbf{X_A})P(\mathbf{s})} = 1$. This implies $P(\mathbf{X_A}, \mathbf{s}) = P(\mathbf{X_A})P(\mathbf{s})$, indicating independence between the one-hot encoded categories and the spread outcome.

- When BL > 1, we have $\frac{P(\mathbf{s}|\mathbf{X_A})}{P(\mathbf{s})} > 1$, which implies $P(\mathbf{X_A}, \mathbf{s}) > P(\mathbf{X_A})P(\mathbf{s})$. This suggests that $\mathbf{s}$ positively depends on $\mathbf{X_A}$, i.e., the specific combination of one-hot encoded categories is positively associated with the spread outcome.

- When BL < 1, we have $\frac{P(\mathbf{s}|\mathbf{X_A})}{P(\mathbf{s})} < 1$, which means $P(\mathbf{X_A}, \mathbf{s}) < P(\mathbf{X_A})P(\mathbf{s})$. This implies that $\mathbf{s}$ negatively depends on $\mathbf{X_A}$, i.e., the specific combination of one-hot encoded categories is negatively associated with the spread outcome (Tian et al., 2013).

This specialized framework, focusing on single spread outcomes, provides a powerful tool for analyzing the relationships between combinations of bond characteristics (represented through one-hot encoded categories in $\mathbf{X_A}$) and spread behavior in green bond markets. The BL metric, in particular, offers clear interpretation of the strength and direction of these relationships, facilitating both theoretical understanding and practical application in bond market analysis.

Appendix E.3. Bayesian Model Selection and Connection to Model Formulation

To identify the most relevant association rules, we use **Bayesian Model Selection**, which ranks candidate models by evaluating their posterior probabilities. Consider a set of candidate rule models $\{\mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}}\}_{k\in\mathcal{K}}$, where each model $\mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}}$ contains a specific subset of association rules linking one-hot encoded bond attribute categories to spread outcomes. Here, $k \in \mathcal{K}$ indexes the space of all possible models, with each k representing a unique configuration of rules.

The posterior probability of a model given observed data D can be written as:

$$P(\mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}} \mid D) = \frac{p(D \mid \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}})P(\mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}})}{\sum_{\ell\in\mathcal{K}} p(D \mid \mathcal{M}_\ell^{\mathbf{X_A}\to\mathbf{s}})P(\mathcal{M}_\ell^{\mathbf{X_A}\to\mathbf{s}})}, \tag{E.12}$$

where the marginal likelihood $p(D \mid \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}})$ can be computed as a sum (since we have discrete rules) over all possible rule configurations θ_k in model k:

$$p(D \mid \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}}) = \sum_{\theta_k} p(D \mid \theta_k, \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}})p(\theta_k \mid \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}}) \tag{E.13}$$

Here, θ_k represents a specific configuration of rules within model k, and the likelihood $p(D \mid \theta_k, \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}})$ can be computed using the Bayesian Support and Confidence measures defined earlier:

$$p(D \mid \theta_k, \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}}) = \prod_{r\in\theta_k} \mathrm{Supp}_{\mathrm{Bayes}}(r) \cdot \mathrm{Conf}_{\mathrm{Bayes}}(r) \tag{E.14}$$

where r represents individual rules in configuration θ_k.

The prior probability $p(\theta_k \mid \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}})$ reflects our beliefs about rule configurations within a model. For categorical variables in our green bond context, these priors can be informed by domain knowledge:

$$p(\theta_k \mid \mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}}) = \prod_{x_{i,k_i}\in\theta_k} P(x_{i,k_i}) \tag{E.15}$$

where $P(x_{i,k_i})$ represents our prior belief about the importance of category k_i of attribute i. For example, if tax-exempt status (category $k = 1$ of the tax attribute) is believed to strongly influence spread behavior, we might assign a higher prior probability to $P(x_{tax,1})$.

The model prior $P(\mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}})$ can be chosen to reflect preferences about model complexity. A common choice is:

$$P(\mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}}) \propto \exp(-\alpha|\theta_k|) \tag{E.16}$$

where $|\theta_k|$ is the number of rules in configuration θ_k and $\alpha > 0$ is a complexity penalty parameter. This exponential penalty implements Occam's Razor by assigning higher prior probabilities to configurations with fewer rules, thereby favouring parsimonious explanations unless the data strongly supports more complex combinations of attribute categories.

Thus, our complete posterior probability can be written as:

$$\begin{aligned} P(\mathcal{M}_k^{\mathbf{X_A}\to\mathbf{s}} \mid D) \propto \sum_{\theta_k} &\left[\prod_{r\in\theta_k} \mathrm{Supp}_{\mathrm{Bayes}}(r) \cdot \mathrm{Conf}_{\mathrm{Bayes}}(r)\right] \\ &\times \left[\prod_{x_{i,k_i}\in\theta_k} P(x_{i,k_i})\right] \exp(-\alpha|\theta_k|) \end{aligned} \tag{E.17}$$

This framework balances three key components in model selection. First, the predictive power of rules is captured through their Bayesian Support and Confidence measures, reflecting how well each rule explains the observed relationships between one-hot encoded categories and spread outcomes. Second, domain expertise is incorporated through the prior probabilities of individual categories $P(x_{i,k_i})$, allowing us to weight certain categories more heavily based on market knowledge. For instance, we might assign different prior probabilities to different coupon range categories ($P(x_{cpn,1})$ for 0-3%, $P(x_{cpn,2})$ for 3-5%, etc.) based on their expected importance in determining spread behavior. Third, the complexity penalty ensures that additional rules are only included when their explanatory power justifies the increased model complexity.

In practice, for our green bond analysis, this Bayesian model selection framework enables us to identify the most relevant combinations of one-hot encoded categories for predicting spread outcomes. The selected models represent those configurations of rules that best explain the observed data while maintaining a balance between model complexity and incorporation of market expertise through informative priors on bond characteristics. This approach is particularly well-suited to our categorical bond attributes, as it naturally handles the discrete nature of our one-hot encoded representation while incorporating domain knowledge about the relative importance of different attribute categories in determining spread behavior.

Appendix E.4. Labeling and attributes categories

The following table shows the details of the labeling process for each attribute.

Table E.1: Labeling process and details

Attribute	**Labelling Approach**	**Labels**
SPREAD MEDIAN	Sign of the annual median spread for each bond	S+, S-, S0
EXTREME SPREAD	lower quantile of the annual spread distribution as the extreme negative, and the upper quantile as the extreme positive	$S(\text{extreme}^{-})$, $S(\text{extreme}^{+})$
MUNI_TAX_PROV	No revision was applied.	Tax: AMT/ST TAX-EXEMPT, FED & ST TAX-EXEMPT, FED BQ/ST TAX-E, FED TAX-EXEMPT, FED TAXABLE, FED TAXABLE/ST TAX-EXEMPT, FED TAXABLE/ST TAXABLE
SELF_REPTD_GREEN_INSTR_INDCTR	"NA" was labelled as NA.	Self Rep. Gr: NA, YES
ISSUER_BULK	No revision was applied.	Issuer Name: Not included in the table
BB_COMPOSITE	"NA" was labelled as NA.	BB_rating: A, A-, A+, AA, AA-, AA+, AAA, BBB, BBB+, NA
MUNI_LONG_INDUSTRY_TYP	No revision was applied.	Issuer Sector: Airport, Bond Bank, Development, Education, General, General Obligation, Higher Education, Housing, Medical, Mello-Roos, Multifamily Hsg, Nursing Homes, Pollution, Power, School District, Transportation, Utilities, Water
CALLABLE	False: Non-Callable, True: Callable	Call: Callable, Non-callable
MARKET_ISSUE	No revision was applied.	Market: CERTIFICATE PARTICIPATION, GENERAL OBLIGATION LTD, GENERAL OBLIGATION UNLTD, REVENUE BONDS, SPECIAL ASSESSMENT, SPECIAL TAX, TAX ALLOCATION
MUNI_PURPOSE	No revision was applied.	UOP: ADVANCE REFUNDING, CIVIC CONVENTION CENTER, CURRENT REFUNDING, ELEC. LT. & PWR. IMPTS., GREEN PURPOSE, HLTH-HOSP- NURSHOME IMPS, MISC. PURPOSES, NURSING HOMES, PARKING FACILITY IMPS., PRT- AIRPRT & MARINA IMPS, PUBLIC FACILITIES, PUBLIC IMPS., RECREATIONAL FAC. IMPS., REFUNDING BONDS, REFUNDING NOTES, REPAYMENT OF BANK LOAN, RESOURCE RECOVERY IMPS., SCHOOL IMPS., SEWER IMPS., STATE MF HSG, STUDENT HOUSING, TRANSIT IMPS., UNIV. & COLLEGE IMPS., WATER UTILITY IMPS.
FINANCING_TYPE	No revision was applied.	FIN. Type: NEW MONEY, REFUNDING & NEW MONEY
CPN_Group	Labelled based on quartiles	CPN: 0-3%, 3-5%, 5-8.5%
SPREAD_AT_ISSUANCE_TO_WORST	Labelled based on quartiles	S_OID: less than 17, 17-57, 57-100, Higher than 100, NA
DUR_ADJ_MID	Labelled based on quartiles	DU_Adj_MID: less than 2.8, 2.8-4.8, 4.8-7, Higher than 7, NA
YIELD_ON_ISSUE_DATE	Labelled based on quartiles	Y_OID: less than 1.7, 1.7-2.4, 2.4-3.2, Higher than 3.2, NA
PRICING TYPE	Labelled based on the price compared to 100	Pricing TYP: At Discount, At Par, At Premium, NA
AMT_ISSUED	Transformed into log scale and labelled based on the quantiles	Issued Amt: less than 14, 14-15, 15-16, higher than 16, NA
MUNI_ISSUE_SIZE	Transformed into log scale and labelled based on the quantiles	Muni Issued Size: 17-18.7, 18.7-19.5, higher than 19.5, less than 17
MTY-YEARS	Labelled based on quantiles	Maturity OID: less than 8, 12.3-17, 8-12.3, higher than 17
REMAINING YEARS TO MATURITY	The year fraction between the calculation date and the maturity date of the bonds is computed and labeled according to the distribution quantiles of the calculated year fraction.	R Ys to Maturity : Less than 4.7, 4.7_9.4, 9.4_14.4, Higher than 14.4
ACTIVE YEARS	The year fraction between the calculation date and the issue date of the bonds is computed and labeled according to the distribution quantiles of the calculated year fraction.	Active years: Less than 1.1, 1.1_2.3, 2.3_4.1, Higher than 4.1

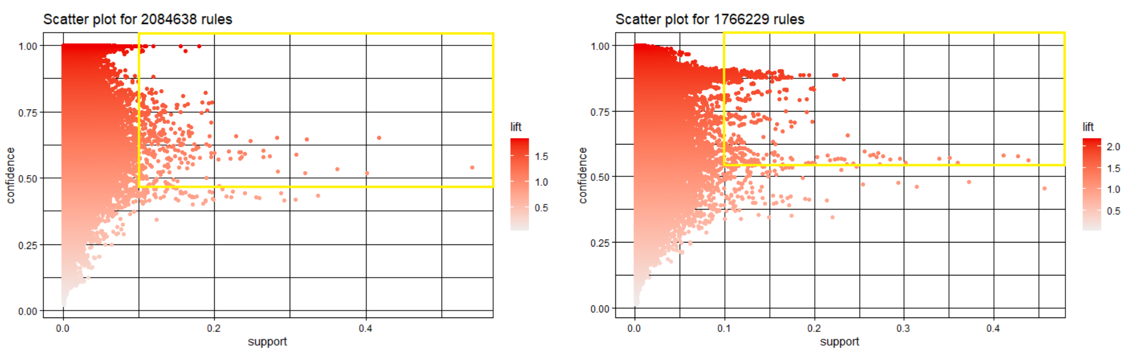
Appendix E.5. Labeling process, parameters setting and model order selection

To apply ARL we perform the following preparation steps. First, we categorize the attributes (defined in Appendix A) according to their characteristics and assign appropriate labels through a systematic labeling process for categorical and numerical attributes, as detailed in Table E.1. The next step involves determining suitable thresholds that generate possible rules for positive and negative spreads. These thresholds are based on the first confidence quantiles of all rules in our itemsets, allowing us to discover rules with the highest confidence level within the most frequent ones in the dataset, which we refer to as strong rules.

The process of generating general rules and selecting parameter thresholds is detailed in next subsection. We set a min_supp of 0.1 for both positive and negative spreads and a min_conf thresholds of 0.56 for positive rules, and 0.45 and 0.49 for negative rules, respectively. After implementing the corresponding thresholds for positive and negative spreads, we obtained subsets of rules for positive and negative spreads, spanning orders 2 to 5. We use version 1.7-7 of the arules package in R, which includes the Apriori, Eclat, and FP-Growth algorithms for mining association rules and frequent itemsets. More specifically, we used the Apriori algorithm in this research (Hahsler et al., 2023). Within the context of the *arules* package in R, "order 2" signifies rules with one item in the antecedent (lhs) and one item in the consequent (rhs).

Appendix E.6. Threshold setting for the Apriori Algorithm

We initially set the parameters quite low, at 0.001 for both support and confidence, to generate all possible rules. The scatterplots below depict these rules at the specified parameter levels for positive and negative spreads.



(a) All generated rules for positive spread using S_1.

(b) All generated rules for negative spread using S_1.

Figure E.2: **All generated rules.**

Incorporating a 10 percent threshold for support, we adopt a dynamic approach that utilizes the first quantile of the confidence distribution for positive and negative spreads. This consistent approach across spreads helps address the different characteristics of them. This adjustment ensures the identification of a suitable number of rules for both positive and negative scenarios, enhancing the robustness of our findings.

Table E.2 shows the descriptive statistics for confidence levels.

Table E.2: Summary statistics for all generated confidence levels

Steps	Descriptive Statistics					
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
S(+)	0.02	0.49	0.65	0.69	1.00	1.00
S(-)	0.01	0.53	0.73	0.71	0.89	1.00

Appendix F. ANOVA method and results

To verify the statistical significance of the ARL results (order 2), we perform a one-way ANOVA test on the frequently identified attributes. To define subsamples for the ANOVA test, we categorize according to maturity (Maturity_OID), which is a common frequent attribute previously identified.[8] We then investigate whether the contributions of the levels of other strong attributes to the spread variation are statistically significant.[9].

We find that the null hypothesis for all ANOVA tests for the top seven attributes can be rejected in favor of the alternative hypothesis (H1). More specifically, the contribution of tax status and at-par pricing type remains relatively stable across maturities. However, the contribution of the low coupon range to spread variation is more pronounced in short maturities and decreases as maturities increase. The S_OID and Y_OID play a greater role in spread variation in short maturities, whereas bonds with higher duration contribute more to spread variation in longer maturities, and vice versa. Furthermore, for longer-term bonds, the callability contributes to the spread variation. The issued amount only contributes to the variance in spreads at higher maturities, and the null hypothesis of the ANOVA test cannot be rejected for the first category of spreads.

Appendix F.1. Samples

In this step, we take Maturity_OID to separate green bonds in different groups (indexed by i) in order to investigate the variance contribution of identified attributes (indexed by

[8] We categorized green bonds based on their maturity since this characteristic affects the Investor's decision to manage convexity risk in fixed income investments. Also, maturity impacts the liquidity, with longer-term bonds experiencing higher demand due to reduced trading activity and investor preference for longer-term investments.

[9] Since we categorized green bonds according to Maturity_OID, we didn't include the other maturity-related attribute, R_Ys_to_Maturity, in the ANOVA test.

j) to the median of spreads within maturity categories. We have four maturity categories: *Less than 8*, *8_12.3*, *12.3_17*, and *Higher than 17*, and eight attributes: *MUNI_TAX_PROV*, *Pricing_TYP*, *CPN_range*, *S_OID*, *Y_OID*, *DAM*, *CALLABLE*, and *Issued_Amt*.

For each combination of maturity category and attribute, we conduct an ANOVA test to determine if there is a statistically significant difference contributed by attribute categories to each maturity group. Our sub-samples consist of bonds belonging to category k of attribute j within maturity group i, and we define their mean as μ_k^{ij}, where $k = 1, 2, \ldots, c$ and c is the number of categories within attribute j.

Appendix F.2. Hypotheses

We define the null and alternative hypotheses for the ANOVA test as follows:

$$H_0 : \mu_1^{ij} = \mu_2^{ij} = \ldots = \mu_c^{ij} \quad \text{for each pair of } i \text{ and } j$$

$$H_1 : \text{At least one pair of means is not equal for each pair of } i \text{ and } j$$

The following table summarizes the p-values obtained from the test, indicating the contribution of attributes to spread variation within each specific maturity group.

Attributes	Maturity: Less than 8 ys	Maturity: 8_12.3 ys	Maturity: 12.3_17 ys	Maturity: Higher than 17 ys
MUNI_TAX_PROV	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***
Pricing_TYP	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***
CPN_range	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***
S_OID	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***
Y_OID	1.61e-249***	< 2.2e-16***	< 2.2e-16***	< 2.2e-16***
DAM	8.83e-7***	8.79e-63***	< 2.2e-16***	< 2.2e-16***
CALLABLE	5.36e-56***	5.91e-55***	< 2.2e-16***	< 2.2e-16***
Issued_Amt	0.093·	0.001**	0.048*	1.515e-09***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix F.3. Boxplot for spread variations

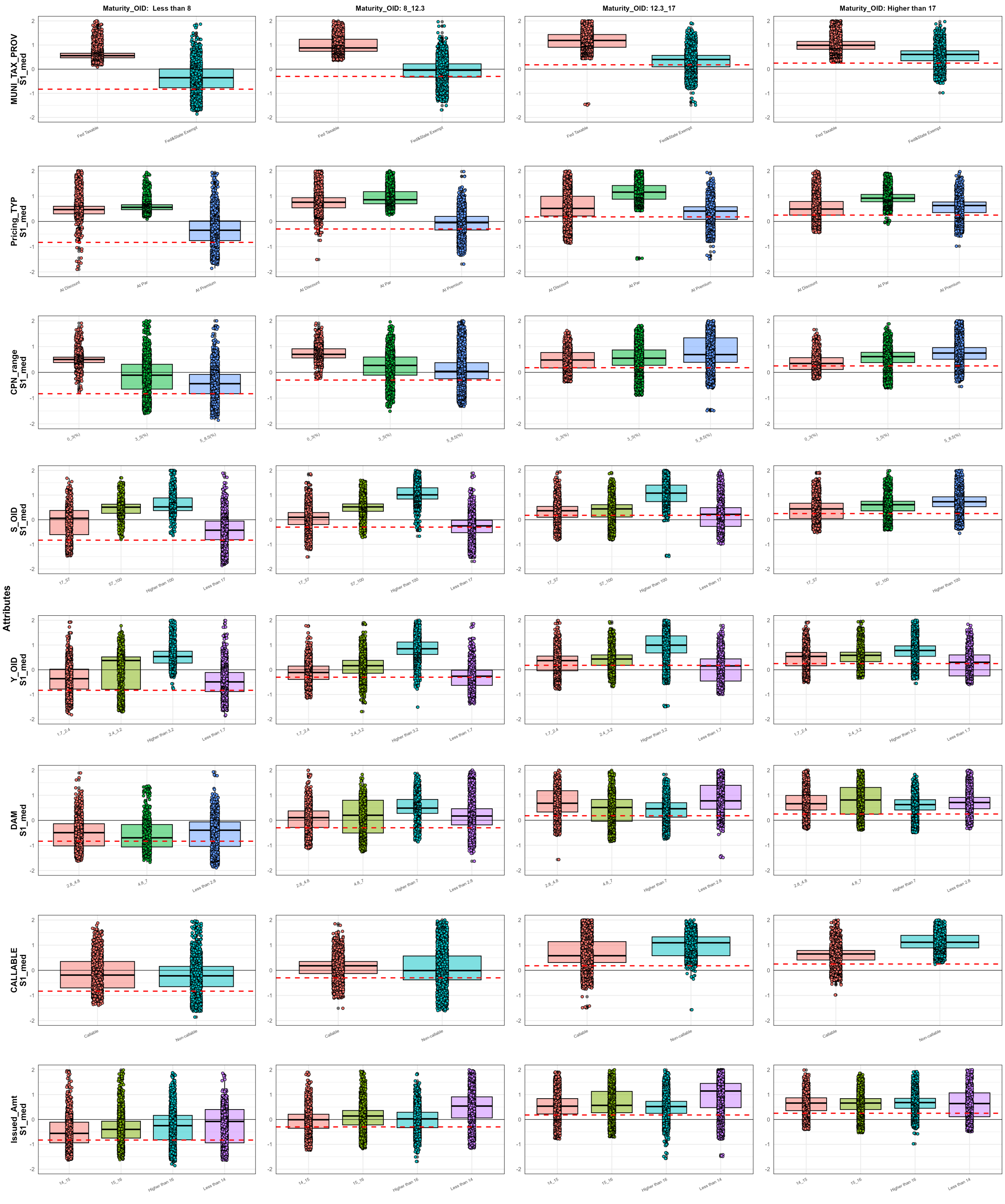


Figure F.1: Box plots for spread variation based on attributes

We find that the null hypothesis for all ANOVA tests for the top seven attributes can be rejected in favor of the alternative hypothesis (H1). More specifically, the contribution of tax status and at-par pricing type remains relatively stable across maturities. However, the contribution of the low coupon range to spread variation is more pronounced in short maturities and decreases as maturities increase. The S_OID and Y_OID play a greater role in spread variation in short maturities, whereas bonds with higher duration contribute more to spread variation in longer maturities, and vice versa. Furthermore, for longer-term bonds, the callability contributes to the spread variation. The issued amount only contributes to the variance in spreads at higher maturities, and the null hypothesis of the ANOVA test cannot be rejected for the first three maturity categories.

Appendix G. Higher-order and nested rules

This appendix presents the higher-order and nested association rules identified by the ARL procedure, providing supplementary evidence on multi-attribute interactions underlying the spread dynamics and portfolio groupings examined in the paper. Nested rules build on previously identified strong associations by incrementally adding attribute states to form higher-order rules, enabling a comprehensive assessment of multi-attribute interactions in green bond pricing.

Table G.1: The support, confidence, and lift of nested higher-order attribute associations of positive and negative green bond yield spreads.

LHS	Support	Confidence	Lift
Positive Spread ($S(+)$)			
Second-order			
Market : REVENUE BONDS, Tax : FED TAXABLE/ST TAX-EXEMPT	0.106	0.990	2.101
Self Rep. Gr. : YES, Tax : FED TAXABLE/ST TAX-EXEMPT	0.112	0.982	2.084
Market : REVENUE BONDS, Pricing TYP : At Par	0.102	0.970	2.058
Pricing TYP : At Par, Tax : FED TAXABLE/ST TAX-EXEMPT	0.143	0.968	2.053
Pricing TYP : At Par, Self Rep. Gr. : YES	0.108	0.959	2.035
Third-order			
Pricing TYP : At Par, Self Rep. Gr. : YES, Tax : FED TAXABLE/ST TAX-EXEMPT	0.104	0.981	2.081
Call : Callable, S OID : Higher than 100, Self Rep. Gr. : YES	0.101	0.667	1.415
Negative Spread ($S(-)$)			
Second-order			
R Ys to Maturity : Less than 4.7, Tax : FED & ST TAX-EXEMPT	0.165	0.876	1.679
Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7	0.170	0.876	1.678
CPN : 5_8.5(%), R Ys to Maturity : Less than 4.7	0.119	0.873	1.672
R Ys to Maturity : Less than 4.7, S OID : Less than 17	0.115	0.863	1.652
Maturity OID : Less than 8, Tax : FED & ST TAX-EXEMPT	0.165	0.856	1.639
Third-order			
Call : Non-callable, CPN : 5_8.5(%), R Ys to Maturity : Less than 4.7	0.109	0.891	1.707
Call : Non-callable, R Ys to Maturity : Less than 4.7, Tax : FED & ST TAX-EXEMPT	0.151	0.888	1.701
Call : Non-callable, Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7	0.157	0.888	1.700
R Ys to Maturity : Less than 4.7, Self Rep. Gr. : YES, Tax : FED & ST TAX-EXEMPT	0.140	0.882	1.690
Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7, Self Rep. Gr. : YES	0.144	0.882	1.689
Fourth-order			
Call : Non-callable, R Ys to Maturity : Less than 4.7, Self Rep. Gr. : YES, Tax : FED & ST TAX-EXEMPT	0.128	0.891	1.707
Call : Non-callable, Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7, Self Rep. Gr. : YES	0.133	0.891	1.706
Call : Non-callable, CPN : 5_8.5(%), Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7	0.104	0.890	1.704
Call : Non-callable, CPN : 5_8.5(%), R Ys to Maturity : Less than 4.7, Tax : FED & ST TAX-EXEMPT	0.102	0.889	1.703
Call : Non-callable, Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7, Tax : FED & ST TAX-EXEMPT	0.149	0.888	1.700

Table G.2: The support, confidence and lift of higher-order attribute associations of positive and negative green bond yield spreads.

LHS	Support	Confidence	Lift
Positive Spread ($S(+)$)			
Third-order			
Pricing TYP : At Par, Self Rep. Gr. : YES, Tax : FED TAXABLE/ST TAX-EXEMPT	0.104	0.981	2.081
Call : Callable, R Ys to Maturity : Higher than 14.4, Self Rep. Gr. : YES	0.130	0.667	1.420
Call : Callable, Market : REVENUE BONDS, R Ys to Maturity : Higher than 14.4	0.107	0.667	1.420
Fourth-order			
Call : Callable, Maturity OID : Higher than 17, R Ys to Maturity : Higher than 14.4, Self Rep. Gr. : YES	0.111	0.650	1.380
Call : Callable, Pricing TYP : At Premium, R Ys to Maturity : Higher than 14.4, Self Rep. Gr. : YES	0.101	0.650	1.390
Call : Callable, R Ys to Maturity : Higher than 14.4, Self Rep. Gr. : YES, Tax : FED & ST TAX-EXEMPT	0.120	0.640	1.370
Fifth-order			
Call : Callable, CPN : 5_8.5(%), Market : REVENUE BONDS, Pricing TYP : At Premium, Tax : FED & ST TAX-EXEMPT	0.122	0.530	1.120
Call : Callable, CPN : 5_8.5(%), Market : REVENUE BONDS, Self Rep. Gr. : YES, Tax : FED & ST TAX-EXEMPT	0.108	0.520	1.110
Call : Callable, CPN : 5_8.5(%), Market : REVENUE BONDS, Pricing TYP : At Premium, Self Rep. Gr. : YES	0.109	0.520	1.110
Negative Spread ($S(-)$)			
Third-order			
Call : Non-callable, CPN : 5_8.5(%), R Ys to Maturity : Less than 4.7	0.109	0.891	1.707
Call : Non-callable, R Ys to Maturity : Less than 4.7, Tax : FED & ST TAX-EXEMPT	0.151	0.888	1.701
Call : Non-callable, Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7	0.157	0.888	1.700
Fourth-order			
Call : Non-callable, R Ys to Maturity : Less than 4.7, Self Rep. Gr. : YES, Tax : FED & ST TAX-EXEMPT	0.128	0.891	1.707
Call : Non-callable, Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7, Self Rep. Gr. : YES	0.133	0.891	1.706
Call : Non-callable, CPN : 5_8.5(%), Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7	0.104	0.890	1.704
Fifth-order			
Call : Non-callable, R Ys to Maturity : Less than 4.7, Self Rep. Gr. : YES, Tax : FED & ST TAX-EXEMPT, Pricing TYP : At Premium	0.128	0.890	1.700
Call : Non-callable, Market : REVENUE BONDS, Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7, Tax : FED & ST TAX-EXEMPT	0.100	0.880	1.690
Call : Non-callable, Maturity OID : Less than 8, Pricing TYP : At Premium, R Ys to Maturity : Less than 4.7, Self Rep. Gr. : YES	0.110	0.880	1.680

Appendix H. Portfolio Implementation and Curve-Based Rate Extraction

This appendix documents implementation details related to the extraction of discount rates from fitted green bond yield curves and the numerical treatment of noisy or unstable yield observations for the portfolio extraction. These aspects are embedded in the empirical code but are not discussed explicitly in the main text.

Appendix H.1. Extraction of term-matched green discount rates

For each green bond observation and each synthetic bond used in portfolio construction, discount rates are extracted from the fitted green yield curve at maturities corresponding to the remaining cash-flow dates of the bond. Specifically, for a bond with remaining cash-flow times $\{\tau_{i,1}, \ldots, \tau_{i,N}\}$, the green discount rates $r_t^{(G)}(\tau_{i,n})$ are obtained by evaluating the fitted partition-specific yield curve at each $\tau_{i,n}$.

Because the fitted green curves are represented as linear B-splines, rate extraction is performed by identifying the spline segment covering each $\tau_{i,n}$ and evaluating the corresponding linear basis function. No interpolation across partitions is performed; each bond is discounted exclusively using the yield curve associated with its structural partition.

When a cash-flow maturity coincides with a knot point, rates are evaluated using the left-continuous spline segment to preserve temporal consistency across trading days and to avoid artificial jumps in extracted rates due to time-varying knot placement.

Appendix H.2. Construction of equivalent zero-coupon yields

The extracted term-specific green rates are used to discount each coupon and principal cash flow of the bond. An equivalent zero-coupon yield is then computed as the constant yield that equates the discounted value of all remaining cash flows to the bond's face value over its remaining life.

This transformation ensures that coupon-bearing green bonds and zero-coupon Treasury benchmarks are evaluated on a fully comparable basis. Equivalent zero-coupon yields are recomputed daily, allowing changes in curve-level pricing to propagate smoothly into yield spreads and portfolio valuations.

Appendix H.3. Treatment of noisy green yield observations

Green municipal bond yields exhibit substantial noise due to episodic trading, dealer-specific pricing, and infrequent quote updates. Several numerical safeguards are implemented to mitigate the impact of such noise on curve estimation and rate extraction.

First, raw yield observations are filtered using a two-step Z-score-based outlier detection procedure prior to curve fitting, as described in Appendix B. This procedure removes extreme yield realizations that are unlikely to reflect market-clearing prices.

Second, curve estimation is only performed when a minimum number of observations is available within a partition. On days with insufficient data, the most recently estimated curve for that partition is carried forward unchanged. This prevents unstable refitting driven by transient or anomalous observations.

Third, discount rates are extracted only within the maturity domain supported by observed data in the corresponding partition. No extrapolation beyond the convex hull of observed maturities is performed, avoiding amplification of noise at the short and long ends of the curve.

Appendix H.4. Stability of extracted rates over time

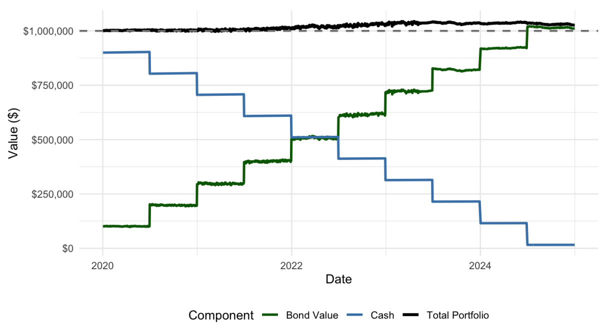
Because discount rates are extracted from a piecewise linear functional representation and curve updates occur only when sufficient data are available, the resulting rate series evolve smoothly over time even when individual bond yields are sporadically observed. Consequently, changes in extracted green discount rates reflect systematic shifts in curve-level pricing rather than mechanical responses to isolated noisy observations.

The extracted rates therefore provide a stable and economically meaningful input for the computation of green bond yield spreads and for portfolio valuation in illiquid municipal bond markets.

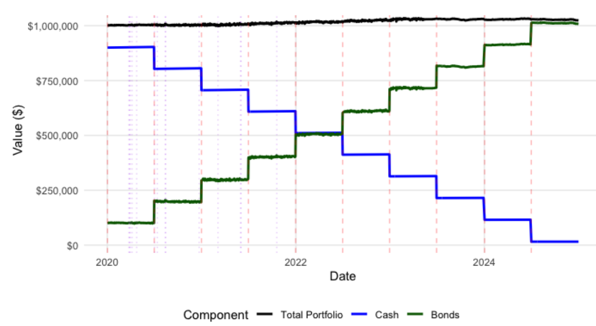
Appendix H.5. Portfolio composition and allocation dynamics

Figure H.1 illustrates the allocation dynamics for Group 1 under both portfolio construction approaches. The plots show the evolution of bond holdings (green), cash allocation (blue), and total portfolio value (black) over the 2020–2024 period with semi-annual rebalancing.

The duration-targeted strategy (Panel a) exhibits greater variation in bond holdings due to dynamic maturity selection to maintain the 9–11 year duration target, while the tenor-matched approach (Panel b) shows more regular patterns from fixed 10-year maturity purchases. Cash reserves decline systematically as capital is deployed into green bonds at each rebalancing date. The stable total portfolio value throughout the sample period, particularly during the 2022–2023 tightening cycle, demonstrates the downside protection provided by the screened green bond portfolios.



(a) Duration-targeted portfolio



(b) Tenor-matched portfolio

Figure H.1: Portfolio composition dynamics for Group 1 screened green bonds (2020–2024). Panel (a) shows the duration-targeted strategy maintaining 9–11 year duration exposure, while panel (b) displays the tenor-matched approach with fixed 10-year maturity purchases. Bond holdings (green), cash allocation (blue), and total portfolio value (black) are tracked under semi-annual rebalancing with 10% wealth allocation per purchase.

References

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases, in: Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pp. 207–216. doi:10.1145/170035.170072.

Cestau, D., Hollifield, B., Li, D., Schürhoff, N., 2019. Municipal bond markets. Annual Review of Financial Economics 11, 65–84. doi:10.1146/annurev-financial-110118-123034.

Chen, Z., Mao, C.X., Wang, Y., 2010. Why firms issue callable bonds: Hedging investment uncertainty. Journal of Corporate Finance 16, 588–607. doi:10.1016/j.jcorpfin.2010.06.008.

Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364. doi:10.1016/j.jeconom.2005.03.005.

Hahsler, M., Buchta, C., Gruen, B., Hornik, K., 2023. arules: Mining Association Rules and Frequent Itemsets. R package version 1.7-7 https://CRAN.R-project.org/package=arules.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. volume 2. Springer.

Perlovsky, I., DeMarco, T., 2018. Overview of the taxable municipal market. Boston: Fidelity Capital Markets. https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/fixed-income/retail-taxable-muni-white-paper-2018.pdf.

Tian, D., Gledson, A., Antoniades, A., Aristodimou, A., Ntalaperas, D., Sahay, R., Pan, J., Stivaros, S., Nenadic, G., Zeng, X.j., et al., 2013. A bayesian association rule mining algorithm, in: 2013 IEEE international conference on systems, man, and cybernetics, IEEE. pp. 3258–3264. doi:10.1109/SMC.2013.555.

Yallup, P.J., 2012. Models of the yield curve and the curvature of the implied forward rate function. Journal of Banking & Finance 36, 121–135. doi:10.1016/j.jbankfin.2011.06.010.