

Term Structure with Investor Tax Heterogeneity: Application to the Muni Puzzle *

Kevin Zheyuan Cui[†]

September 12, 2026

Abstract

The municipal puzzle is that short municipal yields match a top-bracket household's after-tax taxable yield, while long yields far exceed this representative-household benchmark. I develop an equilibrium model and provide evidence that investor tax heterogeneity is a first-order force shaping the yield curve and explains this puzzle. Investors choose taxable and tax-exempt bonds jointly across maturities under no-short-sale constraints. High-tax households value exemption most and anchor short municipal yields. Lower-tax institutions participate farther out when liability-hedging benefits offset their tax disadvantage. Because institutions value exemption less, equilibrium long municipal yields exceed the household benchmark. Household portfolio substitution carries a diminishing premium to shorter maturities, while institutional reallocation affects taxable yields. The calibrated model accounts for 86% of the roughly 110-basis-point thirty-year gap. Two empirical designs support the predicted maturity pattern and contradict the representative-tax benchmark: household-specific tax pass-through declines with maturity, while capitalization of an institution-specific tax benefit increases. Policy counterfactuals show that exemption repeal shifts investors across markets, raising municipal financing costs by more than the added tax receipts and increasing taxable borrowing costs.

Keywords: Term Structure, Asset Ownership, Investment Taxes, Municipal Bonds, Insurance Companies, Segmented Markets

JEL Codes: G12, G14, G22, H24, H63

*I am grateful to Moritz Lenel, Wei Xiong, and Motohiro Yogo for their guidance. I also thank Yacine Ait-Sahalia, Markus Brunnermeier, Fousseni Chabi-Yo, Aditya Chaudhry, Daniel Chen, Zhiyu (Julie) Fu, Nicolae Garleanu, Sergei Glebkin, Zhouzhou Gu, Nicolas Hommel, Paymon Khorrami, Jane Li, Shiqi Li, Jonathan Payne, Yinan Qiu, Julia Selgrad, Amy Huber Wang, Lingxuan (Sean) Wu, Yinjie Yu, Yao Zeng, William (Yinhong) Zhao, and Haoxing Zhu, and participants at MPSS Conference and Princeton University Seminars for constructive feedback.

[†]Princeton University, Department of Economics.

1 Introduction

How do taxes shape the yield curve when investors face different tax rates? Consider two otherwise comparable bonds, one taxable and one tax-exempt. The standard benchmark uses the tax rate of a top-bracket household, which benefits most from exemption, to equate after-tax yields at every maturity: $y_T^M = (1 - \tau_H)y_T^C$ (Chalmers, 1998; Longstaff, 2011; Schwert, 2017; Fleckenstein and Longstaff, 2025). Yet households, insurers, banks, and tax-exempt institutions value the same exemption differently under the U.S. tax code. This paper asks how this neglected tax heterogeneity determines relative yields along the curve.

The U.S. municipal market presents a sharp departure from the representative-tax benchmark. Short-term tax-exempt municipal yields closely match the top-bracket household's after-tax corporate yield, but long-term municipal yields are substantially higher. Even after removing callability and the issuance premium, the gap reaches about 110 basis points at thirty years. This widening yield gap is the municipal puzzle, a pattern visible in data dating to the 1950s (Feenberg and Poterba, 1991).¹ The same puzzle can be expressed as a decline with maturity in the yield-implied tax, $\hat{\tau}_T = 1 - y_T^M/y_T^C$. The implied tax equals the household's rate when municipal yields match the benchmark and falls below it when they exceed it. Accordingly, the adjusted curves imply a tax rate of about 39% at two years but only 23% at thirty, against a household benchmark near 41%. With more than \$4 trillion outstanding by 2020, against \$9.8 trillion for corporate bonds, the market finances state and local public investment, making the long-end gap consequential for the cost of funding long-lived public projects.

This paper develops a term-structure model in which investor tax heterogeneity shapes participation and relative bond prices. It builds on two observations: the tax code gives investors different benefits from the same exemption, and institutional investors and lower-tax households hold municipals alongside high-tax households. Investors choose taxable and tax-exempt bonds jointly across maturities under no-short constraints, prices clear both markets, and the short rate is the only persistent state. Tax treatment changes where investors participate; their portfolios connect prices across maturities and markets. This investor-side mechanism complements bond-side explanations based on default, liquidity, and taxable-bond tax timing, and predicts whose tax changes affect which maturities.

The mechanism rests on a tradeoff between tax advantage and duration benefit. A two-investor illustration makes the mechanism transparent: high-tax households and lower-tax

¹The main taxable benchmark is the high-quality corporate curve, which is closer to municipals in credit and liquidity than treasuries. Using corporates also avoids attributing the treasury convenience premium to the municipal tax wedge. Section 4 defines the common rating band and bond-side adjustments.

insurers represent the principal municipal clienteles. High-tax households receive the largest benefit from municipals, so they anchor the fully riskless one-period muni bond. Longer bonds expose their holders to changes in resale prices when interest rates move: duration risk. Lower-tax investors such as insurers can absorb this duration more cheaply because long-duration assets hedge their liabilities. The insurer’s tax disadvantage in holding municipals is the same across maturities, while its duration benefit grows. A lower-tax institution therefore participates in the municipal market when the value of the duration it absorbs offsets the tax advantage it forgoes.

Participation changes required returns through two channels. At jointly held maturities, returns must accommodate investors who receive less benefit from the exemption, reducing the tax advantage reflected in municipal prices. Their long-end participation also changes households’ total duration exposure. Because households choose one portfolio across maturities, that exposure affects the compensation they require on shorter risky bonds as well. An institution’s tax valuation can therefore affect a bond it never holds. The portfolio spillover diminishes toward the short end and vanishes at the fully riskless one-period bond. The equilibrium price recursion carries both channels into yields: the municipal curve starts at the household benchmark and rises relative to it as maturity lengthens.

The analytical benchmark also separates why investors participate from how their participation affects prices. Liability duration is a natural, empirically relevant motive for insurer participation, but the two pricing channels do not depend on it: any holding benefit that does not vary with the rate state sustains the same participation pattern and leaves both channels intact, so the participation of lower-tax investors steepens the municipal curve relative to the household benchmark under the model’s maturity-ordering conditions even when the motive is not liability hedging.²

The quantitative model extends this mechanism to 45 state-level municipal markets, resident households, and three institutional clienteles, capturing the exemption residents receive only on their own state’s bonds. Equilibrium prices are nonlinear because participation can change with the short-rate state. Tax rules, market quantities, sector balance-sheet structure, rate dynamics, and taxable-yield targets discipline the calibration; municipal yields and the empirical test coefficients are untargeted. The model accounts for 94.8 of the 109.7 basis points remaining at thirty years after callability and the issuance premium are removed, or 86% of the long-end puzzle. Its yield-implied tax correspondingly declines from about 42%

²The conditional result allows additional state-independent, per-unit holding benefits while keeping the participating investors, covariance structure, financing rule, and constraints fixed. Section 5.7 and Appendix Proposition 9 establish the unchanged transmission equations; the yield conclusion uses the forward-rate conditions of Result 3 (Section 5.6).

at the short end to 27% at thirty years. In the paper’s common spread accounting, adding the separately estimated default, liquidity, and capital-gains differentials of 18.4 basis points leaves a -3.5 -basis-point residual. The model’s contribution grows with maturity relative to these bond-side adjustments, precisely where the puzzle is largest.

The empirical tests ask whether different investors’ tax benefits are reflected at different maturities. The model predicts that the household benefit is fully reflected at the short boundary and matters less farther out, while an institutional benefit matters primarily at longer maturities. The representative-tax benchmark instead assigns unit pass-through to the household at every maturity and no role to other investors’ taxes. State income taxes provide household-specific variation because residents generally receive an exemption on in-state municipal interest. The estimated pass-through into the yield-implied state tax declines from 0.98 below five years to 0.13 beyond twenty. The nonlinear calibration independently produces a cross-state slope of 1.00 at two years and 0.25 at thirty, close to the estimates.

A statutory discontinuity supplies the complementary institutional test. The bank-qualified rule grants banks a tax benefit on designated bonds of issuers below an annual issuance cap. Around that cap, the designation premium rises from 1.8 basis points at one to two years to 18.4 at ten to twenty years. A benefit available only to banks thus matters farther along the curve, where household tax pass-through has weakened. Together, the two designs connect the maturity pattern to the identity of the investor receiving the tax benefit, as the model predicts.

Two additional designs examine the same mechanism. The 2017 corporate rate cut reduced property and casualty (P&C) insurers’ municipal tax advantage. States with greater pre-reform P&C exposure show a decline in implied tax, with maturity-gradient point estimates in the predicted direction. Within-bond transaction pairs also show that municipal yields respond more to treasury yield movements at longer maturities. Both patterns are consistent with the heterogeneous-tax model and contrast with the representative-tax benchmark, which predicts no effect of P&C insurers’ tax treatment on municipal yields and a constant municipal-to-treasury yield response across maturities.

The model’s policy relevance extends beyond the scale of the municipal market: tax policies change frequently and often target a particular investor group. Federal personal and corporate rate changes, revisions to asset-specific exemptions and deductions, and state tax reforms repeatedly alter the relative value of taxable and tax-exempt claims. Their effects depend on which investors are affected, which maturities they help finance, and how their participation changes in equilibrium. Because investors choose across both markets, a tax change directed at municipal interest can also change corporate borrowing costs.

Two counterfactuals show why tax accounting must include the equilibrium response of

investor participation. Repealing an in-state exemption recovers tax revenue but weakens the local investor base. Residents reduce precisely the home-state positions on which the new tax is collected, while the higher yield needed to attract replacement investors affects all new and refinanced borrowing. Across the eight state experiments, annualized own-state municipal replacement-financing costs rise by \$2.50–\$2.78 per dollar of state investor-side tax receipts. Moving toward equal municipal and corporate tax treatment raises issuer replacement-financing costs by \$1.99 per dollar of added tax receipts, including \$0.73 for corporate borrowers through investor substitution. Applying a new tax rate to the old allocation would miss both the changing tax base and these cross-market financing effects.

Municipal bonds offer a setting with large, observable tax differences and a longstanding relative-pricing puzzle. The heterogeneous-tax model and its underlying logic also extend to other markets where investor-specific taxes interact with portfolio risk and limits to arbitrage: treasuries and corporates under state taxation, taxable and tax-favored accounts, and cross-border coupon withholding. Tax policy can change who bears risk along the yield curve and transmit financing effects beyond the asset directly affected.

Section 2 reviews the literature. Sections 3–4 present the data, institutions, and bond-side adjustments; Section 5 develops the model; Sections 6–7 report calibration and tests; and Sections 8–9 present counterfactuals and conclusions. Proofs and additional results are in the appendix.

2 Related Literature

This paper is closest to the literature on the muni puzzle. Before the Tax Reform Act of 1986, banks, which could deduct carrying costs, received the largest municipal benefit, and early habitat accounts attributed the declining implied tax to banks holding short bonds and households holding long bonds (Mussa and Kormendi, 1979; Hendershott and Koch, 1980; Kidwell and Koch, 1983; Skelton, 1983). The arbitrage tradition instead obtains one marginal tax rate when the transactions enforcing parity are freely available (Miller, 1977; Fama, 1977; Livingston, 1982).

The 1986 Act removed the bank advantage outside a small bank-qualified segment, and banks largely withdrew; the implied-tax gradient persisted. The literature then turned to bond characteristics while retaining a representative household wedge: taxable-bond timing (Green, 1993), default (Chalmers, 1998), and default, liquidity, and related features (Longstaff, 2011; Schwert, 2017; Fleckenstein and Longstaff, 2025). These adjustments are important but do not close the maturity pattern. Appendix A.8 discusses Green’s single-rate benchmark, which changes the taxable comparison portfolio rather than the municipal bond.

Among the bond-side studies, [Schwert \(2017\)](#) addresses yield levels in a non-callable sample concentrated below ten years, whereas the puzzle is sharpest farther out.

This paper supplies an investor-side explanation with endogenous participation and equilibrium term-structure implications. Unlike the early habitat accounts, it does not assign investors to maturities: tax advantages and portfolio risk jointly determine participation, and the empirical tests identify whose taxes matter along the curve. The mechanism also reconciles the gradient’s persistence across tax regimes: the short end reflects investors with the greatest exemption benefit, while other participating investors affect longer maturities, even when the identity of the highest-benefit clientele changes.

The theory builds on preferred-habitat and segmented-market models ([Vayanos and Vila, 2021](#); [Greenwood et al., 2023](#)), replacing primitive maturity habitats with participation governed by tax policy, risk, and no-short constraints. It connects to maturity clienteles ([Pirinsky and Wang, 2011](#); [Brown and Sirmans, 2011](#)), heterogeneous municipal tax exposure ([Babina et al., 2021](#)), limited arbitrage ([Hines and Horton, 2024](#); [Gromb and Vayanos, 2018](#); [Siriwardane et al., 2022](#)), and ownership-based asset pricing ([Kojen and Yogo, 2019](#); [Coppola, 2021](#); [Li, 2026](#)). For the bank-qualified design, [Dagostino \(2025\)](#) establishes effects on municipal yields, issuance, and employment; this paper isolates the maturity profile of the bank-qualified yield premium and relates it to heterogeneous-tax participation. [Babina et al. \(2021\)](#) show how heterogeneous taxes segment municipal markets geographically and limit risk sharing. Here the central object is the maturity profile of participation and joint municipal–corporate pricing.

3 Data and Measurement

3.1 Sources and curve construction

The main window is 2000–2020. Mergent supplies municipal and corporate issuance, call schedules, and ratings; MSRB and TRACE Enhanced supply transactions. Holdings come from NAIC Schedule D, a FOIA-collected pension panel, and the Flow of Funds; IRS Statistics of Income supplies tax-exempt interest income by household income bracket. The taxable curve is FRED’s A-or-above corporate zero curve; the GSW treasury curve provides the rate benchmark. Markit CDS quotes and Tax Foundation state schedules supply credit and tax inputs. Appendix [A.10](#) documents the yield and exposure construction, and Appendix [A.5](#) gives the adjustment constructions.

Table 1: Muni issuance: key descriptive statistics (sample means).

Variable	Full sample	Main sample	
		AA and above (score ≤ 3)	A— and above (score ≤ 7)
Offering yield (%)	3.11	2.82	2.96
Corporate benchmark yield (%)	4.47	4.33	4.41
Maturity at issuance (years)	10.15	10.76	10.86
Offering price	102.88	104.68	103.85
Issue rating (1 = AAA, 3 = AA, 7 = A—)	3.80	2.18	3.55
Implied tax rate $1 - y^M/y^C$	0.29	0.35	0.33
Share callable	0.46	0.51	0.51
Share tax-exempt	0.92	1.00	1.00
Sample size (CUSIPs)	3,014,420	984,210	1,990,791

Notes. Sample means; one observation per municipal issuance (CUSIP). The full sample contains all issues. The AA-and-above and A—-and-above columns restrict to federally tax-exempt, non-AMT issues with average agency score ≤ 3 and ≤ 7 , respectively; the latter matches the A-or-above corporate credit band. The implied-tax row averages the bond-level $1 - y^M/y^C$ ratio (winsorized at the first and ninety-ninth percentiles), so it need not equal one minus the ratio of the two displayed yield means. Each row uses non-missing observations; the rating mean in the full sample uses the 2,292,092 rated issues, while unrated issues are excluded from the two restricted columns.

3.2 Descriptive statistics: muni issuance

Table 1 compares all municipal issues with the federally tax-exempt, non-AMT samples used below. The A—-and-above sample matches the corporate credit band; the tighter AA-and-above sample limits credit heterogeneity in the transaction tests. Ratings use the unrounded mean of available agency scores. About half of the restricted samples are callable, motivating the bond-level adjustment.

4 Institutional Background and Motivating Facts

This section establishes the fact the model must explain. It first summarizes the tax wedges and ownership patterns that make several clienteles relevant, then shows that the municipal-taxable gap is large, widens with maturity, and survives the standard bond-side adjustments.

4.1 A primer on bond taxation and the textbook benchmark

Taxable coupons are ordinary income, whereas federally tax-exempt municipal coupons are excluded from the federal tax base. For institutions, proration can make part of municipal

interest taxable as well. Investor i 's treatment is therefore summarized by a municipal-equivalent corporate keep

$$\xi_i \equiv \frac{1 - \tau_i^C}{1 - \tau_i^M}, \quad \tau_i \equiv 1 - \xi_i. \quad (1)$$

This reduces to $\xi_i = 1 - \tau_i^C$ for a fully exempt household. A lower ξ_i means a larger municipal advantage. With one frictionless marginal investor, after-tax parity implies $y_T^M = \xi_i y_T^C$ at every maturity, so the yield-implied tax $1 - y_T^M/y_T^C$ is flat. The textbook benchmark assigns that role to the top-bracket household and therefore compares municipals with $\xi_H y_T^C$. With heterogeneous keeps, no one yield ratio can make every investor indifferent; the model determines which clienteles participate at each maturity. Appendix A.9 gives the cash-flow tax rules in full.

4.2 Heterogeneous investor taxes and ownership

Municipal ownership spans exemption benefits both across investor types and within households. Table 2 reports $1 - \xi_i$ by investor type: high-tax households benefit most; P&C insurers benefit more than life insurers and allocate about 40% of their corporate-and-municipal bond holdings to tax-exempt municipals, compared with 7–8% for life insurers. These insurer shares use corporate and municipal bonds as the denominator; they exclude other bond classes.

Figure 1 shows the heterogeneity along both dimensions. The left panel reports municipal holdings by sector, with approximate exemption benefits in the legend. The right reports household tax-exempt interest by income bracket. Institutions and households below the highest bracket hold substantial municipal positions despite their smaller tax benefits, so the market is not held only by top-rate households. Appendix A.10.2 documents these series, and Appendix A.11 adds insurer and pension holdings evidence; the quantitative economy retains the clienteles separately.

4.3 Limits to arbitrage

Heterogeneous wedges matter only if arbitrage is limited. That condition is natural here: interest expense used to carry tax-exempt bonds is generally not deductible, lending a municipal bond can extinguish the lender's exemption, and shorting is difficult in a fragmented market. The highest-benefit investors therefore cannot lever without limit to impose their own parity condition on every maturity. The model represents these restrictions with no-short constraints, allowing participation and the marginal tax wedge to vary along the curve.

Table 2: Investor heterogeneity in the municipal market

Investor types: muni tax benefit and muni share of bond holdings.

Investor type	Federal tax wedge $1 - \xi_i$	State exemption	Tax-exempt muni share of bond holdings
High-tax household	0.35–0.43	only on in-state munis	n/a
Lower-tax household	0.22–0.30	only on in-state munis	n/a
P&C insurer (pre-2017 Act)	0.31	none	40%
Life insurer	0.15–0.21	none	7–8%
Other (pension, etc.)	≤ 0.05	none	$\approx 0\%$

Notes. Federal tax wedge $1 - \xi_i$ from statutory rates and proration formulas. The last column is the average tax-exempt muni share, 2010–2017. For insurers, the denominator is their corporate-and-municipal bond portfolio. Before the 2017 Act the ranking (P&C $\gg$ life insurer $\gg$ other institutional) follows the federal wedge; the Act compresses the P&C–life-insurer gap.

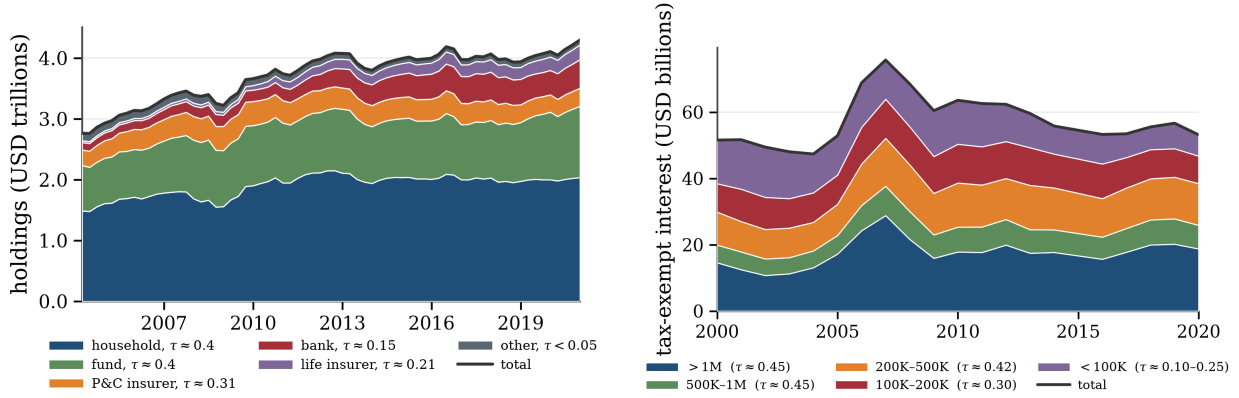


Figure 1: Ownership heterogeneity across and within investor types. Left panel: the stock of municipal holdings by broad investor group, 2004–2020 (USD trillions; Federal Reserve Financial Accounts). Right panel: annual tax-exempt interest reported on taxable U.S. individual income-tax returns by adjusted-gross-income bracket, 2000–2020 (USD billions; IRS Statistics of Income). The legend values $\tau = 1 - \xi$ summarize each group’s approximate benefit from exemption; higher values indicate a larger tax advantage. The panels show ownership across differently taxed institutions and household income groups, rather than exclusively among top-rate households. The right panel measures interest income, not the stock of bond holdings; 2015 is omitted because comparable income-bracket detail is unavailable.

Alt text: Two panels show municipal holdings by investor sector and household tax-exempt interest by income group. Ownership spans institutions and households with different exemption benefits, including households below the highest income groups.

Appendix A.9 gives the institutional details and the historical bank comparison.

4.4 The muni puzzle and what the literature has tried

The municipal puzzle is a widening gap relative to the top-household after-tax corporate benchmark, $\xi_H y_T^C$: short municipal yields closely match this benchmark, whereas long municipal yields substantially exceed it. The yield-implied tax reports the same comparison in tax-rate units:

$$ImpliedTax_{s,t}(T) \equiv 1 - \frac{y_{s,t}^M(T)}{y_t^C(T)}. \quad (2)$$

For a given household benchmark, $\tau_H = 1 - \xi_H$, the two measures satisfy

$$\underbrace{y_{s,t}^M(T) - \xi_H y_t^C(T)}_{\text{municipal yield gap}} = \underbrace{[\tau_H - ImpliedTax_{s,t}(T)]}_{\text{shortfall from the household tax rate}} y_t^C(T).$$

With positive corporate yields, a match to the household benchmark means the implied tax equals the household rate, and a positive yield gap means it falls below. Figure 2 presents both views. Panel A shows a municipal yield near the household benchmark at the short end but about 110 basis points above it at thirty years. Panel B expresses the same relative yields as an implied tax falling from about 39% to 23%, while the household wedge remains near 41%.

Both legs use the A-or-above credit band. The municipal leg is built from offering yields, which provide broad coverage and avoid the secondary-market capital-gains wedge; the corporate leg is the seasoned, option-free A-or-above zero curve. The municipal curve is therefore adjusted for its embedded call and its issuance premium before the two are compared. A corporate benchmark is preferable to treasuries because it is closer in credit and liquidity, but the same declining pattern appears against treasuries and within issuer-matched tax-exempt and taxable municipals. The AA-and-above and unlimited-tax general-obligation samples reproduce the shape (Appendix A.7.1 and A.6.3).

The long-end gap is material for the financing of long-lived public capital and is a relative-pricing puzzle spanning both municipal and taxable debt. Default, liquidity, and taxable-bond tax shelters are the leading existing explanations; the next subsection measures the bond-side margins directly.

4.5 The puzzle survives the bond-side adjustments

Five margins form the accounting: callability, the issuance premium, default, liquidity, and capital-gains taxation. Each is measured on the same spread,

$$Spread(T) = y_T^M - \bar{\xi}_H(T) y_T^C, \quad (3)$$

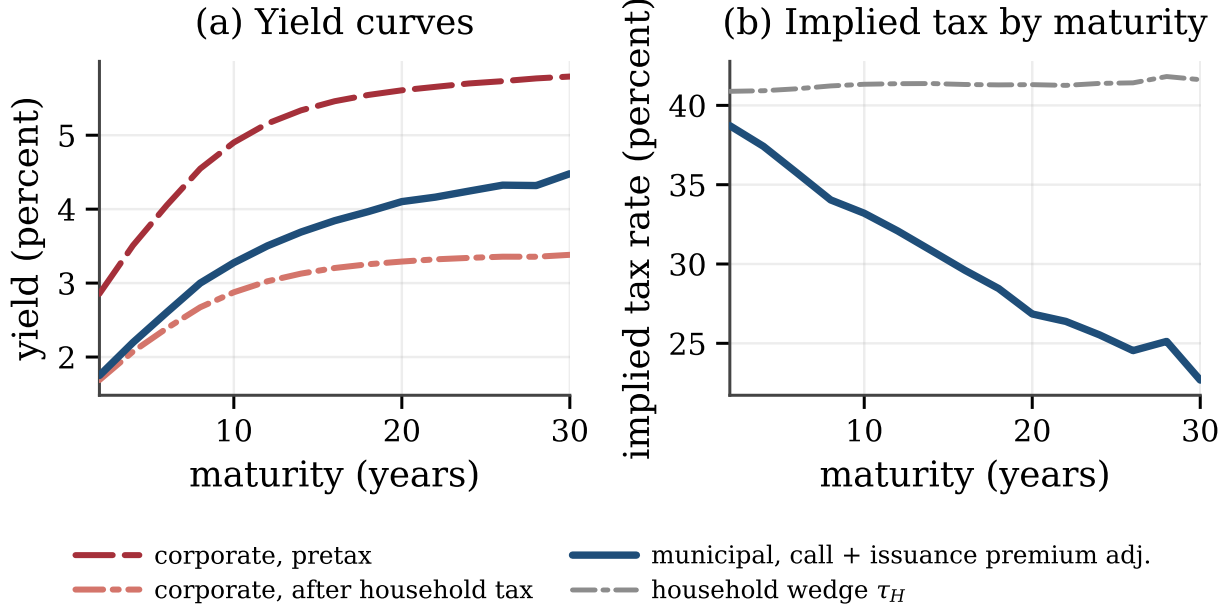


Figure 2: The muni puzzle after the call option and issuance premium are removed. Panel A compares the municipal curve with the pretax corporate curve and the after-household-tax benchmark $\xi_H y_T^C$. Panel B reports $1 - y_T^M / y_T^C$. The observed implied-tax decline is 23.4 percentage points from two to thirty years; the two curve-level adjustments remove 7.4 points and leave the 16.0-point decline shown here. A-or-above zero-coupon curves, 2000–2020. Appendix A.7.1 reports the AA-and-above counterpart.

Alt text: The left panel compares municipal and corporate yield curves after removing municipal callability and issuance premiums. Short municipal yields approach the household after-tax benchmark, but exceed it farther out. The right panel shows the corresponding decline in yield-implied tax.

where $\bar{\xi}_H(T)$ applies each bond’s household tax rate before the same aggregation used for the pretax curve. The effective household wedge averages 41.3% and is nearly flat.

Callability and the issuance premium change the municipal curve itself. Removing them reduces the thirty-year spread from 159.8 to 109.7 basis points. Default, liquidity, and capital-gains treatment are then municipal-minus-corporate differentials that explain part of the remaining spread. Together they contribute 18.4 basis points, or 17% at thirty years, compared with 52% at ten. The remaining spread therefore grows with maturity rather than disappearing. Appendix Table 15 gives the complete ledger; Section 6.3 combines it with the model.

The call adjustment uses bond-level backward induction on contractual semiannual coupon and call calendars, with the non-callable sample disciplining volatility. Dropping callable bonds would discard more than 95% of observations beyond twenty years. The issuance premium compares offering yields with the same bonds’ first-month secondary

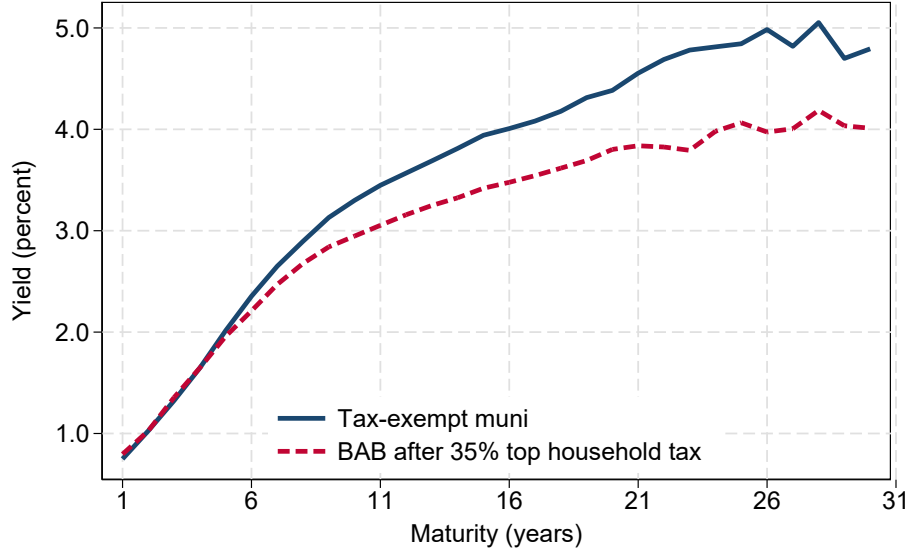


Figure 3: Tax-exempt and Build America Bond offering yields, April 2009–December 2010. BAB yields are multiplied by 0.65, the top-bracket household keep. The curves nearly coincide at short maturities and separate as maturity lengthens. Table 8 reports issuer-by-maturity-by-quarter regressions with and without call-status matching; Appendix A.6.2 gives the sample construction.

Alt text: Average tax-exempt and tax-normalized Build America Bond offering yields are close at short maturities. Tax-exempt yields increasingly exceed the taxable-bond after-tax benchmark as maturity lengthens.

trades, net of market drift; it rises from 7.9 basis points at ten years to 21.1 at thirty. Appendix A.5 gives all five estimators and sensitivity envelopes. Their sum is a first-order accounting of separately estimated contributions.

An issuer-matched comparison confirms the pattern without a corporate benchmark. During 2009–2010, issuers sold both tax-exempt bonds and taxable Build America Bonds (BABs). Figure 3 shows their average offering-yield curves, multiplying BAB yields by 0.65 to apply the top-household keep. The regressions compare bonds within issuer-by-maturity-by-quarter cells. The tax-exempt gap rises by 3.4 basis points per year in the AA-and-above sample and 3.3 in the A-or-above sample; matching call status gives 3.1 and 3.0. Thus the widening gap survives matching on borrower, maturity, timing, and callability. Appendix A.6.2 gives the specification; Table 8 reports the estimates.

5 A Term-Structure Model with Heterogeneous Taxes

Investors with different tax treatment choose municipal and corporate bonds jointly under no-short constraints. Households value the exemption most and anchor the riskless short

bond. Institutions' liabilities let them absorb duration more cheaply, which supports their participation in longer municipal bonds. Participation and required returns are determined together in market clearing. Returns reflect participation directly through the tax mix of holders and indirectly through households' portfolio risk at shorter maturities.

Three economic questions organize the analysis. Why would a less tax-advantaged investor hold municipals? The three-maturity illustration and Result 1 derive the participation incentive from portfolio choice. How do relative yields reflect that participation? Results 2–3 and Section 5.7 derive the direct effect and the spillover in a two-investor affine benchmark. How large are the effects? The full nonlinear economy in Section 6 solves participation and both yield curves jointly, using calibrated balance sheets. It does not impose the analytical restrictions that Section 5.3 states.

5.1 The mechanism first: two investors, three maturities

A three-maturity economy isolates the mechanism and keeps both bond markets. Households H and insurers I jointly choose municipal bonds at $T = 1, 2, 3$ and risky taxable bonds at $T = 2, 3$, all in positive fixed supplies Q_T^a . The one-period taxable investment is cash. Investors cannot short. Cash supplies the opportunity-cost benchmark and imposes no separate wealth constraint. Group i has mass m_i , common risk aversion γ , and relative taxable keep $\xi_H < \xi_I$, so households value the exemption more.

The only state is the short rate, with $z' = \rho z + \sigma_r \varepsilon$, $\varepsilon \sim \mathcal{N}(0, 1)$, and $r(z) = \bar{r} + z$. The illustration uses affine continuation prices on a positive-rate region where the proposed holdings and KKT inequalities are feasible. The full model below restores the floor and state-dependent participation.

Write x_{iT}^a for holdings, μ_T^a for the log expected gross holding return, and $A_{T-1}^a = -\partial_z p_{T-1}^a$ for the rate loading of the bond's next-period log resale price. The bond's return innovation is then $-\sigma_r A_{T-1}^a \varepsilon$. The first claim pays a fixed amount, so $A_0^a = 0$ and the claim is fully riskless. Longer claims carry more rate exposure on the branch studied here. Equilibrium prices determine these loadings, and the last step of the example recovers them.

Joint portfolio choice and KKT conditions. Normalize the municipal keep to one. Taxes then multiply both the mean and the risk of taxable returns by ξ_i . A positive liability loading ℓ_i offsets asset duration, and a negative loading represents outside assets that add

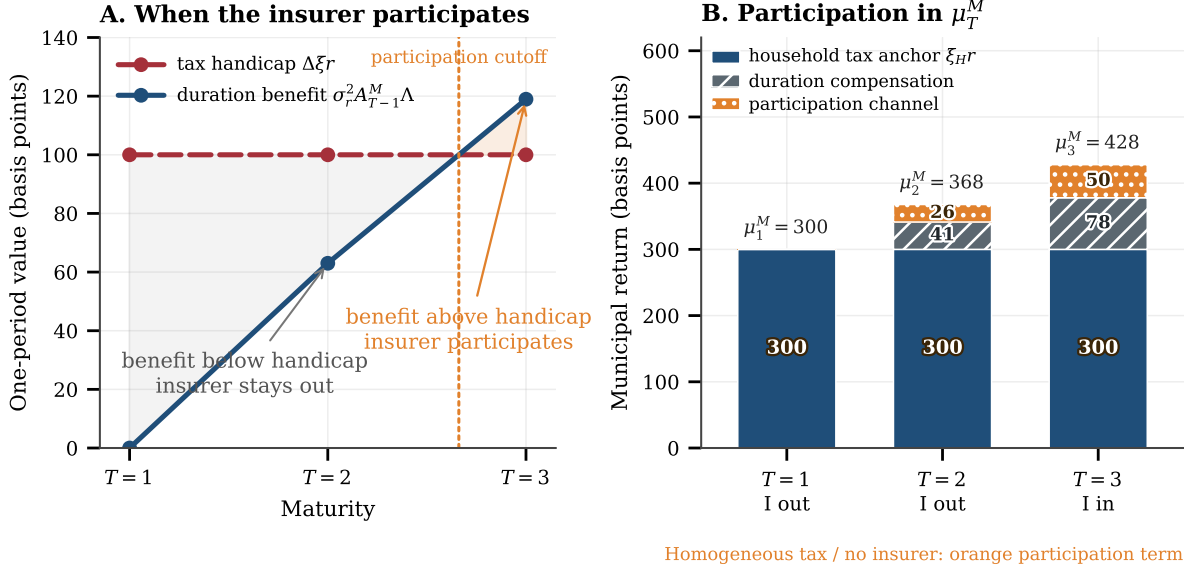


Figure 4: Participation and required returns in the three-maturity economy. Panel A compares the insurer's flat tax penalty with its increasing duration benefit. Panel B separates the household tax anchor, aggregate-duration compensation, and participation: the orange component is a spillover at $T = 2$ and a direct tax-mix effect at $T = 3$. The orange terms vanish under homogeneous taxes or without insurer municipal participation. Values are per-period basis points at the feasible allocation in Appendix A.1.8.

Alt text: The participation panel compares a constant insurer tax disadvantage with a duration benefit that rises across three maturities. The return panel separates the household benchmark, the direct effect at the jointly held long bond, and its spillover to the shorter risky bond.

duration. Investor i solves

$$\max_{x_{iT}^M, x_{iT}^C \geq 0} \sum_{T=1}^3 x_{iT}^M (\mu_T^M - \xi_i r) + \xi_i \sum_{T=2}^3 x_{iT}^C (\mu_T^C - r) - \frac{\gamma \sigma_r^2}{2} \widehat{D}_i^2, \quad (4)$$

where

$$\underbrace{\widehat{D}_i}_{\text{net whole-book rate exposure}} = \underbrace{\sum_{T=1}^3 A_{T-1}^M x_{iT}^M + \xi_i \sum_{T=2}^3 A_{T-1}^C x_{iT}^C}_{=: D_i, \text{ the asset side}} - \underbrace{\ell_i}_{\text{liability hedge}}.$$

A long municipal position, a taxable position, and the liability hedge all change the risk cost of the same portfolio. Differentiating, and dividing the corporate condition by $\xi_i > 0$, gives

$$\underbrace{\mu_T^M - \xi_i r}_{\text{municipal excess return}} \leq \underbrace{\gamma \sigma_r^2 A_{T-1}^M \widehat{D}_i}_{\text{marginal portfolio risk}}, \quad (5)$$

$$\underbrace{\mu_T^C - r}_{\text{taxable excess return}} \leq \gamma \sigma_r^2 A_{T-1}^C \widehat{D}_i,$$

with equality whenever the corresponding holding is positive. Market clearing requires $\sum_i m_i x_{iT}^a = Q_T^a$ in every market. Define the KKT slack ν_{iT}^a as the right side minus the left side. It is zero for a positive holding and nonnegative otherwise.

Participation: subtract the two investors' conditions. At $T = 1$ the risk term vanishes. Clearing selects $\mu_1^M = \xi_H r$, at which households hold the supply and the insurer's slack is $(\xi_I - \xi_H)r > 0$. The short-end anchor therefore follows from optimization.

Now evaluate an insurer purchase at the allocation where households hold the risky munis and insurers hold the risky corporates. Superscript hh denotes this pre-participation allocation. The KKT comparison holds the candidate price loadings fixed and does not re-solve a separate no-participation equilibrium. Substituting the household's municipal equality into the insurer's slack gives

$$\nu_{IT}^{M, hh} = \underbrace{(\xi_I - \xi_H)r}_{\text{tax handicap same at every maturity}} - \underbrace{\sigma_r^2 A_{T-1}^M \Lambda}_{\text{duration benefit of participating in the municipal market}}, \quad \Lambda := \gamma(\widehat{D}_H^{hh} - \widehat{D}_I^{hh}) > 0. \quad (6)$$

Here Λ includes both asset portfolios and liability hedges, $\Lambda = \gamma(D_H^{hh} - D_I^{hh}) + \gamma(\ell_I - \ell_H)$, so a larger insurer liability hedge raises its duration benefit. The tax penalty is constant across maturities, while the benefit rises with A_{T-1}^M . Thus

$$\sigma_r^2 A_1^M \Lambda < (\xi_I - \xi_H)r < \sigma_r^2 A_2^M \Lambda \quad (7)$$

selects participation in the three-period, but not the two-period, municipal bond as a candidate. Panel A of Figure 4 draws the comparison. The duration benefit lies below the penalty at $T = 2$ and above it at $T = 3$, so participation starts from the long end. The final allocation must also satisfy clearing and the remaining KKT inequalities, which Appendix A.1.8 verifies.

The long bond: combine active KKT conditions with clearing. Let $\Delta\xi = \xi_I - \xi_H$, $\omega_i = m_i/(m_H + m_I)$, and $\xi^* = \omega_H \xi_H + \omega_I \xi_I$. Common risk aversion makes the mass weights risk-tolerance weights. At the jointly held three-period municipal bond, subtracting the insurer's equality from the household's gives

$$\underbrace{\gamma \sigma_r^2 A_2^M (\widehat{D}_H - \widehat{D}_I)}_{\text{duration benefit required by the insurer}} = \underbrace{\Delta\xi r}_{\text{insurer's tax handicap}}. \quad (8)$$

Clearing fixes the weighted net duration $\bar{D} = \omega_H \hat{D}_H + \omega_I \hat{D}_I$, whose expression in supplies and liabilities is in the appendix. With $\Delta \hat{D} = \Delta \xi r / (\gamma \sigma_r^2 A_2^M)$, the two equations imply

$$\hat{D}_H = \underbrace{\bar{D}}_{\text{market average}} + \underbrace{\omega_I \Delta \hat{D}}_{\text{extra duration borne by household}}, \quad \hat{D}_I = \bar{D} - \omega_H \Delta \hat{D}. \quad (9)$$

A weighted average of the active three-period KKT equalities then yields

$$\mu_3^M = \underbrace{\xi^* r}_{\text{active-clientele tax mix}} + \underbrace{\gamma \sigma_r^2 A_2^M \bar{D}}_{\text{compensation for aggregate duration}}. \quad (10)$$

The first term accommodates both investors' tax valuations, and the second compensates aggregate duration.

The shorter risky bond: the household carries the spillover. Only households hold the two-period municipal bond. Substituting their solved exposure into their KKT equality gives

$$\mu_2^M = \underbrace{\xi_H r}_{\text{household tax anchor}} + \underbrace{\gamma \sigma_r^2 A_1^M \bar{D}}_{\text{ordinary duration premium}} + \underbrace{\omega_I \frac{A_1^M}{A_2^M} \Delta \xi r}_{\text{spillover from insurer's } T=3 \text{ participation}}. \quad (11)$$

The last term is positive even though insurers never hold this bond. In equilibrium the $T = 3$ yield compensates the insurer's tax disadvantage, and households hold more net duration than the market average. That whole-portfolio risk raises the return households require at $T = 2$.

Read together, (10) and (11) show the two channels in one place. Since $\xi^* r = \xi_H r + \omega_I \Delta \xi r$, both required returns share the household tax anchor $\xi_H r$ and the compensation for aggregate duration. Each also carries one participation term. At $T = 3$ it is the direct tax-mix effect $\omega_I \Delta \xi r$. The insurer is one of this bond's holders, and the clearing return accommodates its tax valuation along with the household's. At $T = 2$ it is the spillover $\omega_I (A_1^M / A_2^M) \Delta \xi r$, and the household's extra net duration raises the compensation it requires on a bond the insurer never holds. Both terms vanish with equal taxes or without insurer participation. Panel B of Figure 4 draws them.

Equilibrium loadings and relative yields. The taxable side adjusts too. Insurers' active corporate KKT equations determine corporate returns, while the household's corporate slack is $\gamma \sigma_r^2 A_{T-1}^C (\hat{D}_H - \hat{D}_I) > 0$. The same allocation therefore satisfies both markets.

Each price is next period's expected resale value discounted at the required return,

$P_T^a(z) = \mathbb{E}_z[P_{T-1}^a(z')]/\exp(\mu_T^a(z))$, and the yield is $y_T^a = -\log P_T^a/T$. A higher required return at any maturity lowers today's price and raises its yield. The price of a long bond embeds the required returns of every maturity it passes through. Let $s_T^a = \partial\mu_T^a/\partial z$ denote the response of a required one-period return to the state, not a yield-curve slope. The loadings then satisfy

$$A_T^a = \rho A_{T-1}^a + s_T^a, \quad a \in \{M, C\}, \quad (12)$$

and $A_1^M = \xi_H$, $A_1^C = 1$. Differentiating the solved return at $T = 2$ gives $s_2^M = \xi_H + \omega_I(A_1^M/A_2^M)\Delta\xi$, so the recursion closes the initially unknown duration:

$$(A_2^M)^2 - \xi_H(1 + \rho)A_2^M - \xi_H\omega_I\Delta\xi = 0.$$

Appendix A.1.8 supplies the positive root, the corporate loadings, and all portfolio checks. Participation and rate exposure are thus solved jointly.

Relative to the household-after-tax corporate benchmark,

$$s_T^M - \xi_H s_T^C = \underbrace{(s_T^M - \xi_H)}_{\text{municipal leg}} + \underbrace{\xi_H(1 - s_T^C)}_{\text{taxable leg}}. \quad (13)$$

Both terms are zero at the riskless boundary and positive at the risky maturities on this branch. On the municipal leg compensation rises, and on the taxable leg the insurer's reduced exposure lowers the corporate required-return response. The recursion accumulates these responses into the loadings. In this illustration the municipal-to-corporate required-return response ratio rises from 0.600 to 0.716 to 0.825. The loading ratio, a positive weighted average of those responses, therefore rises with maturity. This relative rate sensitivity is one ingredient of the yield curve, not the yield ratio itself. Yield levels also reflect price intercepts and term premia, and Result 3 gives the conditions under which the response ordering produces the declining yield-implied tax. The illustration isolates the mechanism and leaves the quantitative fit to the full model.

5.2 The general economy: assets, investors, and equilibrium

The general economy retains joint portfolio choice and no-short constraints while allowing investor-specific taxes, bond-specific risk, and multiple municipal markets. Table 3 collects the notation.

Table 3: Notation for the model

Symbol	Meaning
$z, r(z) = \max\{\bar{r} + z, 0\}$	Latent rate state and the one-period cash return
$k_i^a, \xi_i = k_i^C / k_i^M$	Investor i 's absolute after-tax keep in market $a \in \{M, C\}$ and its relative keep; a smaller ξ_i is a larger municipal advantage
$\xi_H < \xi_I$	Household and insurer keeps in the two-investor benchmark
$p_T^a(z), A_T^a = -\partial_z p_T^a$	Log price of the maturity- T bond in market a and its rate loading, the model's duration; $A_0^a = 0$
$\mu_T^a(z) = \log \mathbb{E}_z[1 + R_T^a],$ $s_T^a = \partial_z \mu_T^a$	One-period required return appearing in the KKT conditions and its response to the state
$x_{i,a}(T, z), Q_T^a, m_i, \gamma_i$	Holdings (written x_{iT}^a in the illustration), supply, investor mass, risk aversion
$\hat{D}_i, \ell_i$	Investor i 's tax-adjusted rate exposure net of the liability loading ℓ_i
$\omega_{i,T}, \mathcal{U}_T$	Risk-tolerance share of an active holder and the active set at maturity T
$\tau_T^{Y,\text{impl}} = 1 - y_T^M / y_T^C$	Yield-implied tax, the empirical object of Sections 4 and 7

Step 1: state, assets, and tax treatment. One aggregate state drives interest rates. It follows a stationary AR(1), and the taxable cash rate is bounded below at zero:

$$z_{t+1} = \rho z_t + \sigma_r \varepsilon_{t+1}, \quad r_t = r(z_t) := \max\{\bar{r} + z_t, 0\}, \quad \rho \in (0, 1), \quad \varepsilon_{t+1} \sim \mathcal{N}(0, 1). \quad (14)$$

The floor changes cash payoffs but leaves the Gaussian state transition intact, so expectations retain the ordinary normal law.

There is one corporate market and a municipal market for each issuer state $s \in \mathcal{S}$. Assets are market-maturity pairs $a \in \mathcal{A} = \{M_s : s \in \mathcal{S}\} \cup \{C\}$, $T = 1, \dots, T_{\max}$. Home-state and out-of-state exemptions can differ for each resident household. The corporate $T = 1$ claim is cash. Its price P_1^C remains the continuation value of the two-period bond, but independently risky corporate positions begin at $T = 2$.

Let $P_{T,t}^a$ be the price of a zero-coupon bond with T periods remaining. Its one-period gross return is

$$1 + R_{T,t+1}^a = \frac{P_{T-1,t+1}^a}{P_{T,t}^a} \exp(\sigma_{\eta,T}^a \eta_{T,t+1}^a), \quad (15)$$

where the shocks $\{\eta_{T,t+1}^a\}_{a,T}$ are mutually independent standard normal variables, independent of the common rate innovation ε_{t+1} . They capture bond-specific risk that the rate factor does not span. Their loadings satisfy $\sigma_{\eta,T}^a \geq 0$ for $T \geq 2$ and $\sigma_{\eta,1}^a = 0$ in every market. A one-period claim therefore has neither rate exposure nor bond-specific risk, so it is fully riskless. Bond-specific risk at longer maturities prevents perfect substitution. Appendix A.2.4

reports numerical accuracy and sensitivity to the riskless boundary.

Investor type i has mass $m_i > 0$, risk aversion $\gamma_i > 0$, and keeps the fraction $\xi_i^a > 0$ of a pre-tax payoff in market a . In the one-municipal-market analytical benchmark, normalize the municipal keep to one and write

$$\xi_i^M = 1, \quad \xi_i^C = \xi_i := \frac{1 - \tau_i^C}{1 - \tau_i^M}.$$

A smaller ξ_i thus means a larger relative municipal tax advantage. The two-investor benchmark assumes $\xi_H < \xi_I$, so the household values the exemption more than the insurer. This normalization rescales payoff units. It does not assume that every investor pays zero municipal tax. The preference parameters below are expressed in the same units.³

Step 2: the investor's joint portfolio problem. Write $g_{T,t+1}^a := \log(1 + R_{T,t+1}^a)$, define $\mu_T^a(z_t) := \log \mathbb{E}_t[1 + R_{T,t+1}^a]$, and center the log-return innovation as $u_{T,t+1}^a := g_{T,t+1}^a - \mathbb{E}_t[g_{T,t+1}^a]$. The after-tax excess return over taxable cash is

$$\tilde{\mu}_{i,a}(T, z_t) := \underbrace{\xi_i^a \mu_T^a(z_t)}_{\text{after-tax bond return}} - \underbrace{\xi_i^C r(z_t)}_{\text{after-tax cash return}}, \quad (16)$$

so, under the one-market normalization,

$$\tilde{\mu}_{i,M} = \mu_T^M - \xi_i r, \quad \tilde{\mu}_{i,C} = \xi_i (\mu_T^C - r). \quad (17)$$

If $x_{i,a}(T, t)$ is investor i 's holding, her after-tax portfolio innovation, including a rate-sensitive liability hedge, is

$$\tilde{\Pi}_i(t+1) = \sum_{a \in \mathcal{A}} \sum_T \xi_i^a x_{i,a}(T, t) u_{T,t+1}^a + \underbrace{\ell_i \sigma_r \varepsilon_{t+1}}_{\text{liability hedge}}, \quad (18)$$

³If the absolute municipal keep used as numeraire is $k_i^M > 0$, absolute keeps and the absolute liability hedge are k_i^M times their normalized counterparts. Setting $\gamma_i = k_i^M \gamma_i^{\text{abs}}$ and $\ell_i = \ell_i^{\text{abs}} / k_i^M$ makes the absolute objective equal to k_i^M times (19), so the optimizer is unchanged. If a reform changes k_i^M , holding absolute preferences fixed requires transforming γ_i and ℓ_i accordingly. Relative keeps alone do not determine tax payments in dollars. Appendix A.9.2 states the additional requirements.

because rising rates reduce bond-like liabilities as well as asset prices. A positive ℓ_i offsets asset duration. Investor i chooses all positions jointly:

$$\max_{x_i \geq 0} \sum_{a \in \mathcal{A}} \sum_T x_{i,a}(T, t) \tilde{\mu}_{i,a}(T, z_t) - \frac{\gamma_i}{2} \text{Var}_t \left(\tilde{\Pi}_i(t+1) \right). \quad (19)$$

The risk cost depends on the whole balance sheet, so a liability hedge can support long-maturity participation despite a smaller tax benefit.

In this problem the keep multiplies $\log \mathbb{E}[1+R]$ and the log-return innovations. This yield-unit convention applies one keep to the whole holding return and does not distinguish coupons from realized resale gains. Taxing simple interest instead introduces a second-order difference of about 1.6 annual basis points at the two-year household benchmark. Appendix A.9 states the actual tax rules, and the separate capital-gains adjustment evaluates differences in timing and tax treatment.

Step 3: KKT conditions and market clearing. The no-short constraint turns this problem into one comparison for every investor, market, and maturity:

$$\begin{aligned} x_{i,a}(T, z) > 0 &\implies \tilde{\mu}_{i,a}(T, z) = \gamma_i \xi_i^a \text{Cov}_z \left(u_{T,t+1}^a, \tilde{\Pi}_i(t+1) \right), \\ x_{i,a}(T, z) = 0 &\implies \tilde{\mu}_{i,a}(T, z) \leq \gamma_i \xi_i^a \text{Cov}_z \left(u_{T,t+1}^a, \tilde{\Pi}_i(t+1) \right). \end{aligned} \quad (20)$$

Equality means compensation matches marginal portfolio risk, and under strict inequality the investor holds zero. Dividing by the positive asset keep recovers the normalized municipal and corporate comparisons of the illustration.

Finally, every market and maturity clears:

$$x_{i,a}(T, z) \geq 0, \quad \sum_i m_i x_{i,a}(T, z) = Q_T^a, \quad (21)$$

where Q_T^a is the per-capita supply of market- a bonds at maturity T . An equilibrium consists of state-contingent prices and holdings that satisfy the return identities, the KKT conditions, and clearing simultaneously. No investor is assigned to a maturity.

5.3 Solving the equilibrium: nonlinear price functions and the affine benchmark

The quantitative model solves the same portfolio and clearing conditions at every rate state. The affine benchmark restricts the rate rule and continuation prices to obtain analytical

results.

Step 4: nonlinear price functions. The quantitative object is a log-price function:

$$p_{T,t}^a = p_T^a(z_t), \quad p_0^a(z) \equiv 0, \quad a \in \mathcal{A}. \quad (22)$$

Let $Z' = \rho z + \sigma_r \varepsilon'$. The bond's exact conditional log expected gross return is

$$\mu_T^a(z) := \log \mathbb{E}_z [\exp\{p_{T-1}^a(Z')\}] + \frac{1}{2} \sigma_{\eta,T}^2 - p_T^a(z), \quad (23)$$

or, equivalently,

$$p_T^a(z) = \log \mathbb{E}_z [\exp\{p_{T-1}^a(Z')\}] + \frac{1}{2} \sigma_{\eta,T}^2 - \mu_T^a(z). \quad (24)$$

At every state, the return $\mu_T^a(z)$ must satisfy the KKT conditions in (20) and market clearing in (21). The appendix writes the stacked system and its exact covariance matrix in (58)–(59).

Three features make the resulting price function nonlinear. The rate floor kinks the cash opportunity cost. State-dependent participation changes which KKT equalities and inequalities bind as z moves. Resale makes today's price an average over future states that can lie on different sides of either boundary. Even at a state where the current active set is unchanged, the continuation value in (24) therefore incorporates future nonlinear regions. Two consequences follow. The yield curve and the implied tax depend on the rate state, so the quantitative results are read at a reporting state. The magnitudes also cannot be taken from the affine benchmark, which fixes the participation pattern, and with it the tax mix and the spillover, at every maturity. The nonlinear solution supplies the magnitudes, and the affine benchmark supplies the mechanism.

To connect prices to the mechanism, define the bond's local rate loading and residual price level by

$$A_T^a(z) := -\frac{\partial p_T^a(z)}{\partial z}, \quad C_T^a(z) := -p_T^a(z) - A_T^a(z)z. \quad (25)$$

At differentiability points,

$$y_T^a(z) = -\frac{p_T^a(z)}{T} = \frac{A_T^a(z)z + C_T^a(z)}{T}, \quad \frac{\partial y_T^a(z)}{\partial z} = \frac{A_T^a(z)}{T}.$$

The loading $A_T^a(z)$ equals maturity times the yield's rate response, and in the calibration it is positive and rises with maturity. The residual $C_T^a(z)$ records the price level and adds no state variable.

Finally, let $s_T^a(z) := \partial \mu_T^a(z) / \partial z$ be the response of the bond's required one-period return.

Differentiating (24) gives the exact accumulation rule

$$A_T^a(z) = s_T^a(z) + \rho \frac{\mathbb{E}_z[\exp\{p_{T-1}^a(Z')\} A_{T-1}^a(Z')]}{\mathbb{E}_z[\exp\{p_{T-1}^a(Z')\}]}. \quad (26)$$

Continuation loadings receive more weight in states where the bond is more valuable. The KKT conditions determine s_T^a , the recursion accumulates it into A_T^a , and the full log price divided by maturity gives the yield. Equation (26) retains the floor, participation changes, and nonlinear continuation.

5.3.1 Step 5: the analytical affine benchmark

The nonlinear solution delivers the magnitudes. To expose the mechanism in closed form, the analytical benchmark imposes two restrictions:

- (i) Replace the floored cash rate with the linear rule $r(z) = \bar{r} + z$, which removes the zero lower bound.
- (ii) Solve a fixed set of active KKT equalities at all states, using the resulting affine continuation functions, and verify the holdings and inactive inequalities on the state region under study.

Active equalities, inactive inequalities, nonnegative holdings, and clearing must hold on the stated region. Appendix A.1.2 gives the feasibility conditions.

Start from affine candidate prices at all maturities. Gaussian integration makes their log continuation values affine in z and their return covariances constant. With active sets fixed and the stacked KKT-and-clearing matrix nonsingular, the simultaneous linear system returns affine holdings and expected returns. The price recursion therefore preserves the affine class exactly, and a coefficient fixed point solves this restricted system without linearization. Its KKT inequalities identify the current-state region on which the proposed participation pattern is feasible, conditional on the global affine continuation. The nonlinear economy also integrates over future changes in the floor and in participation. Appendix A.1.2 makes this distinction precise.

The two-investor analytical economy. Results 1–3 use a household and an insurer with $\xi_H < \xi_I$ in one municipal and one corporate market. Results 1–2 suppress the affine superscript, and their holdings, loadings, and derivatives refer to this benchmark, not the unrestricted quantitative economy.

Prices now have constant coefficients,

$$p_T^{a,\text{aff}}(z) = -A_T^{a,\text{aff}} z - C_T^{a,\text{aff}}, \quad (27)$$

and investor i 's whole-portfolio rate exposure, net of the liability hedge, is

$$\hat{D}_i(z) := \sum_j x_{i,M}(j, z) A_{j-1}^M + \xi_i \sum_j x_{i,C}(j, z) A_{j-1}^C - \ell_i. \quad (28)$$

A positive liability loading reduces net duration, hence the minus sign. Because A_{j-1}^a is constant in this benchmark, $\partial \hat{D}_i(z)/\partial z$ measures only the change in exposure caused by equilibrium trading. The exact loading recursion becomes

$$A_T^{a,\text{aff}} = \rho A_{T-1}^{a,\text{aff}} + s_T^{a,\text{aff}}, \quad (29)$$

so one-period required-return responses accumulate into the full bond-price loading. Appendix equation (75) gives the intercept recursion, including the price of rate risk, convexity, and the idiosyncratic inventory term.

The riskless one-period tax anchor holds in both economies when $r(z) > 0$. Appendix A.1.2 states the affine branch's feasibility and uniqueness conditions. Appendix Figure 12 illustrates the continuation-price variation that the affine restriction omits. Quantitative magnitudes use the nonlinear solution.

5.4 Result 1: insurers participate from the long end

Begin in the two-investor affine benchmark at the household-only municipal allocation, where the household holds all municipal supply and the insurer holds no munis. The insurer holds the entire corporate book, the household holds no corporates, and both retain their liability hedges (Assumption 1). Throughout this result, $\hat{D}_H$ and $\hat{D}_I$ denote whole-portfolio net durations at that allocation, and $\hat{D}_I$ need not be zero.

Switch off bond-specific municipal risk for the moment. The household's active KKT equality and the insurer's condition for a profitable infinitesimal purchase are

$$\underbrace{\mu_T^M - \xi_H r}_{\text{offered to the household}} = \underbrace{\gamma_H \sigma_r^2 A_{T-1}^M \hat{D}_H}_{\text{household marginal risk cost}}, \quad (30)$$

$$\underbrace{\mu_T^M - \xi_I r}_{\text{offered to the insurer}} \geq \underbrace{\gamma_I \sigma_r^2 A_{T-1}^M \hat{D}_I}_{\text{insurer marginal risk cost}}. \quad (31)$$

If the second comparison fails, the insurer remains at zero. Equality marks marginal indifference, and strict inequality makes a positive position profitable at the margin. Substituting the household equality into the insurer comparison cancels the common bond return and leaves one condition:

$$\underbrace{\sigma_r^2 A_{T-1}^M \Lambda}_{\text{relative duration benefit (grows with maturity)}} \geq \underbrace{(\xi_I - \xi_H) r_t}_{\text{relative tax penalty (flat in maturity)}}, \quad \Lambda := \gamma_H \hat{D}_H^{\text{hh}} - \gamma_I \hat{D}_I^{\text{hh}}. \quad (32)$$

Assumption 1 imposes $\Lambda > 0$, meaning that the household is more duration-burdened after adjusting for risk tolerance. The tax penalty on the right is the same at every maturity. At the household-only benchmark, the appendix proof shows that the municipal loading begins at $A_0^M = 0$ and strictly increases with maturity. The duration benefit therefore starts at zero and rises with rate exposure, so the comparison is single-crossing in maturity.

Proposition 1 (Participation from the long end). *In the two-investor affine benchmark, at the household-only municipal allocation with $\Lambda > 0$, in the rate-only specification and at any state with $r_t > 0$, the maturities at which the insurer weakly prefers to participate form the upper interval $\{T : T \geq T^*(z_t)\}$, where*

$$T^*(z_t) = \min \left\{ T \geq 1 : A_{T-1}^M \geq \frac{(\xi_I - \xi_H) r_t}{\sigma_r^2 \Lambda} \right\}, \quad (33)$$

with $\min \emptyset = +\infty$. Strictly profitable participation uses a strict inequality; equality need not imply a positive holding.

Proof. See Appendix A.1 (proof of Proposition 1). □

The participation incentive is endogenous, and no investor is assigned a habitat. A larger tax disadvantage moves its threshold outward, while a larger duration advantage moves it inward. The proposition identifies profitable deviations from the household-only allocation. Clearing and the remaining KKT inequalities determine the resulting equilibrium holdings, as in the three-maturity illustration. Appendix A.1 restores bond-specific risk and gives sufficient conditions for upper-interval participation in the full affine equilibrium. The nonlinear calibration instead solves participation with exact conditional covariances and displays the same long-maturity pattern.

The baseline calibration uses only this rate-duration motive. The default-risk extension in §A.3.2 supplies an additional long-end diversification motive, which Result 1 does not need.

5.5 Result 2: participation changes prices through two channels

Result 1 gives the insurer's participation comparison. Appendix Theorem 1 extends the result to positive bond-specific risk. A positive equilibrium duration gap and nondecreasing predecessor loadings and inventory terms $\sigma_{\eta,T}^{M2} Q_T^M$ imply that the insurer's strictly positive municipal holdings form an upper interval. Result 2 isolates how the return required at each maturity reflects that participation. Write

$$s_T^M := \frac{\partial \mu_T^M(z)}{\partial z}.$$

The derivative s_T^M is the response of the bond's required one-period return to the current rate state. The central result is that this response has only two economic components, the tax valuations of the investors who hold the bond and the change in those investors' whole-portfolio duration.

Fix maturity T , and let $\mathcal{U}_T^M \subseteq \{H, I\}$ denote the investors with positive holdings of that municipal bond. Each active investor satisfies the same KKT equality,

$$\underbrace{\mu_T^M - \xi_i r}_{\text{after-tax return offered to } i} = \gamma_i \left[\underbrace{\sigma_r^2 A_{T-1}^M \hat{D}_i}_{\text{whole-portfolio rate risk}} + \underbrace{\sigma_{\eta,T}^{M2} x_{i,M}(T)}_{\text{bond-specific risk}} \right]. \quad (34)$$

Multiply investor i 's equation by m_i/γ_i , sum over the active holders, and differentiate with respect to z . The bond-specific terms cancel because fixed supply implies $\sum_i m_i \partial x_{i,M}(T, z)/\partial z = 0$. What remains is

$$\left(\sum_{i \in \mathcal{U}_T^M} \frac{m_i}{\gamma_i} \right) s_T^M - \sum_{i \in \mathcal{U}_T^M} \frac{m_i \xi_i}{\gamma_i} = \sigma_r^2 A_{T-1}^M \sum_{i \in \mathcal{U}_T^M} m_i \frac{\partial \hat{D}_i(z)}{\partial z}. \quad (35)$$

Define only the two objects needed to read the result:

$$\omega_{i,T}^M := \frac{m_i/\gamma_i}{\sum_{j \in \mathcal{U}_T^M} m_j/\gamma_j}, \quad E_T^M := \frac{\sum_{i \in \mathcal{U}_T^M} m_i \partial \hat{D}_i(z)/\partial z}{\sum_{i \in \mathcal{U}_T^M} m_i/\gamma_i}.$$

The first is investor i 's share of the active holders' risk-bearing capacity, and the second is their aggregate portfolio-exposure response, scaled by that capacity. The appendix uses the same aggregation for both markets and for state-specific tax keeps.

Derivatives are evaluated on a regular affine branch. A branch is regular when price moments are differentiable, active risky holdings and inactive KKT slacks are strictly pos-

itive, and the stacked active KKT-and-clearing Jacobian is nonsingular. These conditions keep the participating investors fixed locally and make the following derivatives well defined (Appendix A.1.1).

Proposition 2 (Two-channel required-return response). *In the two-investor affine benchmark, for any fixed maturity- T active set $\mathcal{U}_T^M \subseteq \{H, I\}$,*

$$s_T^M = \underbrace{\sum_{i \in \mathcal{U}_T^M} \omega_{i,T}^M \xi_i}_{\substack{\text{direct channel:} \\ \text{tax mix of this bond's holders}}} + \underbrace{\sigma_r^2 A_{T-1}^M E_T^M}_{\substack{\text{spillover channel:} \\ \text{change in their portfolio exposure}}}, \quad (36)$$

exactly. The direct term depends on who holds maturity T . The spillover term depends on how equilibrium trading changes those investors' exposure across their entire portfolios.

Proof. See Appendix A.1 (proof of Proposition 7, of which this is the affine specialization).

□

Direct channel: the tax mix of active holders. If the household alone holds maturity T , the first term is ξ_H . If both investors hold it, define their risk-tolerance weights

$$\omega_i := \frac{m_i / \gamma_i}{m_H / \gamma_H + m_I / \gamma_I}, \quad i \in \{H, I\}.$$

The direct term is then $\omega_H \xi_H + \omega_I \xi_I$. It changes only at maturities the insurer holds.

Spillover channel: portfolio reallocation across maturities. In the segmented benchmark, differentiate the household's and insurer's KKT equations at the jointly held long maturities and impose market clearing. Appendix Proposition 9 solves that system and gives

$$\frac{\partial \hat{D}_H(z)}{\partial z} > 0, \quad \frac{\partial \hat{D}_I(z)}{\partial z} = -\frac{m_H}{m_I} \frac{\partial \hat{D}_H(z)}{\partial z} < 0.$$

When the rate state rises, the insurer's relative tax disadvantage grows through $(\xi_I - \xi_H)r_t$. Her reallocation reduces total municipal duration exposure, and fixed supply makes the household absorb that duration. Individual jointly held positions need not all fall, because the signed result concerns the duration-weighted portfolio. The household then requires more compensation on every risky municipal bond it prices, including shorter bonds the insurer never owns. Where the household alone holds a risky maturity $T \geq 2$, equation (36) becomes

$$s_T^M - \xi_H = \gamma_H \sigma_r^2 A_{T-1}^M \frac{\partial \hat{D}_H(z)}{\partial z} > 0. \quad (37)$$

The spillover is zero for the one-period municipal bond because $A_0^M = 0$. It grows with the rate exposure of the bond being priced.

The segmented benchmark therefore reduces to one line:

$$s_T^M = \begin{cases} \xi_H + \gamma_H \sigma_r^2 A_{T-1}^M \frac{\partial \widehat{D}_H(z)}{\partial z}, & T < T^*, \\ \omega_H \xi_H + \omega_I \xi_I, & T \geq T^*. \end{cases} \quad (38)$$

Below the participation threshold, the insurer affects prices only indirectly, through the household's portfolio. Both investors hold the bond at and above the threshold, where the direct tax mix appears. In this benchmark, market clearing makes their aggregate exposure response cancel at the jointly held maturities, which leaves the direct term there. The spillover is thus a smooth, duration-scaled effect below the cutoff, whereas the direct channel starts at the maturity where participation begins.

5.6 Result 3: the two channels steepen the municipal curve

Result 1 explains the incentive to participate, and Result 2 shows how a feasible participation pattern transmits tax differences across maturities. A yield embeds all of these pricing decisions between today and maturity. The mapping from required returns to yields has two steps:

$$\begin{aligned} \underbrace{s_T^a}_{\text{KKT and clearing}} &\implies \underbrace{A_T^a = \rho A_{T-1}^a + s_T^a}_{\text{accumulate current and continuation responses}}, \\ &\implies \underbrace{y_T^a = (A_T^a z + C_T^a)/T}_{\text{divide the full price by maturity}}. \end{aligned}$$

The empirical object is the yield-implied tax,

$$\tau_T^{Y,\text{impl}} := 1 - \frac{y_T^M}{y_T^C}, \quad \text{as defined in Section 4.4.}$$

The relevant comparison is between the municipal response and the household-after-tax corporate response. When the insurer alone holds risky corporates, applying Result 2 to munis and its corporate counterpart (Appendix Proposition 8) gives

$$s_T^M - \xi_H s_T^C = \underbrace{(\bar{\xi}_T^M - \xi_H)}_{\text{direct channel: active holders' tax mix}} + \underbrace{\sigma_r^2 (A_{T-1}^M E_T^M - \xi_H A_{T-1}^C E_T^C)}_{\text{spillover channel: how trading changes both portfolios}}, \quad (39)$$

where $\bar{\xi}_T^M := \sum_i \omega_{i,T}^M \xi_i$ is the tax mix of the municipal bond's active holders and E_T^M is

defined in Result 2. Its corporate counterpart is $E_T^C := (\sum_{i \in \mathcal{U}_T^C} m_i \partial_z \hat{D}_i) / (\sum_{i \in \mathcal{U}_T^C} m_i / \gamma_i)$, with $\mathcal{U}_T^C$ the corporate bond's active holders. At one period there is no rate exposure and the household prices the municipal bond, so both terms are zero. Below the insurer's participation threshold, the equilibrium with its long-muni holdings has households requiring more compensation on munis and insurers requiring less on corporates than the no-participation benchmark. The two sides of the spillover therefore reinforce each other. At and beyond the threshold, the insurer also participates directly in the municipal tax mix. Under the stability and monotonicity conditions stated in the proposition below, this positive sequence of relative pricing responses accumulates through (29) into a rising municipal-to-corporate loading ratio. Appendix equation (68) gives the exact nonlinear covariance counterpart.

The long-end statement uses two limits. The tax keep of the active long-muni clientele is $\bar{\xi}_\infty^M := \omega_H \xi_H + \omega_I \xi_I$, and the insurer-priced corporate required-return response is $s_\infty^{C,\text{aff}} := \lim_{T \rightarrow \infty} s_T^{C,\text{aff}} \in (0, 1]$. Relief from municipal duration lowers the insurer's marginal cost of holding corporate duration. The corporate limit can therefore be below one.

Proposition 3 (Relative muni duration rises and the yield-implied tax declines). *In the affine two-investor benchmark, specialize Proposition 9 to the upper-interval joint-holding set $J = \{T^*, \dots, T_{\max}\}$: households hold short munis, both investors hold long munis, and the insurer holds corporates. Under the loading-stability and monotonicity conditions (S) and (M), stated in (73)–(74), and the duration advantage $\hat{\Lambda} := \gamma_H \hat{D}_H - \gamma_I \hat{D}_I > 0$:*

(i) Price response. *The municipal-to-corporate loading ratio is non-decreasing with maturity. Extending the same branch recursions to longer maturities with their solved exposure response held fixed, it begins at the household keep and converges to the long-muni clientele keep relative to the corporate response:*

$$\frac{A_1^{M,\text{aff}}}{A_1^{C,\text{aff}}} = \xi_H \quad \text{and} \quad \lim_{T \rightarrow \infty} \frac{A_T^{M,\text{aff}}}{A_T^{C,\text{aff}}} = \frac{\bar{\xi}_\infty^M}{s_\infty^{C,\text{aff}}}. \quad (40)$$

Against the slope-one treasury reference, $A_T^{\text{tsy}} = \sum_{j=1}^T \rho^{T-j}$, the same argument gives a non-decreasing municipal-to-treasury loading ratio from ξ_H to $\bar{\xi}_\infty^M$. This is the comparison used in Prediction P3.

(ii) Yield curve. *Under the forward-rate conditions of Proposition 11, positive corporate forward rates and a municipal-to-corporate forward-rate ratio that is non-decreasing in maturity, at the stationary mean $z = 0$ the yield ratio y_T^M / y_T^C is non-decreasing. Equivalently, the yield-implied tax $\tau_T^{Y,\text{impl},\text{aff}}$ is non-increasing.*

(iii) Long-end level. *Extend the branch recursions to longer maturities and suppose the loadings, risk-premium coefficients π_T^a , and inventory terms ι_T^a converge, with a positive*

limiting corporate forward. Then

$$\lim_{T \rightarrow \infty} \tau_T^{Y, \text{impl}, \text{aff}} = 1 - \frac{\bar{\xi}_\infty^M \bar{r} + A_\infty^M \pi_\infty^M - \frac{1}{2} \sigma_r^2 (A_\infty^M)^2 + \iota_\infty^M}{\bar{r} + A_\infty^C \pi_\infty^C - \frac{1}{2} \sigma_r^2 (A_\infty^C)^2 + \iota_\infty^C}, \quad (41)$$

where π_T^a and ι_T^a are defined by the level KKT aggregation in Appendix equation (75). All terms in this endpoint are affine-branch quantities.

Proof. See Appendix A.1. □

Part (i) is the price mechanism, in which relative municipal rate sensitivity rises first through the spillover and then through the direct channel. A loading result alone is not enough to order yields, so Part (ii) states the additional forward-rate condition that carries the ordering from price responses into yield levels. Part (iii) explains why the long-end implied tax differs from one minus the average tax keep of long-muni holders. Relative term premia, convexity, and bond-specific inventory costs also affect the yield. In the segmented benchmark, the household's larger duration burden raises the municipal duration premium relative to the corporate premium. Whether the full implied tax lies below the clientele-mix benchmark also depends on the remaining endpoint terms. The difference in risk-premium coefficients equals the household's share of the duration-premium gap,

$$\pi_\infty^M - \pi_\infty^C = \omega_H \sigma_r^2 \hat{\Lambda}, \quad \hat{\Lambda} := \gamma_H \hat{D}_H - \gamma_I \hat{D}_I > 0, \quad (42)$$

with $\hat{D}_i = \hat{D}_i(0)$ the net portfolio duration at the stationary-mean segmented allocation. Thus $\hat{\Lambda}$ is the equilibrium duration gap, a different object from the household-only Λ of Assumption 1. Suppose corporate loading feedback $b := \gamma_I \sigma_r^2 (m_H/m_I) \partial \hat{D}_H / \partial z$ is small relative to mean reversion ($b \ll 1 - \rho$), convexity and inventory terms are negligible, and the denominator is positive. To first order in the neglected feedback, convexity, and inventory terms, these conditions give the interpretable benchmark endpoint

$$\tau_\infty^{Y, \text{impl}, \text{aff}} \approx \underbrace{(1 - \bar{\xi}_\infty^M)}_{\text{clientele-mix tax}} - \underbrace{\frac{\bar{\xi}_\infty^M \omega_H \sigma_r^2 \hat{\Lambda}}{(1 - \rho) \bar{r} + \pi_\infty^{C, \text{aff}}}}_{\text{muni term-premium excess}}. \quad (43)$$

The household's larger duration burden thus lifts long muni yields and pushes the benchmark implied tax below $1 - \bar{\xi}_\infty^M$.

The endpoint holds at every fixed state because $A_T^a z / T \rightarrow 0$. Finite-maturity monotonicity is conditional on the full forward-ratio ordering. Proposition 11 gives the nonlinear counterpart, and Appendix A.2.4 assesses numerical maturity profiles on their own discretization.

5.7 Beyond liability hedging: joint participation and curve-wide tax effects

Joint participation is economically consequential even when its motive is not liability hedging. In this two-investor affine benchmark, insurers and households share some risky municipal maturities, households hold the remaining munis, and insurers hold corporates. The joint-holding set need not be an upper interval. The result is curve-wide. Because households choose one portfolio across maturities, an insurer's tax valuation affects the required return on a risky municipal bond it never holds.

When the short rate rises, the insurer's relative tax disadvantage grows, so insurers absorb less municipal duration and households absorb the difference. This raises the compensation households require on other risky municipal bonds, including shorter bonds insurers never hold. Jointly held bonds also reflect insurers' tax valuation directly. Proposition 9 in Appendix A.1.4 proves that, under its regularity and positive-loading assumptions, these channels give a positive relative required-return response at every risky maturity. The price recursion then gives, for $T \geq 2$,

$$\underbrace{\frac{\partial y_T^M}{\partial z} - \xi_H \frac{\partial y_T^C}{\partial z}}_{\text{relative yield response}} = \frac{1}{T} \sum_{j=2}^T \rho^{T-j} \underbrace{\left(\frac{\partial \mu_j^M}{\partial z} - \xi_H \frac{\partial \mu_j^C}{\partial z} \right)}_{\text{direct tax mix and portfolio spillover}} > 0. \quad (44)$$

Here μ_j^a is the required one-period return and y_T^a is the yield, in model-period units. Both sides are zero at the fully riskless one-period bond. Tax differences therefore affect relative yield responses even at maturities insurers never hold.

Motives other than liability hedging can also sustain this participation. Let κ_{iT} be an additional per-unit holding benefit that does not vary with the rate state. Such a benefit shifts the insurer's net disadvantage at a given rate but leaves unchanged how that disadvantage responds to rates:

$$\frac{\partial}{\partial z} \left[\underbrace{(\xi_I - \xi_H)r(z)}_{\text{relative tax penalty}} - \underbrace{(\kappa_{IT} - \kappa_{HT})}_{\text{other relative holding benefit}} \right] = \xi_I - \xi_H > 0. \quad (45)$$

If these benefits sustain the same feasible participation pattern, with covariances, financing, and constraints held fixed, the transmission equations are unchanged. The benefits can change return levels and participation incentives without changing this conditional transmission mechanism. The active-KKT derivation for [alternative holding benefits](#) is in Ap-

pendix [A.1.4](#).

Equation (44) concerns yield responses to the rate state. The maturity-ordering and forward-rate conditions in Result 3 connect this transmission to a rising municipal-to-corporate yield ratio and a declining yield-implied tax. Appendix [A.1.5](#) contains the proofs, including Proposition 11. In the baseline, liability hedging explains participation, and shared portfolio risk transmits its pricing consequences across maturities.

5.8 From mechanism to empirical predictions

The three predictions isolate whose tax matters and how rate risk is priced. P1 concerns household taxes. Their full pass-through at the riskless short boundary gives way to pricing by several clienteles at longer maturities. Institutional taxes, the subject of P2, have no effect at that boundary and affect longer bonds through participation and portfolio reallocation. P3 concerns the relative rate sensitivity of municipal and treasury yields. A constant representative-investor tax wedge produces none of these maturity gradients.

Larger tax benefit for	Short end	Longer maturities
Households	Full pass-through	Weaker pass-through
Institutions	No effect	Long-bond yields respond

The representative-tax benchmark fills both rows with the same entry at every maturity, full household pass-through and no institutional effect.

The tax-shock endpoints and the shape between them. At the fully riskless one-period boundary, the municipal-to-corporate yield ratio equals the household keep whenever $r(z) > 0$ (Appendix Proposition 4). Household-tax pass-through is therefore one, and a strictly less tax-advantaged insurer's pass-through is zero. Longer bonds embed pricing by more than one clientele. To isolate each clientele's direct contribution, hold the participation pattern and risk-tolerance weights fixed and switch off the exposure-response terms in the affine recursions. Put

$$W_S(T) := \frac{\sum_{j=1}^{\min(T, T^*-1)} \rho^{T-j}}{\sum_{j=1}^T \rho^{T-j}}, \quad W_L(T) := 1 - W_S(T).$$

These are the loading weights on the household-only and jointly held parts of the maturity ladder. Differentiating the clientele-weighted loading sum gives

$$\begin{aligned} \mathcal{D}_H(T) &= W_S(T) + \omega_H W_L(T) : \quad 1 \longrightarrow \omega_H, \\ \mathcal{D}_I(T) &= \omega_I W_L(T) : \quad 0 \longrightarrow \omega_I. \end{aligned} \tag{46}$$

Once insurer participation begins, the weight on the early household-only block shrinks, so the household’s direct component declines and the insurer’s rises. Because the two components sum to one, the direct mechanism changes whose tax is reflected in the bond’s rate sensitivity. Appendix Proposition 12 derives the formulas and distinguishes them from the full equilibrium derivative. A maturity ordering of implied-tax levels alone would not prove this ordering of tax responses.

Exposure feedback can transmit an insurer’s tax even to risky maturities it does not hold, and yield responses also include changes in term premia, convexity, and the corporate leg. The direct components in (46) are therefore a mechanism decomposition, not the full nonlinear tax derivative. Proposition 13 gives the exact price-response identity and the additional sign and relative-leg conditions. The baseline equilibrium cross-state comparison in Section 7.1 supplies the quantitative P1 profile. The single-state policy re-solves in Section 8 are a distinct exercise, and Appendix A.2.4 reports the grid sensitivity of the numerical curve. The empirical tests concern the broad maturity gradients. They do not examine strict ordering at every numerical node.

Prediction (P1: household state-tax pass-through declines with maturity). A household’s own state tax passes into the yield-implied state tax one-for-one at the short boundary, and its pass-through declines as other clienteles participate. The leading affine long-end component approaches the household’s risk-tolerance weight among the active long-end holders; the nonlinear response also includes participation, covariance, and term-premium adjustments. Proposition 15 gives the formal multi-state result. The representative-tax benchmark predicts a pass-through of one at every maturity.

Prediction (P2: institutional tax benefits affect longer maturities). A strictly less tax-advantaged institution’s tax has zero effect at the riskless short boundary. Its direct contribution becomes positive as it participates, approaching its active risk-tolerance weight along the affine long-end branch. The bank-qualified design tests whether a bank-specific benefit is capitalized farther along the curve. Complementary TCJA evidence tests for a larger response at longer maturities and where predetermined P&C exposure is greater. The representative-tax benchmark assigns neither institutional benefit a role at any maturity.

Prediction (P3: relative rate sensitivity rises with maturity). The local municipal-to-treasury yield response is

$$\frac{\partial y_T^M / \partial z}{\partial y_T^{\text{Tsy}} / \partial z} = \frac{A_T^M}{A_T^{\text{Tsy}}}.$$

In the affine slope-one treasury benchmark, this is a positive-weighted average of municipal required-return responses. Under condition (M), it rises from the household keep toward

the long-maturity clientele mix (Proposition 3). The representative-tax benchmark predicts a constant response. Transaction-level treasury movements provide a descriptive test; treasuries are an external rate benchmark, not a third market in the calibrated equilibrium.

Appendix A.1.10 illustrates how the direct and spillover components accumulate on a longer maturity ladder.

Multiple clienteles and scope. Taxes, liability hedges, and portfolios determine participation, and with several investors the same KKT comparison applies pairwise. Different balance sheets can make a less tax-advantaged investor participate earlier, so thresholds need not be tax ordered. Appendix A.1.7 gives the conditional multi-clientele formulas. The quantitative economy solves all clienteles and state markets jointly, without assigning maturity habitats.

Investor-specific taxes, common portfolio risk, and limits to arbitrage also arise in taxable and tax-favored accounts, in state taxation of treasuries and corporates, and in cross-border withholding. Section 5.7 separates conditional transmission from the motive that sustains participation, and Appendix A.3 discusses these applications.

6 Calibration and Model Results

The calibration asks whether tax rules and balance sheets generate the municipal curve without targeting municipal yields or the empirical test coefficients. The full nonlinear economy has 45 municipal markets, one corporate market, and 48 investor types. One period is two calendar years, with an exactly riskless first-period claim and maturities through thirty years. Appendix A.2 gives the equilibrium algorithm and checks.

6.1 Calibration inputs and targets

The municipal markets comprise 44 taxed jurisdictions, including DC, and a block combining seven no-income-tax states. Resident households are fully exempt on home-state municipals and pay their own state’s statutory top rate on out-of-state municipals and corporates. The supply-weighted state rate is 5.9%, and the no-tax block represents 18% of supply. Household masses are proportional to municipal supply and sum to 0.30. Their federal keep of 0.62 corresponds to a sample-average top rate of approximately 38%, including the net investment income tax.

Three institutional clienteles complete the economy: P&C insurers, life insurers, and a residual sector that is largely tax-exempt. Table 4 gives their masses, relative tax keeps,

and liability exposures. Flow-of-Funds ownership disciplines the sector mapping. P&C taxes follow statutory proration, and the life-insurer keep is a maintained effective rate. A reduced calibration to the corporate long-end level, with no municipal yield among its targets, fixes P&C mass. The one-keep convention of Section 5.2 applies asset-specific keeps to both expected holding returns and their innovations.

Annual rate persistence is 0.85 and innovation volatility is 1.8%, both aggregated to the two-year grid. The annualized latent cash-rate level, 2.6222%, and common risk aversion, $\gamma_i = 8.9732$, match the continuously compounded two- and thirty-year corporate yields of 2.6222% and 5.7935% at $z = 0$. These targets set the taxable level and the scale of risk prices. Neither uses municipal yields or tax pass-through.

Liability exposures represent the structure of sector balance sheets. Life insurers receive a larger liability hedge than P&C insurers, consistent with their longer-lived contracts and greater sensitivity to long rates (Brunetti et al., 2023). The normalized institutional loadings are $\ell_{I_1} = 0.50$, $\ell_{I_2} = 2.22$, and $\ell_{I_3} = 0$. Households' outside assets add duration, with $\ell_H = -1.387$. The loadings are calibrated approximations to net exposure and do not measure liability durations in years. Appendix A.2.6.2 defines the balance-sheet mapping. Municipal quantities determine state supply shares, and the corporate-to-municipal supply ratio facing the modeled clientele is maintained at 5.8. Mean maturity is ten years for municipal supply and 11.7 years for risky corporate supply. Table 14 distinguishes measured inputs, corporate targets, and maintained mappings.

The taxable targets discipline the overall return level, and balance sheets determine how investors share duration across markets. At the same reporting state and yield convention as the 86% result, the $\pm 20\%$ life-liability comparisons explain 75.8–94.4% of the thirty-year gap, and varying life-insurer mass gives 69.3–99.0%. The comparisons change one input at a time and re-solve the equilibrium without refitting the corporate targets. Appendix Table 13 reports both the share and the resulting corporate-target error. Across all ten comparisons, the implied tax falls by 12.1–16.5 percentage points between two and thirty years. Changing the idiosyncratic-risk profile also preserves the pattern (Appendix A.2.5). The maturity mechanism thus survives substantial perturbations. Its quantitative contribution depends on risk-bearing capacity and on the externally anchored taxable curve.

6.2 Yield curves and the implied tax: model versus data

Figure 5 reproduces the empirical pattern of Figure 2. Model municipal yields start near the household benchmark and exceed it by more at longer maturities. The model's implied tax is 41.84% at two years, 33.53% at ten, and 27.01% at thirty, a 14.83-percentage-point decline.

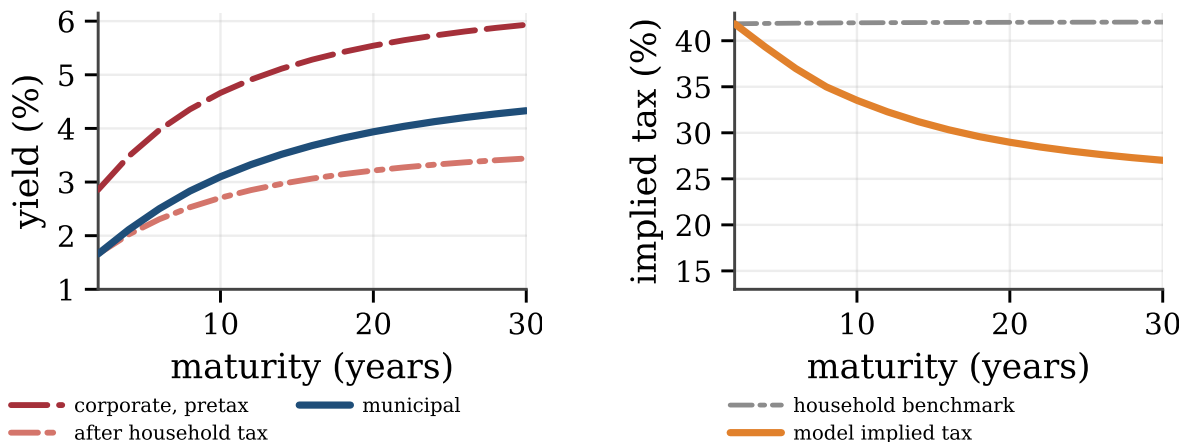
Table 4: Baseline calibration

Symbol	Description	Value
<i>A. Rate process</i>		
$T_{\max}$	Maximum maturity (years)	30
$\bar{r}_{\text{ann}}$	Annualized latent cash-rate level	0.0262
ρ_{ann}	Annual AR(1) persistence of the rate state	0.85
$\sigma_{r,\text{ann}}$	Annual state innovation volatility	0.018
<i>B. Idiosyncratic risk</i>		
σ_{idio}	Risky-node idiosyncratic return volatility (scale)	4×10^{-3}
$\kappa(T)$	Risky-node curve ($T \geq 4$ years); variance at $T = 2$ is zero	$1 + 7\sqrt{(T-1)/29}$
<i>C. State-tax cross-section (statutory rates from data)</i>		
$\{\tau_s^{\text{st}}\}$	Taxed states (with DC)	44
No-tax block	Zero-income-tax states	7 (18% supply)
$\bar{\tau}^{\text{st}}$	Supply-weighted mean state surcharge	5.9%
<i>D. Investor types (all $\gamma_i = 8.9732$)</i>		
H_s	$\xi_{H_s}^{\text{fed}}$ / surcharge / total household mass / ℓ	$0.62 / \tau_s^{\text{st}} / 0.30 / -1.387$
I_1	ξ_{I_1} / mass / ℓ	$0.686 / 0.10 / +0.50$
I_2	ξ_{I_2} / mass / ℓ	$0.79 / 0.40 / +2.22$
I_3	ξ_{I_3} / mass / ℓ	$1.00 / 1.80 / 0$
<i>E. Aggregate supply</i>		
$\bar{T}^M, \bar{T}^C$	Mean maturity of supply profile	10, 11.7
$\int S^M$	Total muni supply facing the modeled clienteles	0.30
$\int S^C$	Total corp. supply ($\approx 5.8 \times$ muni)	1.73

Notes. Corporate targets are continuously compounded yields at $z = 0$: 2.6222% at two years and 5.7935% at thirty. No municipal yield is targeted. The first period is fully riskless; the default-shock extension is inactive in the baseline. Positive ℓ_i denotes bond-like liabilities; negative ℓ_i denotes outside-asset duration. Household mass is split across the 45 resident types. Section 6.1 defines the clienteles and tax keeps; Table 14 gives sources and Appendix A.2.6.1 records the units and extension inputs.

At thirty years, municipal and corporate yields are 4.330% and 5.933%. The reporting state $z^* = 0.004257$ matches the observed two-year corporate curve within 0.05 basis point without changing a parameter. Under the reporting-state and nominal-semiannual conventions, the model's own thirty-year corporate yield of 5.933% compares with 5.794% in the observed benchmark. The explained-share calculation uses the same observed corporate curve in both the data and model comparisons.

Household taxes and the riskless boundary pin the short yield. Beyond that boundary, joint participation activates the direct tax mix and portfolio spillover. The untargeted nonlinear solution produces both a rising relative municipal rate response and a declining implied tax. It also retains the state-dependent covariance feedback that is absent from the



Panel A: Model yield curves.

Panel B: Model implied tax by maturity.

Figure 5: The calibrated yield curves and implied tax. Both panels use nominal semiannual yields at the reporting state z^* . The household-tax benchmark applies the short-end keep to the continuous corporate yield before conversion; municipal yields are averaged across states before conversion. As in the empirical curves of Figure 2, municipal yields track the household benchmark at the short end and exceed it farther out. No municipal yield is targeted.

Alt text: The left panel shows model municipal and corporate yield curves: the municipal curve matches the household after-tax benchmark at the short end and increasingly exceeds it at longer maturities. The right panel shows the corresponding decline in implied tax.

analytical benchmark.

6.3 The model, bond-side adjustments, and observed curves

Figure 6 places the model and the bond-side adjustments on a common spread over the household-after-tax corporate benchmark. At thirty years, callability and the issuance premium account for 50.1 of the observed 159.8 basis points. The remaining 109.7 basis points are directly comparable to the model's curve, which contains no call option and no issuance premium. The model supplies 94.8 basis points, or 86%. The separately estimated capital-gains, liquidity, and default differentials add 18.4 and leave a residual of -3.5 basis points.

In Table 5, the model share rises from 56% at ten years to 79% at eighteen and 86% at thirty. The three differentials stay near twenty basis points, so the increase reflects the model's growing contribution. The root-mean-square residual over the full curve is 6.6 basis points, and Appendix Table 15 reports every component. The comparison is an accounting of separately estimated effects, and their cross-friction interactions are not estimated.

The result concerns both size and shape. A calibration that does not target municipal yields generates a substantial gap, and that gap strengthens toward the long end. The

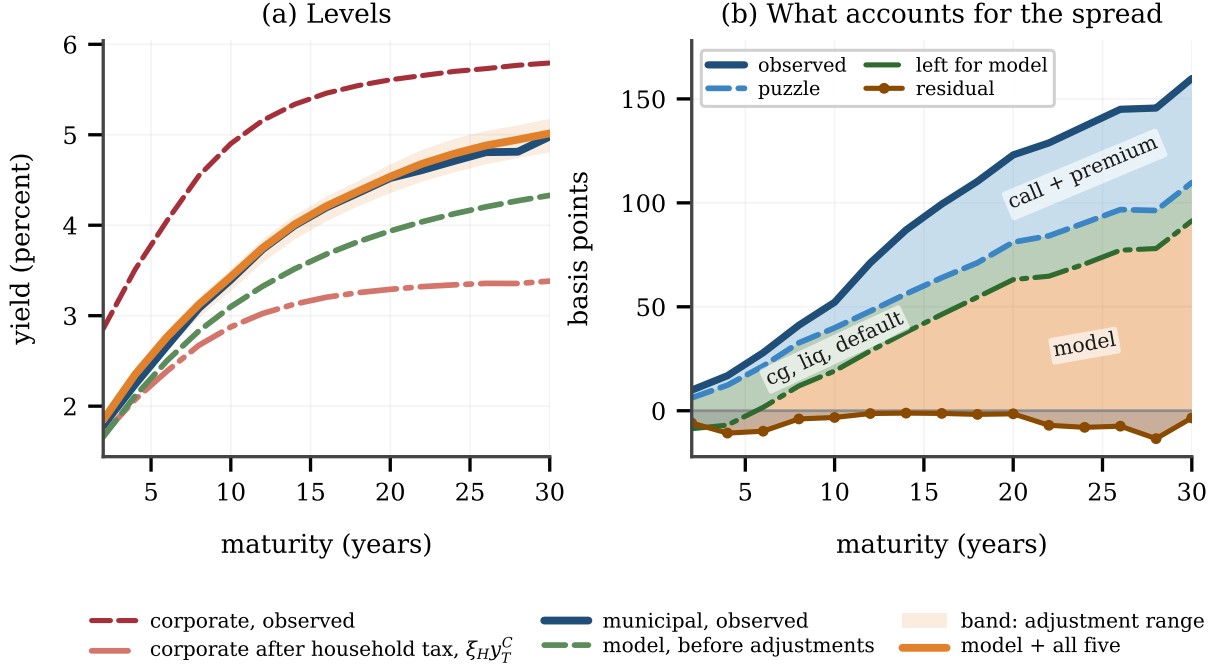


Figure 6: The model and bond-side adjustments. Panel (a) compares observed curves with the calibrated municipal curve and the model plus five adjustments; shading gives the adjustment range. Panel (b) decomposes the observed spread over the household-after-tax corporate benchmark into callability, issuance premium, three muni-minus-corporate differentials, the model, and the residual. Near two years the differentials exceed the small puzzle, so the bands record over-explanation. The household benchmark is constructed as in Figure 2. Appendix Table 15 gives the full ledger. No parameter is chosen to reduce its residual; call and capital-gains calculations use equilibrium discount surfaces, while the other adjustments are data-based.

Alt text: The figure compares the municipal curve with the calibrated model and the model plus separately measured bond-side adjustments. The residual panel displays the remaining maturity-specific difference on a basis-point scale.

empirical tests next examine the investor-specific transmission behind the profile.

7 Empirical Predictions and Tests

The main tests use two tax changes that benefit different investors, household state taxes and bank-qualified designation. They ask whether the yield response changes with both the recipient of the benefit and maturity. Each design tests one prediction of Section 5.8. P1 covers the household component and P3 the relative rate response; the bank-qualified design and the insurer-tax shock cover the institutional component. Under the designs' identifying assumptions, a maturity-dependent credit or liquidity premium that is unchanged by these tax treatments cannot itself explain that joint pattern. The TCJA insurer-tax shock and

Table 5: The calibrated model’s contribution to the municipal yield gap. Removing callability and the issuance premium from $\text{Spread}(T)$ leaves s_{adj} , the gap comparable to the model’s option-free curve. The model column reports its basis-point contribution and share of this adjusted gap. The remaining columns give the capital-gains, liquidity, and default differentials (cg/lq/df) and the residual.

T	$\text{Spread}(T)$	curve-level adj.		s_{adj}	explained by the model		not explained by it	
		call	prem.		bp	share	cg/lq/df	resid.
2	9.9	0.1	3.6	6.2	−2.7	—	14.8	−5.9
6	27.9	0.6	5.8	21.5	11.5	53%	19.9	−9.8
10	52.2	4.5	7.9	39.9	22.3	56%	20.8	−3.2
14	87.0	18.5	12.3	56.2	38.9	69%	18.4	−1.1
18	110.3	23.9	15.3	71.2	56.4	79%	16.5	−1.8
22	128.8	26.8	17.9	84.1	71.7	85%	19.4	−7.0
26	145.0	27.8	20.5	96.7	84.6	87%	19.5	−7.4
30	159.8	29.0	21.1	109.7	94.8	86%	18.4	−3.5

Notes. Basis points; ratios use unrounded values. Shares are omitted when the adjusted spread is below 20 basis points. The model curve is from the fully riskless-boundary equilibrium and is converted to the same nominal semiannual yield convention as the observed curves. The empirical curves and five adjustment estimates are held fixed. The three columns on the right sum to the adjusted spread, but the components are not jointly estimated or jointly repriced; this is an accounting comparison, not an identified additive decomposition.

transaction-level rate responses provide complementary evidence. None of these coefficients is a calibration target.

The curve accounting in Sections 4 and 6 uses the A-or-above municipal sample (agency score seven or better), which matches the credit band of the corporate benchmark. The transaction-level tests use the narrower AA-and-above sample (agency score three or better) to limit credit heterogeneity that could affect both yields and trading patterns. Each exercise is also reported on the wider sample, and Appendix A.7.2 shows that the rating cutoff does not drive the maturity ordering of the point estimates. P1 and the bank-qualified discontinuity supply treatment-level evidence, while the TCJA and treasury-response designs are complementary tests. Each table states its sample restrictions. Below, m_b denotes remaining maturity in years and is distinct from the investor masses m_i in the model.

7.1 P1. State-tax pass-through declines with maturity

Most states exempt residents’ in-state municipal interest while taxing out-of-state municipal and corporate interest. This wedge supplies P1’s variation, and it is specific to households because institutions generally do not receive the same resident benefit. Single-state funds preserve the wedge, whereas national funds attenuate it (Appendix A.9.3.2).

The outcome removes the federal layer from $q_{s,t}(T) := y_{s,t}^M(T)/y_t^C(T)$. Before 2018, federal deductibility of state taxes makes the household keep $\xi^{\text{fed}}(1 - \tau_s^{\text{st}})$. When the post-2018 SALT cap binds, the keep is $\xi^{\text{fed}} - \tau_s^{\text{st}}$. Accordingly,

$$\text{ImpliedStateTax}_{s,t}(T) = 1 - q_{s,t}(T)/\xi^{\text{fed}}, \quad t \leq 2017, \quad (47)$$

$$= \xi^{\text{fed}} - q_{s,t}(T), \quad t \geq 2018. \quad (48)$$

Under household parity, both expressions recover the statutory state rate in the relevant regime. The second applies the binding-cap treatment that is relevant to top-bracket residents in high-tax states.

The model predicts unit pass-through at the riskless boundary, where home households receive the largest benefit, and a declining influence as other clienteles participate. The representative-tax benchmark predicts a pass-through of one throughout. Figure 7 compares this prediction with the cross-state slope of the nonlinear calibration at z^* , including its yield-level corrections. That slope is not a one-state structural partial derivative.

Specification. Partition bonds into five maturity bins by years to maturity ($[0, 5)$, $[5, 10)$, $[10, 15)$, $[15, 20)$, ≥ 20 years) and estimate

$$\text{ImpliedStateTax}_{b,t} = \sum_{g=1}^5 \beta_g (\text{StateTax}_{s(b),t} \times \mathbf{1}\{m_b \in g\}) + \alpha_{t \times m \times \text{callable}} + \delta' X_{b,t} + \varepsilon_{b,t}, \quad (49)$$

where $\text{StateTax}_{s,t}$ is the statutory top rate, $X_{b,t}$ is the mean of the available Moody's, S&P, and Fitch numeric rating scores, and α absorbs time-by-maturity-by-callability effects. Each bond appears once, in its issuance month, with equal weight. Standard errors are clustered by state, the level at which the rate and the instrument vary. Each β_g is the pass-through of the statutory rate into the yield-implied state tax in bin g . The model predicts β_1 near one and a decline across bins, whereas the benchmark predicts $\beta_g = 1$ in every bin. The tests are the unit restriction in the shortest and longest bins and the difference $\beta_1 - \beta_5$.

The IV specification instruments each maturity-bin interaction with that bin's 1990 top state rate from ITEP, following the policy-history approach of Babina et al. (2021). Its exclusion restriction is that, conditional on the controls, historical tax structure affects current implied-tax pass-through only through the current rate. The 1990 vintage balances temporal separation against relevance, and its state-clustered Kleibergen–Paap system F is 10.9. Appendix A.6.6 reports earlier and later vintages and alternative state weights.

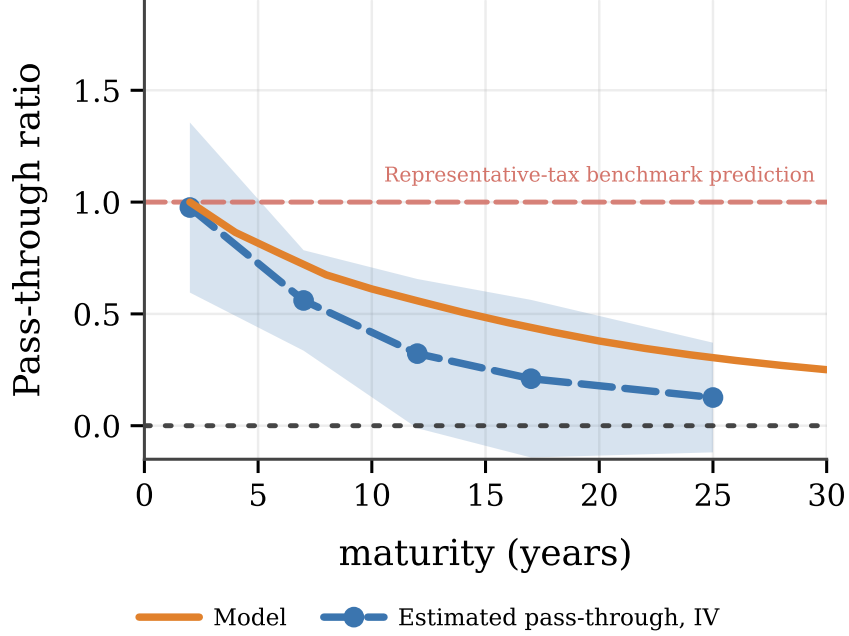


Figure 7: State-tax pass-through by maturity. The blue line and reported 95% band show the IV maturity-bin estimates from (49), on the reported call-adjusted zero-coupon series, unweighted, with the 1990 top state rate as instrument and the band clustered at the state. Repricing inputs are held fixed for this model comparison. The orange line is the riskless calibration’s cross-state OLS slope of implied state tax on the statutory rate, evaluated at z^* using continuously compounded state yields; it is not a one-state structural partial derivative. The dashed unit line, labeled on the figure, is the representative-tax benchmark, which predicts a flat profile at one.

Alt text: Household state-tax pass-through declines across five maturity bins. Empirical estimates and their state-clustered intervals are compared with the model profile and a horizontal unit-pass-through benchmark.

Result. Under IV, pass-through declines from 0.976 below five years to 0.126 beyond twenty, and under OLS it declines from 0.987 to 0.159 (Table 6). A ten-point increase in the statutory rate raises implied tax by about 9.8 points at short maturities and by 1.3 points at the long end. The short-minus-long difference is 0.85 under IV (standard error 0.19) and 0.83 under OLS (0.17). For OLS, the state-level wild-bootstrap test gives $p = 0.020$.

Unit pass-through is consistent with the shortest-bin estimates but rejected beyond twenty years. Without targeting these coefficients, the model produces slopes of 1.00 at two years, 0.61 at ten, and 0.25 at thirty, and its profile lies inside the 95 percent band at all five bin midpoints. The central comparison is the pronounced decline from the household short boundary. This profile is what the mechanism implies. The resident household’s benefit is fully priced where that household is the marginal holder, and it fades where other clienteles share the maturity. The benchmark’s flat unit pass-through fails in the long bins,

which is where the model predicts the dilution.

The profile survives dropping one state at a time, splitting the sample at the SALT reform, changing the instrument vintage and weights, using offering yields or non-call-adjusted zeros, and widening the rating band. With state fixed effects, the short-minus-long difference is 0.752 (standard error 0.150). Appendices A.6.5 and A.6.6 report the estimates and treatment-level inference.

Table 6: P1: implied state-tax pass-through by maturity bin.

	(1) OLS	(2) IV
β_1 : StateTax $\times$ [0,5)y	0.987*** (0.143)	0.976*** (0.194)
β_2 : StateTax $\times$ [5,10)y	0.564*** (0.090)	0.560*** (0.115)
β_3 : StateTax $\times$ [10,15)y	0.347** (0.164)	0.322* (0.170)
β_4 : StateTax $\times$ [15,20)y	0.244 (0.187)	0.210 (0.180)
β_5 : StateTax $\times$ ≥ 20 y	0.159 (0.142)	0.126 (0.125)
H_0 : $\beta_1 = 1$ (t)	-0.09	-0.13
H_0 : $\beta_5 = 1$ (t)	-5.92	-6.98
H_0 : $\beta_1 = \beta_5$ (p)	< 0.001	< 0.001
Wild-cluster bootstrap p , $\beta_1 = \beta_5$	0.020	—
Kleibergen–Paap F	—	10.9
Observations	834,553	834,553
Rating control	Yes	Yes
Weights	None	None
Instrument: 1990 top state rate	No	Yes
SE clustered at state	Yes	Yes

Notes. The dependent variable is the implied state tax built from the call-adjusted zero-coupon yield, and the coefficients reported are those of equation (49). The IV column instruments each bin’s current statutory rate with the same bin’s 1990 top rate. Standard errors, and every test statistic, are clustered at the state (47 clusters), the level at which the statutory rate and its instrument vary. The representative-tax benchmark predicts $\beta_g = 1$ in every bin; the two t -statistics test that restriction in the shortest and longest bins. The decline test is the p -value of $\beta_1 - \beta_5 = 0$, and the bootstrap row is its state-level wild-cluster p -value (Rademacher weights, 9,999 replications). Offering yields and non-call-adjusted zeros give the same declining profile and are reported in Appendix A.6.5; the instrument vintage and the weighting are varied in Appendix A.6.6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7.2 P2. Bank qualification and the maturity profile of institutional tax effects

The bank-qualified design tests the institutional participation prediction with an issuer-level statutory threshold. It asks whether a benefit available only to banks lowers municipal yields equally across maturities or has a larger effect farther out. Within a bank, the annual benefit per dollar of holdings is common across maturities, so a yield effect that rises with maturity connects the bank’s tax treatment to its participation along the curve.

Section 265(b) of the Internal Revenue Code ordinarily denies banks the deduction of interest expense allocated to tax-exempt bonds. For obligations designated “bank-qualified” by a qualifying small issuer, Sections 265(b)(3) and 291 instead allow an 80% deduction. The issuance cap was \$30 million in 2009–10 under the American Recovery and Reinvestment Act and \$10 million otherwise. Designation alters the bank’s tax treatment and leaves the bond’s exemption for households, funds, and insurers unchanged. These small serial issues are an important segment of local bank lending, and [Dagostino \(2025\)](#) studies how the expanded cap affected municipal issuance and employment.

The statute allocates a bank’s interest expense across its assets in proportion to their average adjusted bases. The allocation does not match each bond to borrowing of the same maturity. Let I_{bt} be bank b ’s annual interest expense and $\bar{A}_{bt}$ its average asset basis. A dollar of municipal basis held throughout the year is allocated $c_{bt} = I_{bt}/\bar{A}_{bt}$ of interest expense. Designation reduces the nondeductible fraction from one to 0.2, saving

$$X_{bt} = \underbrace{(1 - 0.2)}_{\text{restored deduction}} \times \underbrace{\tau_{B,t} c_{bt}}_{\text{tax cost of full disallowance}} = 0.8 \tau_{B,t} c_{bt}. \quad (50)$$

A two-year and a twenty-year bond held by the same bank in the same year therefore receive the same annual benefit per dollar of basis. The equality does not assume a flat funding yield curve. Funding costs can change across banks and years, and the present value of future deductions can differ across maturities.

I measure I_{bt} with total interest expense from Schedule RI of the Call Report and take the denominator from the FDIC’s published average-assets series. The annual c_t is the municipal-securities-weighted average of these ratios for all banks with positive municipal-securities holdings, obtained through FDIC BankFind. Appendix [A.6.9.3](#) identifies the report items and the construction. The federal bank rate was 35% before 2018 and 21% thereafter. With the measured funding costs, $X_t = 0.8 \tau_{B,t} c_t$ is about 65 basis points in 2005–08, 14 in 2011–13, 9 in 2014–17, and 11 in 2018–19.

A bond is priced on the deduction over its life, so the scaled estimates use the benefit the

bank expects over the bond's remaining maturity, $X_{tm} = 0.8 \tau_{B,t} \hat{c}(t, m)$. Here $\hat{c}(t, m)$ is an expected funding-cost proxy that uses the available horizon $\min(m, 10)$, and it is built in three steps. The risk-neutral treasury yield at $\min(m, 10)$ years in the issue month summarizes the expected path of the short rate over the bond's life. This yield is net of the term premium estimated by the Kim–Wright model, so a premium that rises with maturity stays out of the denominator. The fitted annual pass-through from the one-year expected rate to the municipal-securities-weighted cost of funds ($\hat{c} = 0.38 + 0.55 r$, $R^2 = 0.75$) then maps that expected average rate to bank funding costs. Because the mapping is linear, $\hat{c}(t, m)$ averages expected funding costs over the available horizon, and bonds longer than ten years use its ten-year value. Finally, the issue year's deduction rate converts the funding cost into an annual benefit per dollar of basis. The resulting benefit has a median of 45 basis points and never falls below 8. Appendix A.6.9.2 gives the full construction, the same measure built from the Adrian–Crump–Moench term premia, and the evidence that the two agree.

Three hypotheses. Let $\pi_k = y_k^N - y_k^Q$ denote the yield premium of a non-designated over a designated bond at maturity k . The annual-carry benchmarks compare otherwise matched bonds and isolate the bank-only deduction. Suppress the date index, and let $y_k^T > 0$ be the taxable comparison yield.

In the representative-household benchmark, the non-designated curve satisfies $y_k^N = (1 - \tau_H)y_k^T$. Consider whether designation would induce a bank to accept that same yield. After the remaining funding-cost disallowance, the bank would earn $y_k^N - 0.2 \tau_{BC}$, compared with $(1 - \tau_B)y_k^T$ on the taxable bond. The taxable bond's advantage is

$$\underbrace{(1 - \tau_B)y_k^T}_{\text{taxable alternative}} - \underbrace{[(1 - \tau_H)y_k^T - 0.2 \tau_{BC}]}_{\text{qualified bond at the household yield}} = (\tau_H - \tau_B)y_k^T + 0.2 \tau_{BC} > 0. \quad (51)$$

The inequality holds whenever $\tau_H > \tau_B - 0.2 \tau_{BC}/y_k^T$, and in particular when households have at least the bank's tax rate and funding costs are positive. Even with designation, the bank would earn more on the taxable alternative. Because designation does not change households' own tax treatment, this benchmark predicts $y_k^Q = y_k^N$ and $\pi_k = 0$ at every maturity.

In the bank-only benchmark, both municipal bonds satisfy the bank's after-tax indifference conditions:

$$\begin{aligned} y_k^N - \underbrace{\tau_{BC}}_{\text{full disallowance}} &= (1 - \tau_B)y_k^T, \\ y_k^Q - \underbrace{0.2 \tau_{BC}}_{\text{qualified disallowance}} &= (1 - \tau_B)y_k^T. \end{aligned} \quad (52)$$

Subtracting the second condition from the first gives $\pi_k = 0.8\tau_{BC} = X$ at every maturity. The taxable yield can vary along the curve but cancels within each pair. The bank-only benchmark therefore predicts a flat premium in basis points and a capitalization share $\pi_k/X = 1$.

The model instead lets tax advantages and duration benefits differ across investors. At the short end, little duration risk needs to be absorbed, so the household tax advantage anchors yields and the bank-only deduction has little effect. At longer maturities, duration benefits support institutional participation despite the tax disadvantage. Applied to banks, this mechanism predicts that the same deduction has a larger yield effect farther out:

$$\pi_k = \underbrace{\theta_k}_{\text{capitalization share}} X, \quad \theta_{\text{short}} \simeq 0, \quad \theta_{\text{long}} > \theta_{\text{short}}. \quad (53)$$

The distinction is sharp. The household benchmark predicts zero along the whole curve, and the bank-only benchmark predicts the full annual benefit at every maturity. Under the model, effects are small at the short end and larger at the long end. The statutory benefit is common across maturities; its capitalization is not. Appendix A.6.9.1 derives θ_k from portfolio demand and market clearing.

Specification. The sample is every tax-exempt new issue in the Mergent file from 2000 to early 2021, with one observation per maturity and its offering yield. Bonds callable at a premium are excluded because they are quoted to the call. The running variable R_{jt} is issuer j 's total tax-exempt issuance in calendar year t , designated or not, minus the cap of that year, in millions of dollars. The statute qualifies an issuer that issues no more than the cap, so eligibility is $E_{jt} = \mathbf{1}\{R_{jt} \leq 0\}$, and B_i is the bond's bank-qualified designation. The estimation window is $0.5 \leq |R_{jt}| \leq 4$. Its half-million donut removes the issuers that bunch just under the cap (Figure 8, panel b) and the 1,034 issuer-years that sit exactly at it, which are eligible and designate at the rate of those below. The maturity bins at issuance are 1–2, 2–5, 5–10, 10–20 and more than 20 years. With $\mathbf{1}_k$ indicating the bond's bin, the first and second stages are

$$\begin{aligned} B_i &= \sum_k \alpha_k E_{jt} \mathbf{1}_k + W_i + e_i, \\ y_i &= \sum_k \beta_k B_i \mathbf{1}_k + W_i + u_i, \quad B_i \mathbf{1}_k \text{ instrumented by } E_{jt} \mathbf{1}_k, \end{aligned} \quad (54)$$

where W_i holds a linear function of R_{jt} on each side of the cap, separately by bin, and log maturity par. It also holds fixed effects for maturity bucket $\times$ year, state $\times$ year, state $\times$

offering month, rating category $\times$ year, security type $\times$ year, use of proceeds, coupon bin $\times$ maturity bucket, and callability $\times$ maturity bucket $\times$ year. The five bins are estimated jointly by exactly identified 2SLS, with standard errors clustered by issuer. The figures also report bin-by-bin reduced-form and first-stage comparisons, and Table 7 reports the joint estimates.

The design is a fuzzy discontinuity with a donut. Its identifying assumption is the exclusion restriction that, conditional on these controls, crossing the cap moves the offering yield only through designation. Because issuer-years bunch just under the cap, the half-million donut drops the issuers that manage their issuance to it. The jump is read across the donut from the local linear functions fitted on either side. The two outcomes are the offering yield in basis points and that yield divided by the bond’s expected benefit X_{tm} , also in basis points. Thus $-\beta_k$ estimates the premium in the first specification and the priced share in the second. In the terms of the three hypotheses, $-\beta_k$ is π_k . The household benchmark predicts $-\beta_k = 0$ in every bin, and the bank-only benchmark predicts a premium equal to the benefit in every bin. The model predicts a small $-\beta_k$ at one to two years that rises with maturity, so the long-minus-short contrast separates the model from both benchmarks. Appendix A.6.9 gives the construction, its relation to [Dagostino \(2025\)](#), and the checks.

Identification and results. Designation changes by 0.42–0.45 at the cap in every bin, so differential first stages do not drive the maturity pattern. Callability and issue structure are balanced at the threshold (Appendix A.6.9). The designation discontinuity follows the statutory cap’s move to \$30 million in 2009–10, while placebo discontinuities are absent or substantially smaller. Appendix A.6.9 reports the residual effects at the \$20 million placebo.

The reduced-form offering-yield change in Figure 8 is -0.5 basis point at one to two years and -8.3 at ten to twenty. Dividing by the corresponding first stages gives separate-bin premia of 1.1 and 18.6 basis points. The joint specification in Table 7 gives 1.8 and 18.4, a long-minus-short difference of 16.5 (standard error 2.7).

Dividing each yield before estimation by the benefit the bank expects over the bond’s life, X_{tm} , gives a priced share of 0.17 at one to two years and 0.41–0.52 at five to twenty. The long-minus-short difference is 0.34 (standard error 0.11). The share is a ratio of two annual yield-equivalent quantities, the premium of designation in the offering yield and the benefit the bank expects each year over the bond’s life. That benefit applies the deduction rate to the average funding cost expected along the bond’s maturity. The benefit computed from the issue year’s funding cost would be the wrong denominator, since the premium in basis points does not track it. That premium is as large in the low-rate years, when that benefit is smallest, as in any other year, and the maturity gradient holds in every rate regime. With

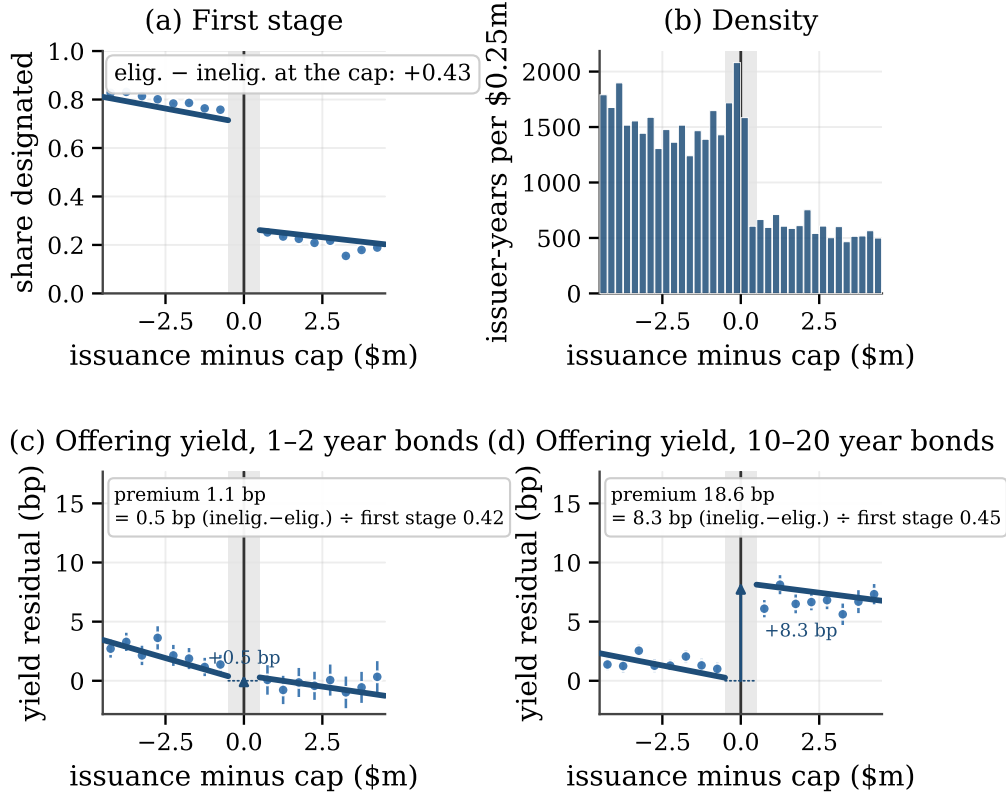


Figure 8: The bank-qualified discontinuity. Panel (a) shows the first-stage drop in designation at the statutory cap; panel (b) shows the donut around issuer bunching. Panels (c)–(d) show the reduced-form offering-yield discontinuities for 1–2 and 10–20 year bonds. Dividing by the bin-specific first stage gives premia of 1.1 and 18.6 basis points. Lines are the estimated local linear functions and bars are 95 percent intervals. Appendix Figure 31 reports the remaining maturities.

Alt text: The panels plot bank-qualified designation, issuance density, and short- and long-maturity offering yields around the statutory annual-issuance threshold. Qualification falls at the threshold; the yield discontinuity is larger for longer bonds.

the Adrian–Crump–Moench term premia in place of Kim–Wright’s, the shares are 0.17, 0.35, 0.46, 0.55 and 0.48 (Appendix A.6.9.2). Figure 9 draws the profile, the mirror image of the household profile in Figure 7.

The positive long-end premium rejects the household benchmark’s zero effect, and the maturity contrast rejects the constant capitalization of the bank-only annual-carry benchmark. Short-end capitalization is indistinguishable from zero. From five years out, the market pays about half of the benefit the bank expects over the bond’s life. Together with P1, the design goes beyond a maturity profile in yields and identifies opposing maturity profiles for benefits accruing to different investors. The model predicts this pattern for an institution-specific benefit. Because the household anchors the short boundary, a deduction that accrues only to banks is capitalized farther along the curve, where banks participate at

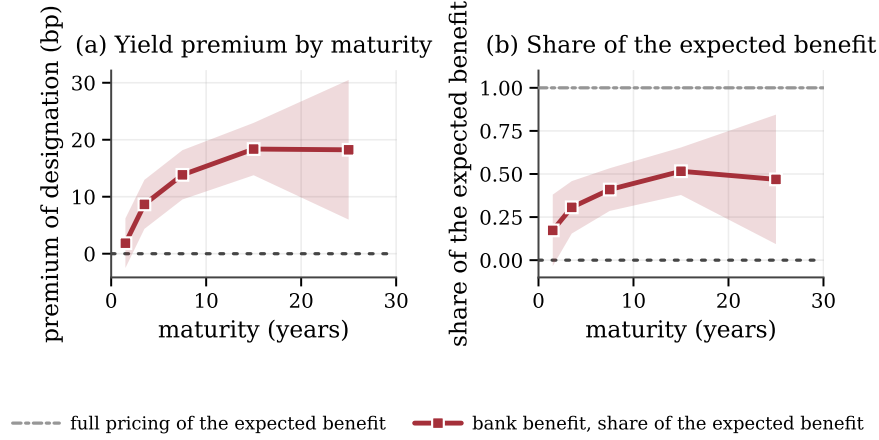


Figure 9: The maturity profile of a bank-only tax benefit. Panel (a) reports the 2SLS premium in basis points. Panel (b) uses yield divided by the benefit the bank expects over the bond’s life, X_{tm} , as the outcome, with the same sample and controls; the horizontal line at one marks full pricing of that benefit. Normalization precedes estimation: panel (b) is not panel (a) divided by a single constant. Funding costs use all banks, weighted by municipal-securities holdings, and the expected path of the short rate comes from the Kim–Wright risk-neutral yields. The bank effect is small at one to two years and larger at five to twenty; the household profile is in Figure 7. Shaded bands are 95 percent intervals.

Alt text: Bank-qualified yield premiums increase from short to longer maturities. The companion panel shows the premium as a share of the benefit the bank expects over the bond’s life, rising from near zero at one to two years to about one half from five years out and staying below the reference line for full pricing, with uncertainty bands.

the margin. The same statutory benefit therefore has a larger yield effect at long maturities than at short ones.

The long-minus-short premium remains 13.5–18.4 basis points across donuts and bandwidths. Issuer fixed effects, credit controls, general-obligation restrictions, alternative outcomes and clustering, and netting refundings preserve the shape (Appendix A.6.9). These comparisons discipline selection explanations. A tax-sensitive clientele model with another participation motive can also produce the opposing profiles. The evidence thus supports investor-specific tax transmission and does not single out liability hedging. Section 5.7 establishes why that transmission matters beyond a particular motive, and Appendix A.6.4 compares the hypotheses. Small short-end capitalization need not mean zero bank holdings.

7.3 Complementary evidence from the TCJA insurer-tax shock

The 2017 corporate rate cut supplies a second institutional comparison. Under P&C pro-ration the municipal keep stayed at about 0.9475, while the corporate keep rose from 0.65 to 0.79. The relative keep therefore rose from 0.686 to 0.834, which reduced the value of

Table 7: P2: the bank-qualified discontinuity by maturity at issuance.

	1–2y	2–5y	5–10y	10–20y	20y+	10–20 minus 1–2
First stage: designation	0.42*** (0.015)	0.42*** (0.015)	0.44*** (0.015)	0.45*** (0.017)	0.43*** (0.027)	
Premium of designation (bp)	1.8 (2.2)	8.7*** (2.2)	13.8*** (2.2)	18.4*** (2.3)	18.2*** (6.3)	16.5*** (2.7)
Priced share of the bank benefit	0.17 (0.11)	0.31*** (0.08)	0.41*** (0.06)	0.52*** (0.07)	0.47** (0.19)	0.34*** (0.11)
Bonds	355,372					
Issuers (clusters)	15,889					

Notes. 2SLS estimates of (54) on tax-exempt new issues of 2000–2021 within \$0.5 to \$4 million of the statutory cap of the year (\$30 million in 2009–10, \$10 million otherwise), with eligibility instrumenting designation bin by bin, the matched fixed effects listed in the text, and standard errors clustered by issuer. The premium is the yield of a non-designated bond minus the yield of a designated one. The priced share divides each bond’s yield by the deduction the bank expects over the bond’s life, $X_{tm} = 0.8\tau_{B,t}\hat{c}(t, m)$, before estimation, with $\hat{c}(t, m)$ the funding cost implied by the average short rate expected over the next $\min(m, 10)$ years in the issue month, the risk-neutral yield of the Kim–Wright term-structure model mapped to bank funding costs by their historical pass-through (Appendix A.6.9.2); a share of one means the market pays the bank’s full expected benefit. The last column is the long-minus-short premium, $-\beta_{10-20} + \beta_{1-2}$. Under the hypothesis that top households alone hold municipal bonds all premia are zero; under the bank-only benchmark the share is one at every maturity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

exemption. In the model, a lower insurer valuation of exemption raises yields and lowers the implied tax at the longer maturities that insurers hold, more so where insurers hold more. The representative-tax benchmark predicts no response, because insurers’ taxes play no role in it. The exposure design compares non-callable municipal transactions in states with different end-2015 P&C holdings, controlling for common time effects, maturity, issuer, resident taxes, and other bond characteristics. Appendix A.6.7 gives the specification, the domicile-based instruments, and the event study.

States with greater insurer exposure show a negative post-reform implied-tax response. The OLS level effect has a state-level wild-bootstrap p -value of 0.044. Maturity-gradient estimates are negative under OLS and both IV specifications, and event-time profiles turn negative around implementation (Appendix A.6.8). The level response has the model’s sign, for a shock to an institution’s tax that the benchmark says should not move municipal yields at all. This insurer-specific shock thus complements the household and bank-qualified maturity-profile tests.

7.4 P3. Descriptive muni-to-treasury duration response by maturity

The affine mechanism predicts that municipal yields respond more to rate movements at longer maturities, whereas the representative-tax benchmark predicts a constant relative response. For consecutive transactions in the same bond, I estimate

$$\Delta y_{b,t}^M = \beta_0 \Delta y_{m_b,t}^{\text{Tsy}} + \beta_1 (\Delta y_{m_b,t}^{\text{Tsy}} \times m_b) + \beta_2 m_b + \varepsilon_{b,t}. \quad (55)$$

The sample contains non-callable, tax-exempt bonds rated AA or above. Trades in a pair are at most ninety days apart, and the maturity-matched treasury yield moves by at least ten basis points in absolute value. Pairs are weighted by the inverse number of candidate pairs for their bond; standard errors are clustered by trade date. In (55), $\beta_0 + \beta_1 m_b$ is the municipal yield response to a treasury move for a bond with m_b years to maturity. The model predicts $\beta_1 > 0$, with the response rising from the household keep at the short end toward the long-maturity clientele mix (Proposition 3). Under the representative-tax benchmark, $\beta_1 = 0$ and the response is the same at every maturity. The matched treasury yield uses each trade’s remaining maturity, so it ages with the bond. Fixing its maturity at the first trade strengthens the gradient (Appendix A.6.10).

The estimated response is $0.346 + 0.007m_b$. Each additional year thus raises the municipal response to a one-percentage-point treasury move by 0.7 basis point. In a binned specification, the response is 0.351 below five years and 0.426 at five years and above. The gradient stays positive in all thirty combinations of trade spacing and minimum treasury move, and in the wider rating band. Appendix Table 35 and Appendix A.6.10 report the estimates and sensitivity checks. The model predicts this positive gradient, and the benchmark rules it out. The transaction responses give descriptive evidence on the maturity amplification without directly estimating a tax rate.

8 Policy Counterfactuals

A tax reform changes both the taxable holder base and the investors who supply replacement demand. The first counterfactual repeals an in-state exemption; the second moves toward equal municipal and corporate tax treatment. Both re-solve the full 45-state non-linear economy with supplies and non-tax preferences held fixed, so yields, holdings, and participation adjust jointly. A static calculation applying new taxes to old holdings would miss the resulting financing terms and cross-market spillovers.

Portfolio choice depends on relative after-tax keeps, but translating equilibrium income

into tax receipts requires absolute rates. The absolute keeps of households, P&C insurers, life insurers, and tax-exempt institutions follow the tax rules described in Section 4.1 and Appendix A.9.1.3–A.9.1.4. For the life-insurer clientele, I use the maintained effective-rate mapping

$$\xi_{I_2} = \frac{1 - \tau_{I_2}^C}{1 - \tau_{I_2}^M}, \quad \tau_{I_2}^M = 1 - \frac{1 - \tau_{I_2}^C}{\xi_{I_2}}, \quad (56)$$

with $\tau_{I_2}^C$ set to the dated corporate rate. In the baseline, $\xi_{I_2} = 0.79$ and $\tau_{I_2}^C = 35\%$ imply $\tau_{I_2}^M = 17.72\%$.

Applying these absolute rates to annualized equilibrium income gives investor-side tax receipts, including federal and all state receipts. On the issuer side, the new yields applied to the full supply give steady-state replacement-financing costs rather than changes in contractual payments on outstanding debt. Market clearing implies

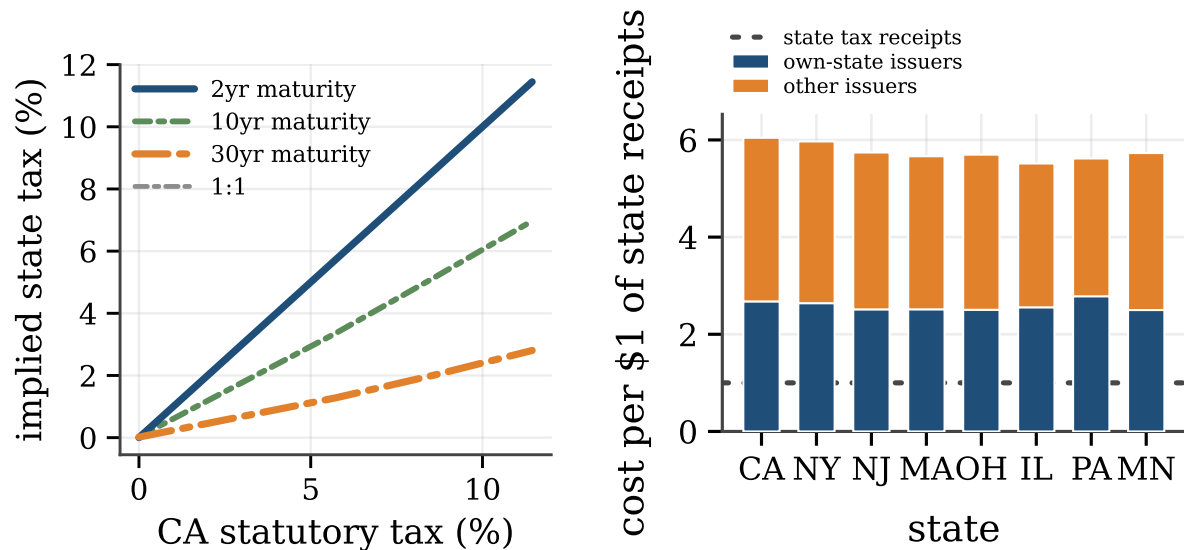
$$\begin{aligned} \Delta(\text{issuer interest flow}) &= \Delta(\text{investor-side tax receipts}) \\ &+ \Delta(\text{after-tax bond income}). \end{aligned} \quad (57)$$

The ledger uses the same yields, holdings, and absolute rates on both sides. The reporting denominator is the repealing state’s own receipts for in-state repeal and total receipts for equal treatment. The resulting ratios measure financing incidence and are neither welfare measures nor a full government budget score. Appendix A.9.2 gives the treatment of refinancing, deductions, and other provisions.

8.1 The in-state exemption and the local investor base

Repeal taxes residents’ home-state holdings but changes financing terms for all new and refinanced debt. Panel A of Figure 10 first traces transmission through separate nonlinear re-solves at different California statutory rates. This single-state exercise is distinct from the cross-state slope of Section 7.1. For each percentage point of statutory change, the implied state tax moves one-for-one at two years but only about 0.61 percentage point at ten years and 0.24 at thirty. Farther along the curve, other clienteles dilute the residents’ influence.

Panel B shows that the increase in financing costs exceeds the added receipts. Across the eight states, own-state municipal replacement-financing cost rises by \$2.50–\$2.78 per dollar of state investor-side receipts, with receipts calculated from the statutory rate and residents’ post-repeal holdings. Adding the costs borne by other municipal and corporate issuers raises the all-issuer ratio to \$5.51–\$6.04, still per dollar of the repealing state’s receipts. The gap between the two ratios measures the financing effect that the common investor base transmits beyond the repealing state’s own borrowers.



Panel A: equilibrium pass-through by maturity.

Panel B: tax-rule-based incidence.

Figure 10: Repealing the in-state exemption in equilibrium. Panel A traces separate full 45-state re-solves as California’s statutory rate changes. The implied state-tax response is one-for-one at two years but about 0.24 per percentage point at thirty years. Panel B removes each listed state’s home exemption and reports annualized issuer replacement-financing cost per dollar of the repealing state’s investor-side tax receipts. Blue denotes that state’s municipal issuers; orange denotes other municipal and corporate issuers. The numerator applies new-equilibrium yields to the full supply, so it is a steady-state replacement-financing flow rather than an immediate change in outstanding coupons.

Alt text: The first panel relates changes in the in-state exemption to equilibrium implied-tax responses at different maturities. The second compares the resulting financing-cost increases inside and outside the reforming state per dollar of added state investor-side tax receipts.

Residents reduce the positions on which the new tax is collected, so the tax base shrinks. Higher yields attract replacement investors, yet issuers pay those yields whether or not the buyer is taxable by the issuing state. Institutions absorb municipal duration by adjusting their whole portfolios, which carries the response to other municipal markets and corporate bonds.

For California, own-state yields rise by 18.6 basis points at two years and 21.8 at ten. Other-state municipal yields move by at most 3.3 basis points, and the thirty-year corporate yield rises by 0.9. Corporate borrowers bear roughly 47% of the aggregate financing increase, because even a small yield response matters on their much larger borrowing base. Static tax accounting misses both this spillover and the resident base’s contraction.

8.2 Near-elimination of the municipal tax advantage

The second counterfactual moves 99.9% toward equal municipal and corporate tax treatment for every investor. It removes household exemptions and institutional proration benefits. This near-equal-treatment endpoint is broader than federal exemption repeal alone. The exercise asks how removing the tax wedge reallocates financing costs and bond income once both markets clear again.

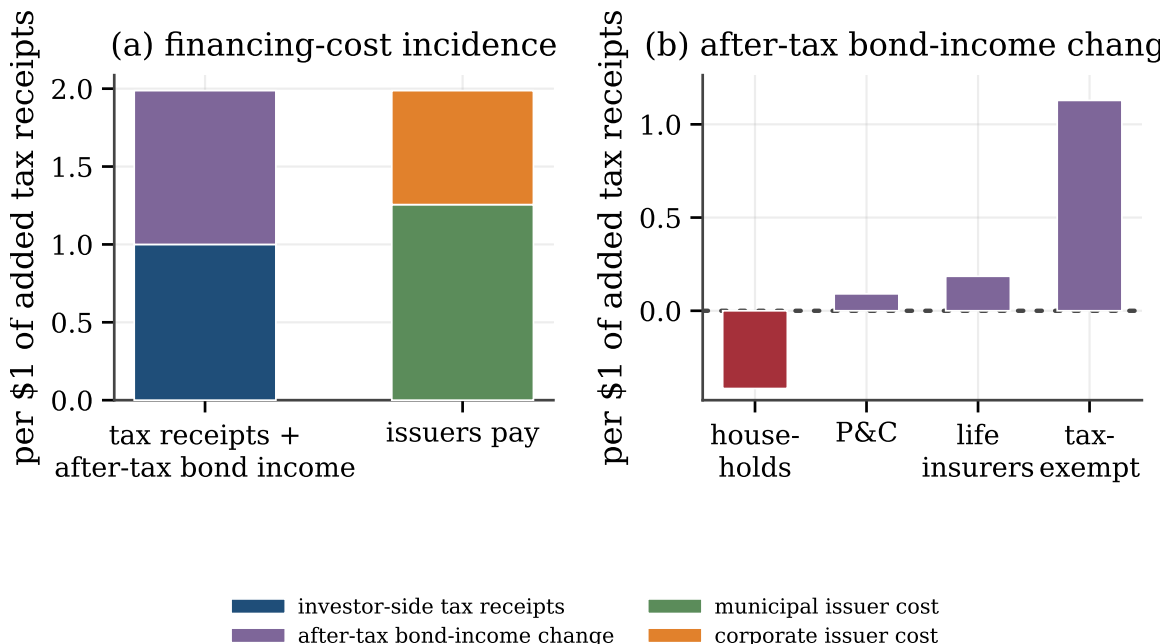


Figure 11: Near-equal municipal and corporate tax treatment. Panel (a) decomposes financing costs using (57); Panel (b) reports after-tax bond-income changes by investor group, each per dollar of added investor-side receipts. The full nonlinear re-solve moves 99.9% toward equal treatment. Absolute tax rates, replacement-financing flows, and the income measure follow the accounting defined above and in Appendix A.9.2.

Alt text: The panels summarize the near-repeal equilibrium: municipal and corporate replacement-financing costs rise relative to added investor-side tax receipts, and gains and losses differ across investor groups. These are financing and tax-accounting comparisons, not welfare measures.

Issuer replacement-financing cost rises by \$1.99 per dollar of added investor-side tax receipts, of which \$1.26 is on the municipal side and \$0.73 on the corporate side. About 37% of the financing response therefore falls on corporate borrowers, although the reform targets the municipal tax advantage. At thirty years, the municipal yield rises by 184.3 basis points and the implied tax falls by 26.45 percentage points.

Of the added receipts, 95.8% accrue federally and 4.2% to states. The remaining \$0.99

of the financing increase is after-tax bond income, with losses for households and gains for institutions (Figure 11, Panel (b)). These income changes are not welfare measures.

As the exemption loses value, households reduce their municipal positions. Institutions absorb that duration at higher yields and adjust their corporate holdings, which raises corporate financing costs as well. Equal tax treatment removes the tax wedge, while bond-specific risk and supply differences remain. The two counterfactuals show why taxable income, replacement investors, and financing terms must be determined jointly.

9 Conclusion

How taxes shape the yield curve depends on who bears risk along it. High-tax households value the exemption most, but they need not absorb duration most cheaply. Lower-tax institutions participate when their duration benefit offsets their tax disadvantage. Required returns move with that participation at the maturities institutions hold and, through households' portfolios, at shorter risky maturities they never hold. The tax treatment reflected in yields is therefore an equilibrium outcome of joint portfolio choice and market clearing.

The municipal market makes this mechanism observable. Without targeting municipal yields, the calibration accounts for 94.8 basis points, or 86%, of the thirty-year puzzle once callability and the issuance premium are removed. Separately estimated bond-side differentials cover the remainder and leave a -3.5 -basis-point residual. Household state-tax pass-through declines with maturity, while a bank-only benefit is increasingly capitalized farther out. These opposing profiles tie the longstanding puzzle to investor-specific tax treatment rather than to a representative rate.

The policy consequences reach beyond municipal debt. Repealing an in-state exemption weakens the local investor base and raises replacement-financing costs by more than the state's added investor-side receipts. Broader repeal also raises corporate borrowing costs through portfolio substitution. With portfolio risk and limited arbitrage, federal and state tax reforms, asset-specific exemptions, and cross-border withholding can reallocate duration and carry financing effects beyond the market directly taxed.

Table 8: Paired issues: the tax-exempt-versus-BAB gap widens with maturity, on both rating cuts and with or without call matching.

	AA and above (score ≤ 3)		A– and above (score ≤ 7)	
	(1) base	(2) call-matched	(3) base	(4) call-matched
<i>Panel A. Continuous maturity</i>				
Tax-exempt	−0.102*** (0.034)	−0.091** (0.036)	−0.062** (0.027)	−0.043 (0.030)
Tax-exempt $\times$ maturity	0.034*** (0.003)	0.031*** (0.003)	0.033*** (0.002)	0.030*** (0.003)
After-tax corporate yield	0.379*** (0.138)	0.376*** (0.135)	0.264* (0.134)	0.261* (0.130)
Observations	29,839	28,890	61,816	59,886
<i>Panel B. Five-year maturity bins</i>				
Tax-exempt	−0.048 (0.033)	−0.050 (0.033)	−0.007 (0.024)	−0.008 (0.024)
Tax-exempt $\times$ [5, 10) y	0.192*** (0.022)	0.191*** (0.022)	0.192*** (0.017)	0.190*** (0.017)
Tax-exempt $\times$ [10, 15) y	0.396*** (0.033)	0.364*** (0.033)	0.368*** (0.026)	0.342*** (0.030)
Tax-exempt $\times$ [15, 20) y	0.450*** (0.041)	0.390*** (0.039)	0.419*** (0.029)	0.403*** (0.039)
Tax-exempt $\times$ [20, 25) y	0.576*** (0.061)	0.540*** (0.066)	0.593*** (0.071)	0.510*** (0.066)
Tax-exempt $\times$ ≥ 25 y	0.755*** (0.079)	0.639*** (0.039)	0.996*** (0.074)	0.804*** (0.060)
After-tax corporate yield	0.380*** (0.138)	0.378*** (0.135)	0.264* (0.134)	0.262** (0.130)
Observations	29,839	28,890	61,816	59,886
Issuer $\times$ maturity $\times$ quarter FE	Yes	Yes	Yes	Yes
$\times$ call status	No	Yes	No	Yes

Notes. Offering yields, 2009Q2–2010Q4; AMT bonds excluded; BAB yields multiplied by 0.65. The omitted bin is [1, 5) years. Standard errors clustered by state. Call-matched columns add the optional-call flag to the absorbed issuer-by-maturity-by-quarter cell. Columns (1)–(2) use AA-and-above bonds and columns (3)–(4) A-or-above bonds. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

- Babina, T., Jotikasthira, C., Lundblad, C., and Ramadorai, T. (2021). Heterogeneous taxes and limited risk sharing: Evidence from municipal bonds. *The Review of Financial Studies*, 34(1):509–568.
- Brown, D. P. and Sirmans, C. F. (2011). Maturity clienteles in the municipal bond market: Term premiums and the muni puzzle. Working paper.
- Brunetti, C., Foley-Fisher, N., and Verani, S. (2023). Measuring interest rate risk management by financial institutions. Finance and Economics Discussion Series 2023-067, Board of Governors of the Federal Reserve System.
- Chalmers, J. M. (1998). Default risk cannot explain the muni puzzle: Evidence from municipal bonds that are secured by US Treasury obligations. *The Review of Financial Studies*, 11(2):281–308.
- Coppola, A. (2021). In safe hands: The financial and real impact of investor composition over the credit cycle. Working paper, Harvard University.
- Dagostino, R. (2025). The impact of bank financing on municipalities’ bond issuance and the real economy. *Journal of Financial Economics*, 166:104022.
- Fama, E. F. (1977). A pricing model for the municipal bond market. Unpublished manuscript, University of Chicago.
- Feenberg, D. R. and Poterba, J. M. (1991). Which households own municipal bonds? evidence from tax returns. *National Tax Journal*, 44(4):93–103.
- Fleckenstein, M. and Longstaff, F. A. (2025). Do municipal bond investors pay a convenience premium to avoid taxes? *The Review of Financial Studies*, page hhaf100.
- Green, R. C. (1993). A simple model of the taxable and tax-exempt yield curves. *Review of Financial Studies*, 6(2):233–264.
- Greenwood, R., Hanson, S., Stein, J. C., and Sunderam, A. (2023). A quantity-driven theory of term premia and exchange rates. *The Quarterly Journal of Economics*, 138(4):2327–2389.
- Gromb, D. and Vayanos, D. (2018). The dynamics of financially constrained arbitrage. *The Journal of Finance*, 73(4):1713–1750.
- Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The U.S. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304.
- Hendershott, P. H. and Koch, T. W. (1980). The demand for tax-exempt securities by financial institutions. *Journal of Finance*, 35(3):717–727.
- Hines, James R., J. and Horton, E. (2024). Banks and tax-exempt debt arbitrage. Working Paper 32647, National Bureau of Economic Research.

- Kidwell, D. S. and Koch, T. W. (1983). Market segmentation and the term structure of municipal yields. *Journal of Money, Credit and Banking*, 15(1):40–55.
- Koijen, R. S. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515.
- Li, Z. (2026). Long rates, life insurers, and credit spreads. *The Review of Financial Studies*. Advance access, August 31.
- Livingston, M. (1982). The pricing of municipal bonds. *Journal of Financial and Quantitative Analysis*, 17(2):179–193.
- Longstaff, F. A. (2011). Municipal debt and marginal tax rates: Is there a tax premium in asset prices? *The Journal of Finance*, 66(3):721–751.
- Miller, M. H. (1977). Debt and taxes. *The Journal of Finance*, 32(2):261–275.
- Mussa, M. L. and Kormendi, R. C. (1979). *The Taxation of Municipal Bonds: An Economic Appraisal*. Number 229 in AEI Studies in Tax Policy. American Enterprise Institute for Public Policy Research, Washington, DC.
- Pirinsky, C. and Wang, Q. (2011). Market segmentation and the cost of capital in a domestic market: Evidence from municipal bonds. *Financial Management*, 40(3):455–481.
- Schwert, M. (2017). Municipal bond liquidity and default risk. *The Journal of Finance*, 72(4):1683–1722.
- Siriwardane, E., Sunderam, A., and Wallen, J. L. (2022). Segmented arbitrage. Technical report, National Bureau of Economic Research.
- Skelton, J. L. (1983). Banks, firms and the relative pricing of tax-exempt and taxable bonds. *Journal of Financial Economics*, 12(3):343–355.
- Vayanos, D. and Vila, J.-L. (2021). A preferred-habitat model of the term structure of interest rates. *Econometrica*, 89(1):77–112.

A.1 Proofs of theoretical results

A.1.1 Exact equilibrium conditions and the analytical boundary

The floorless, fixed-support affine construction used for Results 1–3 supplies the mechanism. It is neither a uniqueness theorem nor an error bound for the nonlinear economy of §5.2. The strict concavity that positive idiosyncratic variances give on the risky coordinates yields neither equilibrium uniqueness nor the Jacobian nonsingularity assumed for derivatives. Throughout the analytical two-investor results, $0 < \rho < 1$, $\sigma_r > 0$, $m_i, \gamma_i > 0$, $0 < \xi_H < \xi_I$, supplies are positive on traded coordinates, the one-period corporate claim is the cash account, and $v_{a,1} = 0$ in every market, where $v_{a,T} \equiv \sigma_{\eta,T}^2$ denotes the bond-specific variance. Positive bond-specific variances are retained where strictly concave risky demands and coupled participation require them, without imposing a uniform idiosyncratic-to-rate variance bound. An investor is eligible for a coordinate when no predetermined trading restriction bars the holding.

Index risky market–maturity coordinates by $j = (a, T)$, write $v_j = (\sigma_{\eta,T}^a)^2$, $K_i = \text{diag}(\xi_i^a)$, and $d_{ij} = \xi_i^C / \xi_i^a$, so that $d_{i,M} = \xi_i$ and $d_{i,C} = 1$ in the one-municipal-market normalization. With $p_{j,-} = p_{T-1}^a$, the exact covariance and the liability credit are

$$\Sigma_{jk}(z) = \text{Cov}_z(p_{j,-}(Z'), p_{k,-}(Z')) + \mathbf{1}_{\{j=k\}} v_j, \quad c_{ij}^L(z) = -\text{Cov}_z(p_{j,-}(Z'), \ell_i \sigma_r \varepsilon'). \quad (58)$$

Dividing the first-order condition by the positive asset keep gives

$$\begin{aligned} \nu_{ij} := d_{ij}r + \gamma_i[\Sigma K_i x_i - c_i^L]_j - \mu_j &\geq 0, & x_{ij} &\geq 0, & x_{ij}\nu_{ij} &= 0, \\ \sum_i m_i x_{ij} &= Q_j. \end{aligned} \quad (59)$$

Concavity of the maintained objective (19) makes these conditions sufficient for each investor.

Whenever holdings are differentiated below, the price moments are continuously differentiable, complementarity is strict on the risky coordinates, and the stacked active KKT-and-clearing Jacobian is nonsingular. The implicit-function theorem then gives a locally fixed support with differentiable holdings and clearing returns. In particular,

$$\sum_{i \in \mathcal{U}_j} m_i \frac{\partial x_{ij}}{\partial z} = 0. \quad (60)$$

Proposition 4 (Deterministic one-period boundary). *Suppose the one-period municipal claim has $v_{M,1} = 0$, its supply is positive, $r(z) > 0$, and the household is eligible with the smallest keep among eligible investors. Then only eligible minimal-keep investors can*

hold it, and

$$y_1^M = \mu_1^M = \xi_H r(z), \quad y_1^C = r(z), \quad \tau_1^{Y, \text{impl}} = \tau_H. \quad (61)$$

If the household is the unique keep minimizer, its tax passes through one-for-one and a strictly higher-keep insurer's tax has zero effect at this boundary. At a tie the household derivative requires perturbing all tied minimal keeps together. At $r = 0$ the yield ratio is undefined.

Proof. Since $p_0^a = 0$ and $v_{a,1} = 0$, the boundary return has zero covariance with every portfolio and liability. The municipal KKT conditions reduce to $\mu_1^M \leq \xi_i r$, with equality for any positive holder. Positive supply implies $\mu_1^M = \min_{i \in \mathcal{E}_{M,1}} \xi_i r = \xi_H r$. The price recursion gives $p_1^M = -\mu_1^M$; the corporate cash claim has $p_1^C = -r$. Differentiating their ratio proves the tax statements. $\square$

A.1.2 Affine closure: what is fixed and what must be verified

Fix a support, set inactive positions to zero, replace the cash floor by $r(z) = \bar{r} + z$, and use affine continuation prices on all of $\mathbb{R}$. Active equalities are solved jointly across maturities; nonnegative holdings and inactive inequalities are checked afterwards.

Proposition 5 (Conditional affine construction). *For affine candidates $p_T^a(z) = -A_T^a z - C_T^a$ under the floorless rule $r(z) = \bar{r} + z$, suppose the active KKT-and-clearing matrix is nonsingular. The price update preserves the affine class. At a coefficient fixed point, all active equalities, clearing equations, and exact continuation equations hold. Its holdings and inactive slacks are affine, so the set on which all KKT inequalities hold is an interval, possibly empty. This is a current-state validity interval conditional on the global affine continuation, not an equilibrium of the unrestricted Gaussian economy confined to that interval.*

Proof. Gaussian integration gives

$$\log \mathbb{E}_z e^{p_{T-1}^a(Z')} = -\rho A_{T-1}^a z - C_{T-1}^a + \frac{1}{2} \sigma_r^2 (A_{T-1}^a)^2. \quad (62)$$

The covariance $\Sigma = \sigma_r^2 A A' + \text{diag}(v)$, where A collects predecessor loadings, and $c_i^L = \ell_i \sigma_r^2 A$ are constant in the current state, so the active system has a constant coefficient matrix whose only state-dependent term is $d_{ij}(\bar{r} + z)$. Its nonsingularity gives affine holdings and μ , and substitution in the price recursion returns affine prices. At a coefficient fixed point the equations hold identically, and the loading equations iterated from $A_0^a = 0$ are

$$A_T^a = \rho A_{T-1}^a + s_T^a = \sum_{j=1}^T \rho^{T-j} s_j^a, \quad s_T^a = \frac{\partial \mu_T^a}{\partial z}. \quad (63)$$

Each reinstated holding or slack inequality defines a half-line, all of $\mathbb{R}$, or the empty set; their intersection is the stated interval. $\square$

If the KKT inequalities hold at every state, this verifies an equilibrium of the constrained floorless economy. Feasibility on an interval alone does not, because the Gaussian transition reaches its complement from every current state. With a deterministic one-period muni and heterogeneous keeps, the household-only boundary has insurer slack $(\xi_I - \xi_H)(\bar{r} + z)$, which fails at negative rates. Reinstating the cash floor changes the continuation problem.

Proposition 6 (Conditional coefficient existence). *Let $\mathcal{F}$ be the affine coefficient update for a fixed support, with nonsingular active systems throughout a nonempty compact convex coefficient set B . If $\mathcal{F}$ is continuous and $\mathcal{F}(B) \subseteq B$, it has a fixed point. If all reinstated KKT inequalities hold on the desired current-state region for every candidate in B , the fixed point is feasible there conditional on its affine continuation. A contraction of this entire update on B gives uniqueness within B only.*

Proof. Brouwer gives a coefficient fixed point; the stipulated inequalities verify its KKT feasibility. Under contraction, two fixed points have distance at most the contraction modulus times that same distance and hence coincide. $\square$

A.1.3 Participation: benchmark incentive and equilibrium support

Assumption 1 (Household duration burden at the benchmark). *At the candidate allocation where the household holds all munis and the insurer all corporates,*

$$\hat{D}_H^{\text{hh}} = \sum_T A_{T-1}^M Q_T^M / m_H - \ell_H, \quad \hat{D}_I^{\text{hh}} = \xi_I \sum_T A_{T-1}^C Q_T^C / m_I - \ell_I, \quad \Lambda := \gamma_H \hat{D}_H^{\text{hh}} - \gamma_I \hat{D}_I^{\text{hh}} > 0.$$

Proof of Proposition 1. Substituting the household's active municipal equality

$$\mu_T^{M,\text{hh}} = \xi_H r + \gamma_H \sigma_r^2 A_{T-1}^M \hat{D}_H^{\text{hh}} + \gamma_H v_{M,T} Q_T^M / m_H$$

into the insurer's zero-holding condition gives the insurer slack

$$\nu_{I,T}^{\text{hh}} = (\xi_I - \xi_H)r - \sigma_r^2 A_{T-1}^M \Lambda - \gamma_H v_{M,T} Q_T^M / m_H, \quad (64)$$

nonpositive where the insurer weakly prefers to participate and negative where a positive position is strictly profitable; in the rate-only case it is (32). Benchmark holdings are state

invariant, so $s_T^M = \xi_H$ and $A_T^M = \xi_H \sum_{q=0}^{T-1} \rho^q$ is strictly increasing. If $v_{M,T}Q_T^M$ is nondecreasing, (64) decreases in maturity and the profitable-participation set is an upper interval. In the rate-only case, dividing by $\sigma_r^2\Lambda > 0$ gives $\nu_{I,T}^{\text{hh}} \leq 0$ if and only if $A_{T-1}^M \geq (\xi_I - \xi_H)r/(\sigma_r^2\Lambda)$. With strictly increasing loadings this is (33), at least two for $r > 0$ because $A_0^M = 0$. The inventory term weakly expands the set at fixed benchmark loadings and exposures, without comparing re-solved equilibria.

At given affine prices in the rate-only benchmark, buying one muni and selling $A_{T-1}^M/(\xi_I A_{T-1}^C)$ corporates preserves after-tax duration and gains

$$\mathcal{G}_i = \mu_T^M - \xi_i r - \frac{A_{T-1}^M}{A_{T-1}^C}(\mu_T^C - r)$$

per unit. Hence $\mathcal{G}_I - \mathcal{G}_H = -(\xi_I - \xi_H)r$ at every maturity, because the corporate keep cancels against the duration-matching quantity. At the household-only benchmark the insurer's corporate KKT equality gives $\mathcal{G}_I = -\nu_{I,T}^{\text{hh}}$ without idiosyncratic risk. The household generally cannot make this trade, which shorts corporates.

Theorem 1 (Upper-interval insurer participation in the affine equilibrium). *At any state of a two-investor affine solution with positive municipal supplies, suppose $v_{M,T} > 0$ on the risky maturities considered. Define the equilibrium, rather than benchmark, duration gap $\hat{\Lambda} = \gamma_H \hat{D}_H - \gamma_I \hat{D}_I$ and*

$$\Psi_T = (\xi_I - \xi_H)r - \sigma_r^2 A_{T-1}^M \hat{\Lambda} - \gamma_H v_{M,T} Q_T^M / m_H. \quad (65)$$

Then $x_{I,M}(T) > 0$ if and only if $\Psi_T < 0$. If $\hat{\Lambda} > 0$, the predecessor loadings are nondecreasing, and $v_{M,T}Q_T^M$ is nondecreasing, the insurer's positive support is an upper interval, possibly empty. This identifies the cutoff within a given equilibrium, not uniqueness across equilibria.

Proof. If the insurer is inactive, clearing gives $x_{H,M}(T) = Q_T^M / m_H$; substituting the household equality into the insurer inequality gives $\Psi_T = \nu_{I,T} \geq 0$. If both are active, subtracting the insurer equality from the household equality and clearing give

$$\Psi_T = -v_{M,T}(\gamma_I + \gamma_H m_I / m_H) x_{I,M}(T) < 0.$$

If the insurer alone is active, its equality and the household inequality imply $(\xi_I - \xi_H)r - \sigma_r^2 A_{T-1}^M \hat{\Lambda} \leq -\gamma_I v_{M,T} Q_T^M / m_I < 0$, so again $\Psi_T < 0$. This proves the equivalence. Finally,

$$\Psi_{T+1} - \Psi_T = -\sigma_r^2 (A_T^M - A_{T-1}^M) \hat{\Lambda} - \frac{\gamma_H}{m_H} (v_{M,T+1} Q_{T+1}^M - v_{M,T} Q_T^M) \leq 0.$$

The first strictly negative index is the positive-support cutoff; equality is a zero-holding boundary. $\square$

With $R_T = (\xi_I - \xi_H)r - \sigma_r^2 A_{T-1}^M \hat{\Lambda}$, $K = \gamma_H m_I + \gamma_I m_H$ and $v_T := v_{M,T} > 0$, subtracting the two active KKT equalities and clearing give

$$x_{H,M}(T) = \frac{\gamma_I Q_T^M + m_I R_T / v_T}{K}, \quad x_{I,M}(T) = \frac{\gamma_H Q_T^M - m_H R_T / v_T}{K}. \quad (66)$$

Both holdings are nonnegative when $-\gamma_I Q_T^M / m_I \leq R_T / v_T \leq \gamma_H Q_T^M / m_H$. This fixes one maturity given common durations, which must still equal the resulting whole-portfolio exposures. With a nonzero tax penalty and distinct positive limiting loadings, no fixed block of two or more jointly active maturities survives every $v_T \downarrow 0$. Bounded holdings would need $R_T \rightarrow 0$ at each maturity, which assigns different values to one common duration gap. The positive-variance coupled formula below therefore describes no feasible joint-block equilibrium in that singular limit.

At fixed benchmark loadings and liabilities, raising ξ_I raises the rate-only cut-off loading $(\xi_I - \xi_H)r / (\sigma_r^2 \Lambda)$ for $r > 0$ and $\Lambda > 0$. The numerator rises, and $\partial \Lambda / \partial \xi_I = -\gamma_I \sum_T A_{T-1}^C Q_T^C / m_I < 0$. This isolates a benchmark incentive at fixed loadings and liabilities, since the average-maturity derivative in general equilibrium requires re-solved exposures throughout. The conditional average-maturity comparison in Result 1 assumes positive corporate holdings at every maturity on the municipal support, some corporate mass below the cutoff, and a muni-to-corporate holding ratio nondecreasing above it, a condition participation alone does not deliver. Truncation to the municipal support strictly raises the corporate mean and reweighting by the nondecreasing ratio cannot lower it, since $2 \text{Cov}(J, w(J)) = \mathbb{E}[(J - J')(w(J) - w(J'))] \geq 0$ for independent copies J, J' , so the municipal mean is strictly greater.

A.1.4 The two pricing channels from active KKT equations

Proposition 7 (Exact active-clientele aggregation). *At a regular state, let $\Gamma_j = \sum_{i \in \mathcal{U}_j} m_i / \gamma_i$, $\bar{d}_j = \Gamma_j^{-1} \sum_{i \in \mathcal{U}_j} m_i d_{ij} / \gamma_i$, and $\mathcal{P}_j = \Gamma_j^{-1} \sum_{i \in \mathcal{U}_j} m_i [\Sigma K_i x_i - c_i^L]_j$. Then*

$$\begin{aligned} \mu_j &= \bar{d}_j r + \mathcal{P}_j, \\ s_j &= \bar{d}_j r' + \frac{1}{\Gamma_j} \sum_{i \in \mathcal{U}_j} m_i [(\partial_z \Sigma) K_i x_i + \Sigma K_i \partial_z x_i - \partial_z c_i^L]_j. \end{aligned} \quad (67)$$

Above the cash floor the municipal-taxable difference is

$$s_T^M - \xi_H s_T^C = (\bar{\xi}_T^M - \xi_H) + \partial_z \mathcal{P}_T^M - \xi_H \partial_z \mathcal{P}_T^C. \quad (68)$$

Proof. Multiply each active investor's condition $\mu_j - d_{ij}r = \gamma_i[\Sigma K_i x_i - c_i^L]_j$ by m_i/γ_i , add, and divide by Γ_j . Support and weights are locally fixed, so the product rule gives the second line. In the one-market normalization $\bar{d}_{M,T} = \bar{\xi}_T^M$ and $\bar{d}_{C,T} = 1$; subtraction gives the gap. $\square$

Affine specialization: proof of Result 2. Let $\hat{D}_i = A'K_i x_i - \ell_i$. With constant affine covariance and liability credits, (67) becomes

$$s_j = \bar{d}_j + \frac{\sigma_r^2 A_j}{\Gamma_j} \sum_{i \in \mathcal{U}_j} m_i \frac{\partial \hat{D}_i}{\partial z} + \frac{v_j}{\Gamma_j} \sum_{i \in \mathcal{U}_j} m_i \xi_i^a \frac{\partial x_{ij}}{\partial z}, \quad (69)$$

where $A_j = A_{T-1}^a$ is the predecessor loading. For munis in the one-market normalization the final term vanishes by (60), leaving the direct and portfolio-exposure channels of Result 2.

Proposition 8 (Corporate counterpart). *The corporate version of (69) has direct term one. Its idiosyncratic term vanishes if the corporate active set is a singleton, all active corporate keeps are equal, or $v_{C,T} = 0$. In particular, with the insurer the sole corporate holder, $s_T^C = 1 + \gamma_I \sigma_r^2 A_{T-1}^C \partial_z \hat{D}_I$.*

Proof. Set $d_{i,C} = 1$ in (69). A sole holder's quantity is fixed by supply. Common keeps factor out of the inventory sum, which then vanishes by (60). Zero variance removes the term directly. With heterogeneous keeps and several active holders, unweighted clearing alone does not cancel it. $\square$

Proposition 9 (Joint participation and propagation across maturities). *At a regular state with $r(z) > 0$ in the two-investor affine benchmark, suppose the insurer alone holds risky corporates. Let $\emptyset \neq J \subseteq \{2, \dots, T_{\max}\}$ be the fixed set of municipal maturities jointly held by both investors; the household alone holds every municipal maturity outside J , including the riskless one-period claim. Assume $A_{T-1}^M, A_{T-1}^C > 0$ for every $T \geq 2$ and $v_{M,T} > 0$ on J . Define*

$$P_J = \sum_{T \in J} \frac{A_{T-1}^M}{v_{M,T}}, \quad S_J = \sum_{T \in J} \frac{(A_{T-1}^M)^2}{v_{M,T}}, \quad d := \frac{\partial \hat{D}_H}{\partial z}.$$

Then

$$d = \frac{(\xi_I - \xi_H)m_I P_J}{(\gamma_H m_I + \gamma_I m_H)(1 + \sigma_r^2 S_J)} > 0, \quad \frac{\partial \hat{D}_I}{\partial z} = -\frac{m_H}{m_I} d. \quad (70)$$

Writing $\omega_i = (m_i/\gamma_i)/(m_H/\gamma_H + m_I/\gamma_I)$, $\xi^* = \omega_H\xi_H + \omega_I\xi_I$, and $b = \gamma_I\sigma_r^2(m_H/m_I)d$, the slopes are

$$s_T^M = \begin{cases} \xi_H + \gamma_H\sigma_r^2 A_{T-1}^M d, & T \notin J, \\ \xi^*, & T \in J, \end{cases} \quad s_T^C = 1 - bA_{T-1}^C. \quad (71)$$

Consequently, $s_T^M - \xi_H s_T^C > 0$ and $A_T^M - \xi_H A_T^C > 0$ at every risky maturity $T \geq 2$, with equality in both comparisons at $T = 1$.

Proof. Clearing fixes each sole-held position, Q_T^M/m_H for the household outside J and Q_T^C/m_I for the insurer in risky corporates, so only jointly held muni positions vary with the state. Liability loadings are constant, and the fixed corporate book adds no exposure derivative. Differentiating clearing and summing against the municipal predecessor loadings gives

$$m_H \partial_z \widehat{D}_H + m_I \partial_z \widehat{D}_I = 0. \quad (72)$$

At each $T \in J$, subtracting the differentiated insurer municipal equality from the household's gives

$$\xi_I - \xi_H = \sigma_r^2 A_{T-1}^M (\gamma_H \partial_z \widehat{D}_H - \gamma_I \partial_z \widehat{D}_I) + v_{M,T} (\gamma_H \partial_z x_{H,M}(T) - \gamma_I \partial_z x_{I,M}(T)).$$

With $u_T = \partial_z x_{I,M}(T)$ and $k = (\gamma_H m_I + \gamma_I m_H)/m_H$, clearing reduces this to $u_T = -(\xi_I - \xi_H)/(k v_{M,T}) - \sigma_r^2 A_{T-1}^M \partial_z \widehat{D}_I / v_{M,T}$. Since $\partial_z \widehat{D}_I = \sum_{T \in J} A_{T-1}^M u_T$, multiplying by A_{T-1}^M and summing over J gives

$$(1 + \sigma_r^2 S_J) \partial_z \widehat{D}_I = -\frac{\xi_I - \xi_H}{k} P_J < 0,$$

and combining this with (72) gives (70). At a household-only maturity the holding derivative is zero, which gives the first municipal slope. At a jointly held maturity, risk-tolerance-weighted aggregation cancels both the aggregate exposure derivative and the inventory derivative, giving ξ^* . The corporate formula is Proposition 8.

For propagation, write $g_T := s_T^M - \xi_H s_T^C$. Equation (71) gives

$$g_T = \begin{cases} \sigma_r^2 d \left[\gamma_H A_{T-1}^M + \xi_H \gamma_I \frac{m_H}{m_I} A_{T-1}^C \right], & T \notin J, \\ \omega_I (\xi_I - \xi_H) + \xi_H b A_{T-1}^C, & T \in J. \end{cases}$$

Thus $g_T > 0$ for every $T \geq 2$, and $g_1 = 0$ because both predecessor loadings vanish. Iterating

(63) with $0 < \rho < 1$,

$$A_T^M - \xi_H A_T^C = \sum_{j=2}^T \rho^{T-j} g_j > 0, \quad T \geq 2.$$

□

Suppose an investor's objective includes $\sum_T \kappa_{iT} x_{i,M}(T)$, with κ_{iT} independent of the rate state (a state-invariant regulatory holding benefit, say). Her active municipal condition becomes

$$\mu_T^M - \xi_i r + \kappa_{iT} = \gamma_i \left[\sigma_r^2 A_{T-1}^M \hat{D}_i + v_{M,T} x_{i,M}(T) \right].$$

With covariances, the financing rule, and constraints fixed, differentiation eliminates κ_{iT} and any constant liability loading. The preceding proof and the loading fixed-point equations are therefore unchanged along a given branch on the same feasible joint-holding set. Equality across distinct coefficient fixed points would additionally require uniqueness. A state-dependent benefit contributes its derivative to the KKT equations, while a fixed participation fee changes the participation decision itself and adds no per-unit KKT term.

Uniqueness of the loading fixed point on a proposed branch. Given d , generate the municipal loadings from (71) and the loading recursion, and let $\mathcal{F}_J(d)$ be the right side of (70) at those loadings. If $\mathcal{F}_J$ maps $[0, \bar{d}]$ into itself and $\sup |\mathcal{F}'_J| < 1$ on the whole interval, Banach's theorem gives exactly one exposure fixed point, hence one municipal and corporate loading path on this branch. Active holdings and inactive inequalities must still be verified, and another feasible cutoff is not excluded. The derivative is

$$\mathcal{F}'_J(d) = c \frac{P'_J(1 + \sigma_r^2 S_J) - \sigma_r^2 P_J S'_J}{(1 + \sigma_r^2 S_J)^2}, \quad c = \frac{(\xi_I - \xi_H)m_I}{\gamma_H m_I + \gamma_I m_H},$$

with P'_J, S'_J from differentiating the same loading recursions.

Proposition 10 (No-participation loading benchmark). *In the two-investor affine benchmark, if the household alone holds every muni and the insurer alone holds every risky corporate, then $s_T^M = \xi_H$, $s_T^C = 1$, and $A_T^M = \xi_H A_T^C$ at every maturity.*

Proof. Clearing fixes every holding, so both duration derivatives are zero. Apply (69) and iterate (63). This is a loading identity; different intercept risk premia can still affect the yield ratio. □

A.1.5 From slope ordering to loading and yield ratios

On the upper-interval specialization $J = \{T^*, \dots, T_{\max}\}$ of Proposition 9, two sufficient conditions give the maturity ordering:

$$(S') \quad \frac{\gamma_I m_H (\xi_I - \xi_H)}{\xi_H (\gamma_H m_I + \gamma_I m_H)} \leq \rho, \quad (73)$$

$$(M) \quad A_{T^*-1}^M \leq \frac{\xi^*}{1 - \rho}. \quad (74)$$

(S') controls corporate loading feedback; (M) keeps the household branch from overshooting the long-muni loading fixed point. Both are conditions on the segmented affine benchmark; establishing that this support exists requires a separate argument.

Proof of Result 3(i). On the household branch $A_T^M = (\rho + \gamma_H \sigma_r^2 d) A_{T-1}^M + \xi_H$, so A_T^M increases from zero. On the joint branch $A_T^M = \rho A_{T-1}^M + \xi^*$, whose increments have the sign of $\xi^* - (1 - \rho) A_{T^*-1}^M$, nonnegative under (M). Since $A_{T-1}^M \geq \xi_H$ on J , $P_J/S_J \leq 1/\xi_H$, and (70) with (S') gives $0 \leq b \leq \rho$. Therefore

$$A_T^C = \sum_{q=0}^{T-1} (\rho - b)^q, \quad s_T^C = 1 - b A_{T-1}^C \downarrow \frac{1 - \rho}{1 - \rho + b} > 0.$$

The municipal slope rises on the household branch and is constant on the joint branch, with no downward jump when insurer participation begins. By (70), the required inequality $\omega_I(\xi_I - \xi_H) \geq \gamma_H \sigma_r^2 A_{T^*-2}^M d$ is equivalent to $\sigma_r^2 A_{T^*-2}^M P_J \leq 1 + \sigma_r^2 S_J$, which follows from $A_{T-1}^M > A_{T^*-2}^M$ for every $T \in J$, shown above from (M).

For positive b_j and nondecreasing $r_j = a_j/b_j$, the discounted ratio

$$R_T = \frac{\sum_{j \leq T} \rho^{T-j} a_j}{\sum_{j \leq T} \rho^{T-j} b_j}$$

is a positive-weight average of $r_1, \dots, r_T$, hence nondecreasing:

$$R_{T+1} - R_T = \frac{b_{T+1}(r_{T+1} - R_T)}{\rho \sum_{j \leq T} \rho^{T-j} b_j + b_{T+1}} \geq 0.$$

With $a_j = s_j^M$, $b_j = s_j^C$ and (63), this is the loading-ratio ordering; the first ratio is ξ_H . Extended to arbitrarily long maturities with fixed d and eventual joint participation, the branch recursions give $A_\infty^M = \xi^*/(1 - \rho)$ and $A_\infty^C = 1/(1 - \rho + b)$, hence the endpoint in (40); the extension is not an existence assertion for an infinite-supply economy. Taking $b_j = 1$

gives the treasury-reference result.

Proposition 11 (Exact forward-ratio bridge). *For any price functions, affine or nonlinear, let $f_T^a(z) = -p_T^a(z) + p_{T-1}^a(z)$. Fix a horizon N . If $f_j^C(z) > 0$ for every $1 \leq j \leq N$ and $f_j^M(z)/f_j^C(z)$ is nondecreasing on $j = 1, \dots, N$, then, over that full range, $y_T^M(z)/y_T^C(z)$ is nondecreasing and $\tau_T^{Y, \text{impl}}(z)$ is nonincreasing.*

Proof. With $p_0^a \equiv 0$, telescoping gives $Ty_T^a = -p_T^a = \sum_{j \leq T} f_j^a$, so the yield ratio is a weighted average of the forward ratios with weights $f_j^C > 0$ and cannot fall when the added forward ratio is at least every previous one. On a two-year grid T indexes steps and the common step-to-year conversion cancels from the ratio. $\square$

Affine forward levels and Result 3(iii). For coordinate $j = (a, T)$, put $\Gamma_T^a := \sum_{i \in \mathcal{U}_j} m_i/\gamma_i$, $d_{i,a} := d_{ij}$, and

$$c_T^a = \frac{\sum_i (m_i/\gamma_i) d_{i,a}}{\Gamma_T^a}, \quad \pi_T^a = \frac{\sigma_r^2}{\Gamma_T^a} \sum_i m_i \hat{D}_i(0), \quad \iota_T^a = v_j \left[\frac{\sum_i m_i \xi_i^a x_{i,a}(T, 0)}{\Gamma_T^a} - \frac{1}{2} \right],$$

with sums over active holders. At $z = 0$, aggregating the level KKT equations and subtracting the two Jensen terms in the exact Gaussian mean gives

$$C_T^a - C_{T-1}^a = c_T^a \bar{r} + A_{T-1}^a \pi_T^a - \frac{1}{2} \sigma_r^2 (A_{T-1}^a)^2 + \iota_T^a. \quad (75)$$

For munis the weighted holding sum is Q_T^M ; for corporates heterogeneous keeps remain inside it. If the displayed coefficients and loadings converge and the limiting corporate forward f_∞^C is positive, Cesàro convergence gives

$$y_T^a(z) \longrightarrow f_\infty^a = c_\infty^a \bar{r} + A_\infty^a \pi_\infty^a - \frac{1}{2} \sigma_r^2 (A_\infty^a)^2 + \iota_\infty^a \quad (76)$$

at every fixed z , since C_T^a/T converges to the forward limit and $A_T^a z/T \rightarrow 0$; the implied-tax limit is $1 - f_\infty^M/f_\infty^C$. Risk-tolerance-weighted aggregation on the joint muni branch and the sole-insurer corporate branch gives $\pi_\infty^M - \pi_\infty^C = \omega_H \sigma_r^2 (\gamma_H \hat{D}_H(0) - \gamma_I \hat{D}_I(0))$, which is (42).

A.1.6 Tax changes: direct component and equilibrium response

Proposition 12 (Household and insurer direct pass-through). *Define the loading-implied tax $\tau_T^A := 1 - A_T^M/A_T^C$. In the segmented affine benchmark, (i) $\partial \tau_1^A/\partial \tau_H = 1$ and $\partial \tau_1^A/\partial \tau_I = 0$. (ii) Define the direct component by switching off exposure-response terms in the pricing*

recursions and holding the support and risk-tolerance weights fixed. With

$$G_T = \sum_{j=1}^T \rho^{T-j}, \quad W_S(T) = \frac{\sum_{j=1}^{\min(T, T^*-1)} \rho^{T-j}}{G_T}, \quad W_L(T) = 1 - W_S(T),$$

the direct loading-implied pass-throughs are

$$\mathcal{D}_H(T) = W_S(T) + \omega_H W_L(T), \quad \mathcal{D}_I(T) = \omega_I W_L(T). \quad (77)$$

They equal $(1, 0)$ before insurer participation and converge monotonically to (ω_H, ω_I) . (iii) The full derivative also includes exposure feedback and the taxable price response and need not have these maturity profiles.

Proof. For (i), $A_1^M = \xi_H$ and $A_1^C = 1$. For (ii), switching off exposure terms gives $A_T^C = G_T$ and

$$A_T^M = \xi_H \sum_{j=1}^{\min(T, T^*-1)} \rho^{T-j} + (\omega_H \xi_H + \omega_I \xi_I) \sum_{j=T^*}^T \rho^{T-j}.$$

Differentiate $1 - A_T^M/G_T$ using $\xi_i = 1 - \tau_i$; an empty sum is zero. For $T \geq T^* - 1$, $W_S(T+1) = W_S(T)\rho G_T/(\rho G_T + 1)$, and since $G_T < 1/(1 - \rho)$ the multiplier is below $\rho < 1$, so W_S decreases from the participation transition and tends to zero geometrically, proving (ii). At a differentiable full affine solution, by contrast, the total derivative is

$$\frac{d\tau_T^A}{d\tau_i} = \frac{1}{A_T^C} \frac{dA_T^M}{d\xi_i} - \frac{A_T^M}{(A_T^C)^2} \frac{dA_T^C}{d\xi_i}. \quad (78)$$

If the branch has the long-end extension above, its coefficient fixed point is regular and $d_i := \partial d / \partial \xi_i$ is finite. Differentiating its geometric recursions gives

$$\lim_{T \rightarrow \infty} \frac{d\tau_T^A}{d\tau_i} = \omega_i + \frac{c_C}{1 - \rho} (\omega_i d + \xi^* d_i), \quad c_C = \gamma_I \sigma_r^2 m_H / m_I.$$

The equality holds with fixed branch parameters and $|\rho - c_C d| < 1$ near the parameter. Its additional term has no general sign, which proves (iii). $\square$

The channel-off calculation is an accounting decomposition of the affine recursions, silent on whether $d = 0$ is attainable in a jointly active heterogeneous-tax equilibrium.

Proposition 13 (Local nonlinear price response and sign preservation). *Let a differentiable branch of nonlinear equilibrium prices be given near a tax parameter θ , keeping the state process, cash rule, and idiosyncratic variances fixed. Assume the conditional exponential*

moments and their parameter derivatives admit a common integrable envelope. Set $q_T^a = \partial_\theta p_T^a$ and $g_T^a = -\partial_\theta \mu_T^a$, including all equilibrium portfolio and covariance responses. Then

$$q_T^a = \mathcal{K}_T^a q_{T-1}^a + g_T^a, \quad q_0^a = 0, \quad (\mathcal{K}_T^a \psi)(z) = \mathbb{E}^{Q_{T,z}^a} \psi(Z'), \quad (79)$$

where F_z is the conditional Gaussian law of Z' given z , and $dQ_{T,z}^a/dF_z = e^{p_{T-1}^a}/\mathbb{E}_z e^{p_{T-1}^a}$. If every g_j^a has a common weak sign at all states, q_T^a has that sign. If, in addition, some $g_{j_0}^a$ is strictly signed on a set of positive Lebesgue measure, then q_T^a has that strict sign at every state for $T > j_0$. For the yield-implied tax,

$$\partial_\theta \tau_T^{Y,\text{impl}} = \frac{q_T^M - (y_T^M/y_T^C)q_T^C}{Ty_T^C}, \quad y_T^C > 0. \quad (80)$$

Thus a municipal-price sign alone does not sign the ratio; the relative taxable response must also be controlled.

Proof. Dominated differentiation of the log expectation gives $\partial_\theta \log \mathbb{E}_z e^{p_{T-1}^a} = \mathbb{E}^{Q_{T,z}^a} q_{T-1}^a$; applied to the price recursion and iterated, this is (79). Each tilted kernel is a positive probability operator, so every summand preserves a common sign, and its strictly positive Gaussian density turns a strictly signed positive-measure contribution into a strictly signed function after one transition. Differentiating $1 - y_T^M/y_T^C$ with $y_T^a = -p_T^a/T$ gives (80). $\square$

The differentiable branch is a hypothesis the identity itself does not establish. In a chosen function space it follows from the implicit-function theorem when the equilibrium operator is continuously differentiable and $I - D_p \mathcal{T}$ has a bounded inverse, a local condition for which global contraction is sufficient but not necessary. Support changes require a re-solve or directional analysis.

The reported nonlinear P&C tax re-solve raises the effective keep from 0.686 to 0.834; the supply-weighted implied-tax change is approximately zero at two years and -1.285 , -1.861 , and -2.330 percentage points at ten, twenty, and thirty years. This finite reform on a finite grid supports the predicted maturity pattern; it neither identifies the pattern with the direct component (77) nor verifies the all-state infinitesimal sign premises of Proposition 13.

A.1.7 State-specific exemptions and additional clienteles

Proposition 14 (Multi-state aggregation). *For market M_s , put $\chi_{i,s} = \xi_i^C/\xi_{i,s}^M$, $\Gamma_T^{M_s} := \sum_{i \in \mathcal{U}_T^{M_s}} m_i/\gamma_i$, and $\omega_{i,T}^s := (m_i/\gamma_i)/\Gamma_T^{M_s}$. The exact formula is (67) with $d_{ij} = \chi_{i,s}$. In the*

affine construction it becomes

$$\begin{aligned}
s_T^{M_s} = & \sum_{i \in \mathcal{U}_T^{M_s}} \omega_{i,T}^s \chi_{i,s} + \frac{\sigma_r^2 A_{T-1}^{M_s}}{\Gamma_T^{M_s}} \sum_{i \in \mathcal{U}_T^{M_s}} m_i \partial_z \hat{D}_i \\
& + \frac{v_{M_s,T}}{\Gamma_T^{M_s}} \sum_{i \in \mathcal{U}_T^{M_s}} m_i \xi_{i,s}^M \partial_z x_{i,M_s}(T).
\end{aligned} \tag{81}$$

The final term vanishes under common active municipal keeps, a singleton active set, or zero idiosyncratic variance, but not by unweighted clearing alone.

Proof. Apply (69) coordinate by coordinate with the state-specific keeps; the three cancellations are those of Proposition 8. $\square$

Proposition 15 (State-tax anchor and direct pass-through). *Suppose the home household uniquely minimizes $\chi_{i,s}$ and the one-period municipal boundary is deterministic with $r > 0$. Then $y_1^{M_s}/y_1^C = \chi_{H_s,s}$. Under the multiplicative mapping $\chi_{H_s,s} = \xi^{\text{fed}}(1 - \tau_s^{\text{st}})$, define*

$$\tau_{s,T}^{\text{st,impl}} = 1 - \frac{y_T^{M_s}/y_T^C}{\xi^{\text{fed}}}, \quad \tau_{s,T}^{A,\text{st}} = 1 - \frac{A_T^{M_s}/A_T^C}{\xi^{\text{fed}}}. \tag{82}$$

The yield boundary equals τ_s^{st} and has unit pass-through. In the affine channel-off decomposition, if only the home-household effective keep changes and its active weight is one below T_s^* and constant at $\omega_{H_s}^s$ thereafter, the direct loading-state-tax derivative is $W_S^s(T) + \omega_{H_s}^s W_L^s(T)$, where W_S^s, W_L^s are defined in the display preceding (77) with T^* replaced by T_s^* . For $0 \leq \omega_{H_s}^s < 1$, this profile is one before participation and decreases thereafter to $\omega_{H_s}^s$, by the geometric recursion for W_S above. If $\omega_{H_s}^s = 1$, it remains one.

Proof. The normalized boundary KKT minimizes $\chi_{i,s}r$, as in Proposition 4, and substitution in (82) gives the anchor. In the channel-off decomposition the exposure and bond-specific inventory response terms in (81), and their corporate counterparts, are switched off, so $A_T^{M_s} = \sum_{j \leq T} \rho^{T-j} \sum_i \omega_{i,j}^s \chi_{i,s}$ and $A_T^C = G_T$. The direct derivative follows from (77), the common federal keep canceling in numerator and denominator. $\square$

If the perturbation also changes out-of-state keeps or other clienteles, their derivatives add to this direct formula. A different statutory keep mapping changes the implied-state-tax transformation but not the primitive identity $y_1^{M_s}/y_1^C = \chi_{H_s,s}$. The reported shortest-node and finite-maturity state-tax responses are computed from the full multi-state model.

Multiple clienteles and their direct tax effects. Results 1–3 use two investors for closed forms; with N types the portfolio problem and the KKT logic are unchanged, and

one participation threshold becomes several.

Index types by increasing taxable keep, $\xi_1 < \dots < \xi_N$, so type 1 is the household. In the rate-only benchmark, subtracting type 1's active municipal KKT equality from type j 's marginal condition gives the comparison of Result 1.

$$\underbrace{\sigma_r^2 A_{T-1}^M \Lambda_j}_{\text{duration benefit}} \geq \underbrace{(\xi_j - \xi_1) r_t}_{\text{tax penalty}}, \quad \Lambda_j := \gamma_1 \hat{D}_1 - \gamma_j \hat{D}_j. \quad (83)$$

The exposures are evaluated at the candidate allocation, including its corporate holdings, at which type 1 holds the municipal bond and type j is assessed. The comparison does not by itself pin down a unique household-only allocation across the other markets. With positive bond-specific risk, the incumbent's inventory premium and type j 's marginal inventory cost also matter, as in (59), so each type's crossing point depends on its portfolio, liabilities, and covariance structure.

Assumption 2 (Long-end participation, clientele by clientele). *Type 1 is active at every maturity, so $T_1^* = 1$. For each type $j > 1$, the exposure gap Λ_j evaluated at the stated affine equilibrium is positive, and its active municipal set is an upper maturity interval $[T_j^*, T_{\max}]$ with $T_j^* \geq 2$. The thresholds need not be ordered by tax advantage. The calibrated nonlinear holdings provide a numerical comparison, rather than a verification of these exact hard-constraint supports; an unrestricted N -type analogue of the affine two-investor Theorem 1 is not claimed.*

Tax ordering of thresholds is needed only for a theorem that the aggregate leading tax component moves monotonically at every maturity. Clientele-specific pass-through and the short- and long-end comparisons do not require it.

Prediction (P1–P2 generalized: clientele-specific pass-through). Under Assumption 2 the active set expands with maturity: the most muni-advantaged household prices the one-period bond, and the long-end leading tax component is the pooled risk-tolerance-weighted keep of all active clienteles, with the term-premium, convexity, and inventory corrections of Result 3. A later-participating clientele j 's affine direct component is zero below T_j^* and positive thereafter, the discounted average of its active risk-tolerance weights over the bond's maturity history rather than its current weight (Proposition 16). Total pass-through need not be zero below T_j^* , because the portfolio spillover can transmit type j 's tax through investors who hold a shorter risky bond, whereas the deterministic one-period boundary is unchanged by a strictly less tax-advantaged clientele's tax. P1 is the always-active household case and P2 the later-participating institutional case.

Empirically, cross-state household taxes move the always-active household clientele of P1, the bank-qualified design isolates a bank-only benefit, and the TCJA test moves the institutional tax advantage. P&C insurers are major municipal holders with the largest municipal tax advantage among taxable institutions, which makes state-level P&C exposure the sharpest observed measure of the later-participating channel. The calibrated long-end clientele also includes life insurers.

Proposition 16 (Many-clientele direct pass-through). *Under the nested supports of Assumption 2, the channel-off affine derivative $\mathcal{D}_i(T) = \partial\tau_T^A/\partial\tau_i$ for type i , where $\tau_T^A = 1 - A_T^M/A_T^C$, is*

$$\mathcal{D}_i(T) = \frac{\sum_{k=T_i^*}^T \rho^{T-k} \omega_{i,k}}{\sum_{k=1}^T \rho^{T-k}}, \quad (84)$$

with an empty numerator equal to zero. It is positive after participation begins and converges to ω_i if its active weight is eventually constant at ω_i . Ordered participation by increasing keeps makes the direct clientele-weighted keep nondecreasing; it does not by itself order total slopes, loading ratios, or yields.

Proof. Differentiate $1 - A_T^M/A_T^C$ using $A_T^M = \sum_{k \leq T} \rho^{T-k} \sum_i \omega_{i,k} \xi_i$, fixed support and weights, $A_T^C = G_T$, and $\partial\xi_i/\partial\tau_i = -1$. Every block before the weights stabilize is finite and, for fixed k , $\rho^{T-k}/G_T \rightarrow 0$, so all such blocks, including the transitional participation block, receive vanishing weight. Adding a new keep at least as large as all incumbents raises their risk-tolerance-weighted mean. The other slope terms in (81) and the taxable leg remain unrestricted by this argument. $\square$

A.1.8 Three maturities: completing the KKT verification

The main-text example sets $v = 0$ and $\gamma_H = \gamma_I = \gamma$. Clearing both books gives

$$\bar{D} = \frac{A_1^M Q_2^M + A_2^M Q_3^M + \xi_I(A_1^C Q_2^C + A_2^C Q_3^C)}{m_H + m_I} - \omega_H \ell_H - \omega_I \ell_I,$$

and, with the $M3$ duration gap (8), the split (9). The proposed holdings are

$$\begin{aligned} x_{H,M}(1) &= Q_1^M/m_H, & x_{I,M}(1) &= 0, \\ x_{H,M}(2) &= Q_2^M/m_H, & x_{I,M}(2) &= 0, \\ x_{H,M}(3) &= \frac{\hat{D}_H + \ell_H - A_1^M Q_2^M/m_H}{A_2^M}, & x_{I,M}(3) &= \frac{\hat{D}_I + \ell_I - \xi_I(A_1^C Q_2^C + A_2^C Q_3^C)/m_I}{A_2^M}, \\ x_{H,C}(T) &= 0, & x_{I,C}(T) &= Q_T^C/m_I \quad (T = 2, 3). \end{aligned}$$

Both $M3$ holdings must be positive. The $M1$ row reproduces $\mu_1^M = \xi_H r$ and the insurer slack $(\xi_I - \xi_H)r > 0$; substitution in the household's municipal and the insurer's corporate equalities yields the main-text returns. The remaining inequalities,

$$\nu_{I2}^M = (\xi_I - \xi_H)r \left(1 - \frac{A_1^M}{A_2^M}\right) > 0, \quad \nu_{HT}^C = \gamma \sigma_r^2 A_{T-1}^C \Delta \widehat{D} > 0,$$

hold provided $r > 0$, $A_2^M > A_1^M > 0$, and $A_1^C, A_2^C > 0$.

Supplies, loadings, and liability sensitivities are constant in the state on this branch, so differentiating the solved municipal returns gives

$$\begin{aligned} s_1^M &= \underbrace{\xi_H}_{\text{anchor}}, \\ s_2^M &= \underbrace{\xi_H}_{\text{anchor}} + \underbrace{\omega_I \frac{A_1^M}{A_2^M} \Delta \xi}_{\text{long-end spillover}}, \\ s_3^M &= \underbrace{\xi^*}_{\text{active-clientele mix}}. \end{aligned} \tag{85}$$

Equivalently, as an interpolation between the household keep and the long-end tax mix,

$$s_2^M = \left(1 - \frac{A_1^M}{A_2^M}\right) \xi_H + \frac{A_1^M}{A_2^M} \xi^*. \tag{86}$$

The insurer's active corporate KKT conditions give

$$s_1^C = 1, \quad s_2^C = 1 - \omega_H \frac{A_1^C}{A_2^M} \Delta \xi, \quad s_3^C = 1 - \omega_H \frac{A_2^C}{A_2^M} \Delta \xi, \tag{87}$$

and the riskless municipal anchor gives

$$A_1^M = \xi_H. \tag{88}$$

Substituted into (12), these give the municipal closure

$$(A_2^M)^2 - \xi_H(1 + \rho)A_2^M - \xi_H \omega_I (\xi_I - \xi_H) = 0, \tag{89}$$

whose positive root exceeds ξ_H ; the corporate closure gives $A_1^C = 1$ and $A_2^C = 1 + \rho - b$ with $b = \omega_H(\xi_I - \xi_H)/A_2^M$.

For Figure 4, $m_H = m_I = 1$, $\xi_H = 0.60$, $\xi_I = 0.80$, $\rho = 0.80$, $r = 0.05$, $\sigma_r = 1$, $\gamma = 7$, $Q_2^M = Q_3^M = 0.001$, $Q_2^C = Q_3^C = 0.0001074$, and $\ell_H = \ell_I = 0$. The root is

$A_2^M = 1.133$; the municipal responses are $(0.600, 0.653, 0.700)$ and the corporate responses $(1.000, 0.912, 0.849)$. The $M3$ holdings are approximately $(0.0008945, 0.0001055)$, the insurer's $M2$ slack is 0.004704 , and the household corporate slacks are $(0.008826, 0.015109)$. At the household-only allocation evaluated with these loadings, the duration benefits are $(0, 63.0, 119.0)$ basis points against a 100-basis-point tax penalty. The Panel B municipal-return components are the household tax anchor $(300, 300, 300)$, aggregate-duration compensation $(0, 41.3, 78.0)$, and participation $(0, 26.5, 50.0)$ basis points.

A.1.9 What nonlinear continuation adds

For locally absolutely continuous continuation prices whose Gaussian exponential moment and its state derivative have a common integrable envelope near the current state, differentiating under the expectation gives the loading recursion.

$$A_T^a(z) = s_T^a(z) + \rho \mathbb{E}^{Q_{T,z}^a} A_{T-1}^a(Z').$$

The deviation from the affine continuation term is

$$\delta_T^a(z) = \rho \left[\mathbb{E}^{Q_{T,z}^a} A_{T-1}^a(Z') - A_{T-1}^a(z) \right]. \quad (90)$$

Under the differentiability and integrable domination needed for a second derivative, differentiating the tilted average gives

$$\mathcal{V}_T^a(z) = \rho \text{Var}^{Q_{T,z}^a}(A_{T-1}^a(Z')), \quad \partial_z \mathbb{E}^{Q_{T,z}^a} A_{T-1}^a(Z') = \rho \mathbb{E}^{Q_{T,z}^a} \partial_{Z'} A_{T-1}^a(Z') - \mathcal{V}_T^a(z). \quad (91)$$

The covariance term follows because the log tilting weight has derivative $\rho \partial_{Z'} p_{T-1}^a = -\rho A_{T-1}^a$. Both diagnostics vanish for globally affine continuation prices but not for a merely fixed current support, since tomorrow's state can cross a floor or participation boundary.

At the stationary mean of the reduced economy in Figure 12, the averaging term is 6–11 percent of the loading in absolute value and the tilted dispersion 13–25 percent.

A.1.10 The thirty-maturity loading illustration

The two channels on a longer maturity ladder. Figure 13 shows the two pricing channels accumulating along a thirty-maturity affine branch with insurer participation from $T^* = 17$, $\rho = 0.8$, $\sigma_r = 0.025$, $(\gamma_H, \gamma_I) = (8, 5)$, $(\xi_H, \xi_I) = (0.55, 0.78)$, equal investor masses, and municipal idiosyncratic standard deviation $0.004[1 + 7\sqrt{(T-1)/29}]$ for $T \geq 2$.

Panel (a) plots one-period required-return responses. Below T^* the household alone

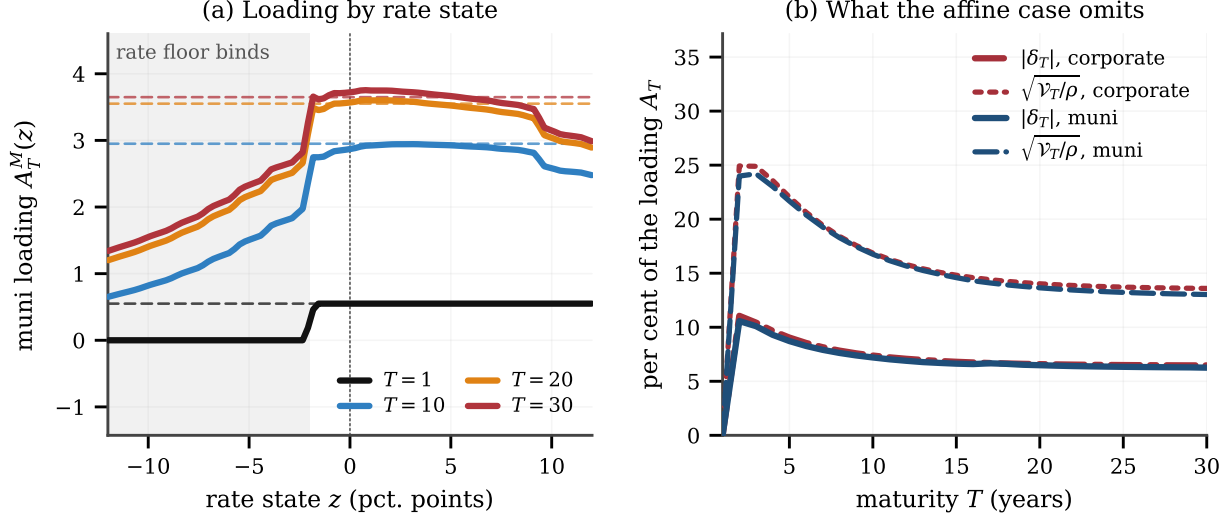


Figure 12: Nonlinear continuation in the reduced three-clientele economy: a household, a P&C insurer, and a tax-exempt institution. Panel (a) compares the municipal loading with the constant affine loading; shading marks the cash floor. Panel (b) reports $|\delta_T^a|$ and $\sqrt{\mathcal{V}_T^a/\rho}$ relative to the current loading at the stationary mean. The computation uses 193 state nodes; its price residual is 4.8×10^{-8} and clearing residual 5.7×10^{-13} . This reduced illustration is distinct from the full 45-state calibration.

Alt text: Price-loading diagnostics vary with the rate state in the nonlinear solution. Their deviations from the affine benchmark show where the rate floor and state-dependent participation alter the fixed-configuration approximation.

prices the municipal bond.

$$s_T^M = \underbrace{\xi_H}_{\text{household tax anchor}} + \underbrace{\gamma_H \sigma_r^2 A_{T-1}^M \frac{\partial \hat{D}_H}{\partial z}}_{\text{portfolio spillover}}, \quad T < T^*. \quad (92)$$

The spillover is positive at every risky maturity, including bonds the insurer never holds. At jointly held maturities the aggregate exposure response cancels in market clearing, leaving

$$s_T^M = \underbrace{\xi_H}_{\text{household tax anchor}} + \underbrace{\omega_I(\xi_I - \xi_H)}_{\text{direct clientele mix}}, \quad T \geq T^*. \quad (93)$$

The corporate response is $s_T^C = 1 - bA_{T-1}^C$; both markets therefore contribute to relative rate sensitivity.

Panel (b) plots the accumulated municipal components. With $G_T = \sum_{j=1}^T \rho^{T-j}$ the slope-one reference loading, d^* the exposure fixed point of (70) verified below, $e_j = \gamma_H \sigma_r^2 A_{j-1}^M d^*$ for $j < T^*$ and $c_j = \omega_I(\xi_I - \xi_H)$ for $j \geq T^*$, both zero otherwise, the loading recursion gives

identically

$$\frac{A_T^M}{G_T} = \xi_H + \frac{\sum_{j \leq T} \rho^{T-j} e_j}{G_T} + \frac{\sum_{j \leq T} \rho^{T-j} c_j}{G_T}.$$

Spillovers accrued below T^* remain embedded in longer bonds. Both components vanish under the no-participation benchmark of Proposition 10 and when the relative tax keeps coincide. Yield levels are not decomposed; they require the intercept recursion (75).

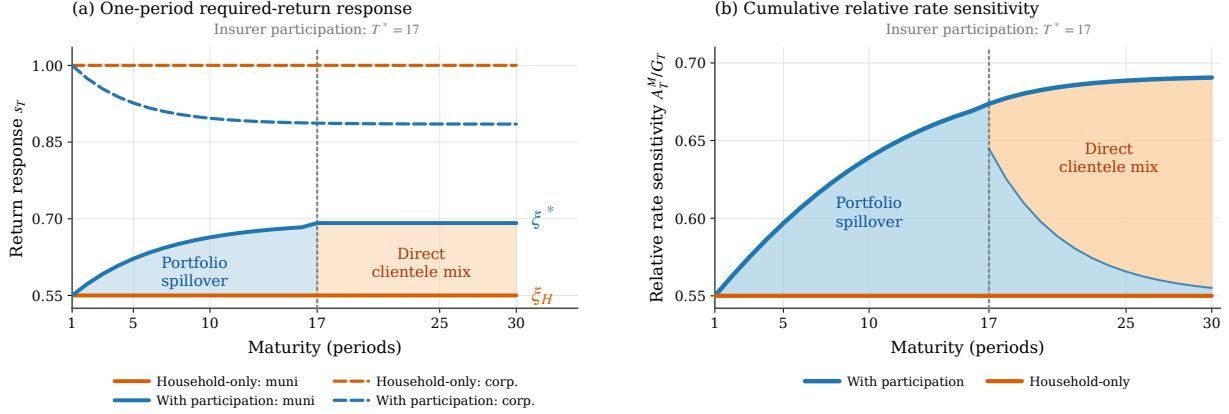


Figure 13: The two channels and their accumulation along maturity. Panel (a): municipal and corporate one-period required-return responses on the specified affine branch. Panel (b): the municipal yield response relative to a slope-one reference, with the household-only benchmark at ξ_H . Blue shading is the cumulated portfolio spillover; orange shading is the cumulated direct clientele contribution. The decomposition follows the affine loading recursion exactly. The vertical line marks the insurer's participation threshold.

Alt text: Two panels distinguish the response of one-period required returns from the accumulated response of relative yields. Insurer participation begins at the long-maturity threshold; portfolio spillovers affect shorter risky bonds, while the riskless boundary retains the household benchmark.

For every $d \geq 0$ the joint-block predecessor loadings exceed $a_{\min} = \xi_H \sum_{q=0}^{15} \rho^q$, so $P_J/S_J \leq 1/a_{\min}$ and the continuous map $\mathcal{F}_J$ takes $[0, \bar{d}]$ into itself, since

$$\bar{d} = \frac{\xi_I - \xi_H}{(\gamma_H + \gamma_I)\sigma_r^2 a_{\min}} < 10.591841.$$

An outward-rounded, forty-digit decimal interval evaluation of $\mathcal{F}'_J$ over 20,000 cells of $[0, \bar{d}]$ gives $\sup |\mathcal{F}'_J| < 0.080131$. Directed evaluations at the endpoints bracket the unique fixed point.

$$8.310258726885486 < d^* < 8.310258733114512.$$

The resulting A_{16}^M lies between 3.251488563972084 and 3.251488564481132, below $\xi^*/(1 - \rho) = 3.457692307692308$, so (M) holds; the left side of (S') is $0.16084 < \rho$. This verifies a unique loading solution on the proposed branch, leaving the level allocation, its support, and

uniqueness across alternative thresholds unverified.

A.1.11 Equilibrium selection and scope

The benchmark enumeration keeps an affine configuration and loading system only if its remaining level equations, a finite linear KKT-and-clearing system, have a unique feasible solution. This restricted affine check falls short of a global uniqueness theorem for the unrestricted nonlinear price-function operator. Exact identities apply to every differentiable equilibrium satisfying their hypotheses; the calibrated maturity pattern is a computed property of the 48-clientele branch, whose independently recomputed finite-grid residuals Appendix A.2.4 reports.

A.2 Nonlinear price-function algorithm and convergence

A.2.1 Moments, clearing, and the price update

On a state grid $\{z_g\}_{g=1}^G$ with Gauss–Hermite nodes and weights $\{(u_q, w_q)\}_{q=1}^{N_q}$, every continuation price is interpolated at $Z'_{gq} = \rho z_g + \sigma_r u_q$. With $\bar{p}_{j,-} := \sum_q w_q p_{j,-}(Z'_{gq})$ for asset–maturity coordinate j , the log continuation expectation and the conditional covariance are

$$m_T^a(z_g) = \log \sum_q w_q \exp\{p_{T-1}^a(Z'_{gq})\} + \frac{1}{2} \sigma_{\eta,T}^2, \quad (94)$$

$$\Sigma_{jk}(z_g) = \sum_q w_q (p_{j,-}(Z'_{gq}) - \bar{p}_{j,-})(p_{k,-}(Z'_{gq}) - \bar{p}_{k,-}) + \mathbf{1}_{\{j=k\}} v_j. \quad (95)$$

The deterministic one-period claims are priced and cleared analytically as in Appendix A.2.4. On the risky coordinates each investor’s barrier KKT condition,

$$K_i \mu(z_g) - \xi_i^C r(z_g) \mathbf{1} - \gamma_i K_i [\Sigma(z_g) K_i x_i(z_g) - c_i^L(z_g)] + \frac{\epsilon}{x_i(z_g)} = 0, \quad x_i(z_g) > 0, \quad (96)$$

and clearing, $\sum_i m_i x_i(z_g) = Q$, are solved by a joint primal–dual Newton step. Because Σ is diagonal plus a quadrature factor of rank at most $N_q - 1$, Woodbury products and a matrix-free Schur complement avoid dense investor Hessians. A fraction-to-boundary rule, a trust radius, and residual backtracking safeguard the step. After clearing every state, prices are updated by

$$\hat{p}_T^a(z_g) = m_T^a(z_g) - \mu_T^a(z_g), \quad (97)$$

and the outer iteration seeks $p = \hat{p}(p)$ with damped, trust-region safeguarded updates. Loadings $A_T^a = -\partial_z p_T^a$ are centered differences of the converged price functions; no affine coefficient is iterated.

A.2.2 Continuation and settings

Undamped iteration fails for three reasons. The cash floor has a kink at $z = -\bar{r}$ that binds with stationary probability of about 0.22, and the no-short constraints make each state's clearing problem combinatorial in the active set. The conditional expectation couples every state to every other. Two continuations address this. A state continuation releases nodes outward from the center, imposing (97) and clearing only where $|z_g| \leq \omega\sigma_z$, with $\sigma_z := \sigma_r/\sqrt{1-\rho^2}$ and $\Omega = (0, 0.6, 1.2, 1.8, 2.4, 3.0, 4.2, 5.4, 6.0)$. Frozen nodes keep their values but stay in every expectation at full weight. A floor continuation replaces $r(z)$ by $r_\lambda(z) = (\bar{r} + z) + \lambda \max\{-(\bar{r} + z), 0\}$ with $\Lambda = (0.1, 0.2, 0.4, 1.0)$. With a terminal refinement of the quadrature from 21 to 41 nodes this makes fourteen stages, run with damping $\theta = 0.8$, trust radius $\delta = 10^{-2}$ and tolerances $\text{tol}_p = 5 \times 10^{-8}$ and $\text{tol}_c = 2 \times 10^{-10}$. Only the terminal stage, with every node released, $\lambda = 1$ and the refined quadrature, carries the equilibrium interpretation. The endpoint is then interpolated onto the kink-aligned 71-node grid, with guard nodes at seven and eight standard deviations, and re-solved with analytic Gaussian moments of the piecewise-linear log-price interpolant and its linear tails. The riskless claims are eliminated analytically, and the last calibration step changes only common risk aversion to restore the thirty-year corporate target.

The barrier replaces the combinatorial active-set problem (689 traded coordinates, 644 of which remain in the risky system after the 45 riskless municipal boundaries are eliminated, and 48 investors) by a smooth Newton system. The barrier is a numerical device for solving that system, not an economic friction. Because $x_{ij}\lambda_{ij} = \epsilon$, small computed holdings do not identify which positions would survive the hard-constraint limit, so the paper reports the barrier value and the smaller-barrier comparisons of Appendix A.2.4 and does not treat computed holdings as exact hard-constraint solutions.

A.2.3 Numerical accuracy diagnostics

With $\lambda_i = K_i\nu_i$ the unnormalized nonnegativity multiplier of (96), the reported grid sets $G = 71$, risky-market barrier $\epsilon = 1.1920928955 \times 10^{-10}$, and common risk aversion $\gamma = 8.9732121$, matching the thirty-year corporate target at $z = 0$ to 0.00017 basis point. The diagnostics below are recomputed independently from stored prices and holdings at the reported riskless-boundary endpoint:

$$\begin{aligned}
\|p - \widehat{p}(p)\|_\infty &= 6.27 \times 10^{-9}, \\
\left\| Q - \sum_i m_i x_i \right\|_\infty &= 5.55 \times 10^{-17}, \\
\text{risky-market barrier stationarity residual} &= 3.25 \times 10^{-10}, \\
\|x_i \odot \lambda_i - \epsilon \mathbf{1}\|_{\infty, \text{risky}} &= 4.72 \times 10^{-12}.
\end{aligned} \tag{98}$$

The first line reconstructs the residual independently of the outer iteration's own stopping statistic. Together the four diagnostics document a high-accuracy approximate solution of the discretized finite-barrier equations, well short of an exact functional equilibrium or a fixed-configuration affine solution. At $z = 0$ the branch has a continuously compounded thirty-year corporate yield of 5.7935%, a supply-weighted municipal yield of 4.2442%, and an implied tax of 26.7424%. The deterministic municipal prices match their analytical anchor within 4.45×10^{-16} with zero omitted-dual violation, and nonholders hold exactly zero of the first-node claims eliminated before the barrier solve.

A.2.4 Numerical error and the riskless boundary

Boundary specification and scope. The maintained model makes the one-period claim riskless ($\sigma_{\eta,1}^a = 0$, $A_0^a = 0$). The reported computation eliminates it analytically:

$$p_1^{M,s}(z) = -r(z) \min_{i \text{ eligible}} \frac{\xi_i^C}{\xi_{i,s}^M}, \quad p_1^C(z) = -r(z).$$

The price and KKT residuals in (98) measure how accurately the algorithm solves the resulting finite-grid, finite-barrier equations. A local iteration derivative or a return from nearby starting points does not by itself establish global uniqueness.

Grid and boundary checks. A 33-node grid adds the floor kink $z = -\bar{r}$, the evaluation state $z^* = 0.004257$, and intermediate knots near the center. A nested 71-node grid extends the guards to eight standard deviations and bisects cells by stationary probability and width. Table 9, with analytic Gaussian moments throughout, holds the primitives and the barrier $\epsilon = 3.0517578125 \times 10^{-8}$ fixed. The riskless rows set the one-period municipal variance to zero and re-solve all longer maturities, while the others keep the auxiliary positive boundary variance. Independently reconstructed nodal price residuals are below 5×10^{-8} , clearing residuals below 10^{-15} , and both riskless solves match the exact boundary price within 4.5×10^{-16} in log-price units. With positive boundary variance, the ten- and thirty-year implied taxes change by -0.0937 and -0.1678 percentage point between the 21- and 71-node grids.

Table 9: Grid and boundary checks: yield-implied tax at z^* (percent)

State grid and boundary	2 years	10 years	30 years
21 nodes, positive boundary variance	41.543	33.424	27.070
33 kink-aligned nodes, positive boundary variance	41.543	33.340	26.925
71 guard-extended nodes, positive boundary variance	41.543	33.330	26.902
33 kink-aligned nodes, fully riskless boundary	41.667	33.356	26.929
71 guard-extended nodes, fully riskless boundary	41.667	33.346	26.906

Imposing the riskless boundary on the 71-node grid changes the two-, ten-, and thirty-year values by +0.1242, +0.0159, and +0.0042 point. All five curves decline at each modeled maturity. Yields are continuously compounded; the nominal-semiannual two-year entry is 41.84%. The table holds $\gamma = 9$.

Integration is exact only conditional on the interpolant. For a standard normal U and an affine segment $p = a + bU$ on $l < U < u$,

$$\mathbb{E}[e^{a+bU} \mathbf{1}_{\{l < U < u\}}] = e^{a+b^2/2} \{\Phi(u-b) - \Phi(l-b)\}.$$

Summing these terms integrates a piecewise-affine log-price function analytically, tails included, removing quadrature truncation but leaving interpolation error intact. Residuals tested at cell midpoints do not bound their supremum between test points.

Tail control requires a price envelope. Let $\tilde{p}$ be the extrapolated continuation and q its exponentially tilted probability outside a chosen domain. For any p that agrees with $\tilde{p}$ inside the domain and satisfies $|p - \tilde{p}| \leq \delta$ outside,

$$|\log \mathbb{E}e^p - \log \mathbb{E}e^{\tilde{p}}| \leq \max \{ \log[1 + q(e^\delta - 1)], -\log[1 + q(e^{-\delta} - 1)] \},$$

by bounding $e^{p-\tilde{p}}$ on each region before taking logs. Interior interpolation error needs a separate bound, and a small unconditional Gaussian tail probability supplies neither. Without such envelopes, a validated active face, and an inverse bound for the coupled price-and-KKT system, none of the residuals, tail diagnostics, or barrier-ladder differences reported here bounds the distance from a continuous-state, hard-constraint equilibrium.

Barrier control uses unnormalized multipliers. The numerical condition (96) adds the barrier to the unnormalized investor objective. If ν_{ij} is the keep-normalized slack in (59),

the unnormalized multiplier is $\lambda_{ij} = \xi_i^a \nu_{ij}$.

$$x_{ij}\lambda_{ij} = \epsilon, \quad \|\min(x, \lambda)\|_\infty \leq \sqrt{\epsilon}.$$

At exact barrier stationarity and fixed prices the gradient of the unpenalized objective is $-\lambda_i$, so concavity bounds the gain from any nonnegative alternative y_i by $\lambda'_i x_i - \lambda'_i y_i \leq n_i \epsilon$ for an investor with n_i constrained positions.

$$0 \leq V_i^* - V_i(x_\epsilon) \leq n_i \epsilon.$$

Reducing the barrier from 3.0518×10^{-8} to 1.9073×10^{-9} on the 21-node grid of Table 9 at fixed primitives moves the ten- and thirty-year implied taxes from 33.42391% and 27.06974% to 33.35408% and 26.90298%.

Barrier stability with the riskless boundary imposed. Table 10 repeats the barrier comparison on the 71-node grid with the one-period claims eliminated analytically. The final sixteenfold reduction meets a 0.1-basis-point successive-solution tolerance in both yield curves at every maturity at z^* . Ownership is less stable. The largest final-step change in a clientele’s aggregate municipal share is 0.730 percentage point.

Table 10: Barrier stability on the riskless-boundary, 71-node grid

Barrier ϵ	Yield-implied tax (%)		Maximum yield change (bp)	
	10 years	30 years	Municipal	Corporate
3.0518×10^{-8}	33.346	26.906	—	—
1.9073×10^{-9}	33.274	26.727	0.9064	0.1984
1.1921×10^{-10}	33.268	26.715	0.0537	0.0176

Notes. Maximum absolute changes are relative to the preceding row, across modeled maturities at the common evaluation state. The two-year implied tax, omitted from the table, is exactly 41.6671% at every barrier because the boundary is solved without a barrier; all three profiles decline with maturity. Independently recomputed nodal price residuals are below 3×10^{-8} and market-clearing residuals below 1.2×10^{-16} . The table holds $\gamma = 9$; the main calibration and the shape comparisons use the final barrier, $\epsilon = 1.1921 \times 10^{-10}$, with common risk aversion re-fitted to 8.9732.

A.2.5 Robustness to the maturity profile of idiosyncratic risk

To check that the assumed maturity profile of bond-specific risk does not generate the calibrated curve, all cases re-solve both markets from the main calibration with the profile

changed at fixed average variance. Let t denote maturity in years, $u_t = (t - 1)/29$, and v_t the idiosyncratic variance of the two-year log return. At $t \in \{4, 6, \dots, 30\}$ the three profiles are

$$v_t^{\text{base}} = (0.004)^2(1 + 7\sqrt{u_t})^2, \quad v_t^{\text{flat}} = c_f, \quad v_t^{\text{linear}} = c_l(1 + 7u_t)^2,$$

the last making volatility linear in maturity, leaving variance quadratic. All profiles set $v_2 = 0$ and apply the same variance to municipal and corporate bonds at each risky maturity; $c_f, c_l > 0$ preserve the pooled, supply-weighted mean variance over risky coordinates.

$$\frac{\sum_{j: t_j > 2} Q_j v_{t_j}}{\sum_{j: t_j > 2} Q_j} = \bar{v} = 0.0004345493.$$

The base variance rises from 0.0001691 at four years to 0.0010240 at thirty years, the flat profile is 0.0004345, and the linear-volatility profile rises from 0.0000824 to 0.0017730.

Table 11: Risk-profile robustness with a fully riskless first node

Risk profile	Implied tax (%)		Decline (pp)
	10 years	30 years	2–30 years
Square-root volatility	33.269	26.725	14.942
Flat volatility	33.228	26.621	15.046
Linear volatility	33.272	26.782	14.885

Notes. Every row is an equilibrium on the same grid at the same barrier, evaluated at $z^* = 0.004257$; the first row is the main calibration, the other two re-solve prices and holdings after the stated primitive change. Yields are continuously compounded, whereas the main empirical accounting uses nominal semiannual yields. The decline is $100(\tau_2^y - \tau_{30}^y)$ in percentage points.

In Table 11, changing the shape moves the thirty-year municipal yield by +0.830 basis point under flat volatility and −0.584 basis point under linear volatility, the largest municipal-yield changes at any maturity. The implied tax declines at every maturity in all three cases, with a two-to-thirty-year decline of 14.88–15.05 percentage points. In all three, independently recomputed nodal price residuals are below 7.3×10^{-9} , clearing residuals below 1.2×10^{-16} , and barrier-stationarity residuals below 3.6×10^{-10} ; the deterministic boundary price holds within 4.5×10^{-16} with no violation of its omitted dual inequalities.

A.2.6 Full nonlinear parameter sensitivity

Because several calibration inputs are measured with error, Table 12 perturbs five non-tax primitives by 20% in either direction from the main riskless calibration, holding the other primitives fixed, and re-solves the complete 45-state nonlinear equilibrium. No yield target

is re-fitted. Yields are supply-weighted, annualized, and continuously compounded at the exact state node $z = 0$. The main empirical accounting reports nominal-semiannual yields at z^* . Figure 14 plots the two state-level inputs to the calibration.

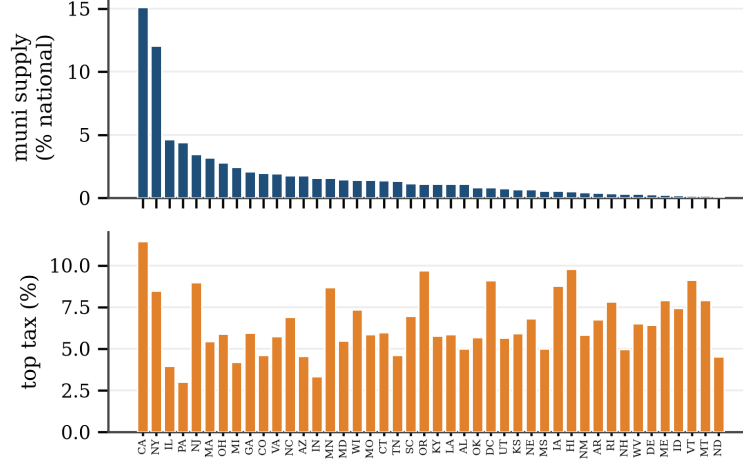


Figure 14: The state-tax cross-section used in the calibration. Top: each state’s share of national municipal supply. Bottom: each state’s statutory top income-tax rate. States are ordered by supply share; the seven no-income-tax states are aggregated into a single federal-only block (18% of supply, not shown).

Alt text: State-level bars display municipal supply shares and top statutory household tax rates. Together they show the geographic quantity and tax heterogeneity used in the calibration.

Table 12: Full nonlinear sensitivity with a riskless first period

Primitive varied by $\pm 20\%$	Two-year tax (%)	Thirty-year tax (%)	Decline (pp)	Thirty-year muni yield change (bp)
Baseline	41.67	26.74	14.92	0.0
Common risk aversion γ	41.67	26.11–27.50	14.16–15.56	−59.4+59.9
P&C mass m_{I_1}	41.67	25.69–28.21	13.46–15.98	−10.4+8.6
Life-insurer mass m_{I_2}	41.67	25.12–29.42	12.25–16.55	−18.5+13.7
Life-insurer liability loading $ \ell_{I_2} $	41.67	25.57–28.31	13.36–16.09	−11.4+8.7
Idiosyncratic volatility σ_η	41.67	26.68–26.81	14.86–14.98	−0.5+0.4

Notes. Each row is the envelope across the -20% and $+20\%$ endpoints, not a confidence interval. The riskless boundary pins the two-year implied tax at 41.6671% in every row. The last column is the thirty-year municipal-yield change from the baseline. Liability loading is perturbed in absolute magnitude; idiosyncratic variance is multiplied by 0.8^2 or 1.2^2 , so volatility changes by 20%. All ten endpoints retain the exactly riskless two-year boundary, the 71-node state grid, analytic Gaussian moments of the log-price interpolant, and $\epsilon = 1.1921 \times 10^{-10}$; primitive continuation and bounded price updates are used where needed, without relaxing the equation checks. Nodal price residuals, recomputed independently, are below 7.30×10^{-9} and market-clearing residuals below 1.12×10^{-16} .

Every endpoint retains a positive two-to-thirty-year decline, between 12.25 and 16.55 percentage points around the baseline 14.92. Idiosyncratic volatility has little effect at this scale; institutional masses, the life-insurer liability loading, and common risk aversion move the long-end level through participation, term premia, covariance feedback, and the allocation of duration. A cancellation in normalized risk-tolerance weights does not cancel in equilibrium yields. Common γ changes risk compensation, and liability loadings change net exposure and can shift participation.

A.2.6.1 Calibration units and extension inputs

On the two-year grid, $\rho = 0.85^2 = 0.7225$, $\sigma_r = 2(0.018)\sqrt{1+0.85^2} = 0.047248$, and $\bar{r} = 0.052444$. The risky-node idiosyncratic variance $(\sigma_{\text{idio}}\kappa(T))^2$ is zero at two years, 1.69×10^{-4} at four, and 1.024×10^{-3} at thirty. The inactive default-risk refinement uses cross-class shock correlation $\varrho_q = 0.40$, a zero first-period boundary, and a schedule increment of 4×10^{-3} in (101), with thirty-year endpoint approximately 0.0143 at $\rho = 0.7225$.

A.2.6.2 Mapping balance sheets into exposure

Let L_i be liability value and V_i^{out} the value of outside assets, with positive log-value loadings a_i^L and a_i^{out} on the same rate state as the bonds, all in the bonds' tax normalization and two-year state units. To first order, the outside-book contribution to net resources is

$$\underbrace{\Delta V_i^{\text{out}} - \Delta L_i}_{\text{outside-book innovation}} = \underbrace{(L_i a_i^L - V_i^{\text{out}} a_i^{\text{out}})_{\ell_i}}_{\ell_i} \sigma_r \varepsilon. \quad (99)$$

The outside book excludes the bond positions chosen in (19), whose exposure the portfolio already counts. Because ℓ_i combines balance-sheet size with exposure to the common rate factor, a reported liability horizon alone does not determine it. The baseline values approximate the sector ordering, in which long-lived life-insurance liabilities support a larger duration benefit than the shorter P&C book (Brunetti et al., 2023), leaving (99) unestimated; masses measure risk-bearing capacity in the model's supply units.

A.2.6.3 Sensitivity of the reported explained share

With the main accounting's nominal-semiannual yields, reporting state $z^* = 0.004257$, and observed benchmark, let B_{30}^{data} be the household-after-tax corporate benchmark and G_{30}^{data} the municipal gap after the call and issuance-premium adjustments. The reported share for

primitives θ is

$$\mathcal{S}_{30}(\theta) = 100 \frac{y_{30}^M(\theta, z^*) - B_{30}^{\text{data}}}{G_{30}^{\text{data}}}. \quad (100)$$

The denominator is 109.7225 basis points, fixed with the benchmark and bond-side adjustments. The baseline numerator of 94.7738 basis points gives the 86.3759% main-text share. Each alternative is the re-solved equilibrium of Table 12 in this convention.

Table 13: The explained share and the taxable calibration target

Primitive varied	Model share (%)	Corporate target error (bp)
Baseline	86.4	0.0
Risk aversion γ	31.0–142.3	–75.9–76.2
P&C mass	76.6–94.4	–2.7–3.3
Life-insurer mass	69.3–99.0	–37.3–41.4
Life liability exposure	75.8–94.4	–24.5–24.9
Idiosyncratic volatility	85.9–86.8	–0.1–0.1

Notes. Each range spans the –20% and +20% endpoints for one primitive, holding the others fixed; it is not a confidence interval. The share uses (100) with no refit or recalculation of the empirical adjustments. Corporate-target error is the re-solved thirty-year continuously compounded yield at $z = 0$ minus the 5.7935% calibration target, in basis points; the displayed baseline error of zero is rounded from –0.00017 basis point. The reporting-state convention affects the share, not the calibration target.

Life liability exposure between 1.776 and 2.664 contributes 83.2–103.6 basis points, or 75.8–94.4% of the gap; life-insurer mass gives 69.3–99.0%, P&C mass 76.6–94.4%, and idiosyncratic volatility 85.9–86.8%. All preserve the declining implied-tax profile, whose nominal-semiannual two-to-thirty-year decline ranges from 12.10 to 16.49 percentage points.

Changing common risk aversion without refitting gives 31.0–142.3% but misses the thirty-year corporate target by about 76 basis points in either direction, an off-target comparative static. Liability and mass changes also move both markets, so the baseline magnitude is assessed jointly with the taxable-curve fit. Stability of the implied-tax decline alone does not establish stability of the share. The one-at-a-time comparisons hold relative supply fixed and do not jointly identify balance-sheet exposures.

Table 14: External sources for the calibrated primitives

Primitive	Value	Role and source
Short-rate ($\rho_{\text{ann}}, \sigma_{r,\text{ann}}, \bar{r}_{\text{ann}}$)	AR(1) 0.85, 0.018, 0.0262	<i>Measured/targeted:</i> persistence and volatility from U.S. short-rate estimates; the annualized latent rate level matches the 2.6222% continuously compounded two-year corporate target at $z = 0$
Risk aversion γ_i	8.9732 (common)	<i>Targeted:</i> matches the 5.7935% continuously compounded thirty-year corporate target at $z = 0$; common γ cancels from normalized risk-tolerance weights, but can still change participation, term premia, and the nonlinear yield-implied tax
Household federal keep ξ_H^{fed}	0.62	<i>Statutory:</i> top federal rate (38% incl. 3.8% NIIT), 2000–2020 average
State surcharges $\{\tau_s^{\text{st}}\}$	44 taxing jurisdictions plus a no-tax block; supply-weighted mean 5.9%	<i>Statutory:</i> top state income-tax rates (Tax Foundation)
P&C keep ξ_{I_1}	0.686	<i>Statutory:</i> corporate rate 35% with §832 proration (15% muni-interest disallowance)
Life-insurer keep ξ_{I_2}	0.79	<i>Maintained:</i> effective relative keep under life-insurer proration
Residual keep ξ_{I_3}	1.00	<i>Sector mapping:</i> tax-exempt institutions (no tax on either bond)
Investor (H, I_1, I_2, I_3)	masses 0.30, 0.10, 0.40, 1.80	<i>Mapped/targeted:</i> sector mapping informed by Flow of Funds (Z.1 L.212); P&C mass fixed by a reduced calibration to the corporate long-end level

Table 14 continued

Primitive	Value	Role and source
Aggregate supply (muni, corp.)	0.3000, 1.7303 ($= 5.7675\times$)	<i>Maintained capacity normalization:</i> relative supply facing the modeled client-les, informed by cross-market ownership rather than the ratio of outstanding par (less than $2.45\times$ using the market stocks reported in §1); max-entropy maturity generators with mean 10 years; excluding the elastically supplied two-year corporate anchor leaves an 11.7-year mean for risky corporate supply
Liability loadings (H, I_1, I_2, I_3)	ℓ_i -1.387, +0.50, +2.22, 0	<i>Calibrated sector mapping:</i> signs and ordering informed by asset–liability structure; normalized exposure mapping and sensitivity in Appendix A.2.6.2
Idiosyncratic risk ($\sigma_{\text{idio}}, \kappa(T)$)	4×10^{-3} ; see Table 4	<i>Maintained:</i> variance schedule, not a fitted yield moment; Appendix A.2.5 changes its shape with a fully riskless boundary

Notes. Companion to Table 4: no primitive is chosen to fit a municipal yield, the municipal–corporate gap, or its maturity gradient; the cash-rate level and common γ are the two corporate-curve level targets. The exact ratio $1 - \xi_H$ holds in the analytical riskless one-period boundary; more generally tax keeps pin the two-year continuously compounded yield ratio, while the long-end magnitude depends on the primitives jointly through participation, term premia, and covariance feedback. Appendix A.2.6 reports the full nonlinear perturbations around the riskless baseline.

A.3 Model extensions and scope of the mechanism

A.3.1 Multiple muni states with state and federal taxes

State income tax is a household-specific channel. It falls on a household’s out-of-state but not home-state muni coupons, and not on institutions (Appendix A.9.3). Statutory top state rates run from zero to about 11%, with a supply-weighted mean of 5.9%, variation that both calibrates the model and identifies P1. The multi-state model replaces ξ_i with the state-pair ratio $\chi_{i,s} := \xi_i^C / \xi_{i,s}^M$ and applies (67) state by state. Proposition 15 derives the declining

direct-channel pass-through in the affine benchmark, and the calibrated model supplies the full-yield pass-through profile used in P1, with level and exposure corrections.

A.3.2 Default-risk diversification

Augment the baseline with class-level default shocks of cross-class correlation $\varrho_q \in (0, 1)$ and a maturity-volatility schedule $\sigma_q(T)$, zero at $T = 1$ and increasing in T , where T counts model periods, so the first boundary is two years on the numerical grid. With the benchmark allocation and rate-risk loadings fixed, Assumption 3 makes the insurer's profitable-deviation set expand, by more at longer maturities, and its cutoff weakly fall (Proposition 17). The schedule is

$$\sigma_q(1) = 0, \quad \sigma_q(T) = \sigma_q^{\text{slope}} + \rho\sigma_q(T-1), \quad T \geq 2, \quad (101)$$

with $\sigma_q^{\text{slope}} > 0$ and $0 < \rho < 1$.

Assumption 3 (Diversification condition). *Let the added pre-tax return innovation be $-\sigma_q(T)\eta_a$, with $\text{Var}(\eta_a) = 1$, $\text{Corr}(\eta_M, \eta_C) = \varrho_q$, and these shocks independent of rate and bond-specific shocks. At the allocation where households hold all munis and insurers all corporates, define after-tax default exposures per investor by*

$$M_H = \sum_T \sigma_q(T) Q_T^M / m_H, \quad C_I = \xi_I \sum_T \sigma_q(T) Q_T^C / m_I.$$

Assume $\gamma_H M_H > \gamma_I \varrho_q C_I$. The maturity volatilities and corporate keep are included in the exposures, not just in the marginal bond's loading.

Proposition 17 (Conditional default-diversification incentive). *Under the diversification condition and a default-volatility schedule that is zero at the one-period boundary and rises with maturity, default diversification weakly lowers the insurer's benchmark deviation cost at the fixed household-only allocation, by more at longer maturities.*

Proof. At the fixed households-only benchmark the household requires the municipal default premium $\sigma_q(T)\gamma_H M_H$ on its exposure M_H . The insurer, holding no municipals, incurs through its corporate book the marginal covariance cost $\sigma_q(T)\gamma_I \varrho_q C_I$ of adding a municipal position. Relative to the baseline without the default factor, its slack changes by

$$\sigma_q(T) [\gamma_I \varrho_q C_I - \gamma_H M_H] = -\sigma_q(T) [\gamma_H M_H - \gamma_I \varrho_q C_I].$$

The second bracket is strictly positive under Assumption 3, so the deviation cost falls by $\Delta_T^{\text{div}} := \sigma_q(T) [\gamma_H M_H - \gamma_I \varrho_q C_I] > 0$ whenever $\sigma_q(T) > 0$. Since $\sigma_q(1) = 0$ the boundary

is unchanged, and since $\sigma_q(T)$ rises with maturity the reduction is weakly larger at longer maturities. The comparison is of the benchmark KKT slack only; it does not assert that the same active set obtains after prices and portfolios re-equilibrate. $\square$

The baseline calibration sets the default factor to zero, so the reported equilibrium and every central result are independent of this extension, and the CDS adjustment of §A.5.1.5 is empirical accounting outside the tax-clientele equilibrium.

A.3.3 Beyond munis: scope of the mechanism

The mechanism needs two bond classes whose relative tax treatment differs across investors, so that the relative keep $\xi_i = (1 - \tau_i^B)/(1 - \tau_i^A)$ varies across clienteles without cancelling at $\xi_i = 1$, and a no-shorting friction that keeps the most advantaged investor from arbitraging the wedge flat. Decomposition (67) then applies, and implied tax declines when the relative forward ratio rises with maturity (Proposition 11).

Munis combine an unusually large wedge (full federal exemption) with a tight short-sale constraint. The same structure arises wherever one class carries an investor-specific tax advantage, producing a term structure priced by a maturity-shifting mixture of clienteles. Interest on treasury bonds and some agency debt (Federal Home Loan Banks, Farm Credit) is exempt from state income tax, while corporate interest and other agency interest (Fannie Mae, Freddie Mac) are not, sorting holders by state rate. A security's coupons are taxed currently in a taxable account but not in an IRA or pension. Withholding on cross-border sovereign coupons varies with the holder's country and tax treaty. The inflation accrual on TIPS is taxed as ordinary income before receipt, a phantom-income cost that taxable investors bear and tax-exempt accounts and nominal treasuries avoid.

A.4 Additional counterfactual: the 2017 TCJA by clientele

This counterfactual separates the household, P&C, and life-insurer components of the 2017 Tax Cuts and Jobs Act (TCJA). Each is a full-equilibrium re-solve of the $|\mathcal{S}| = 45$ nonlinear economy of Section 6 by continuation from the riskless baseline, with supplies and non-tax preference primitives fixed.

The TCJA cut the top household ordinary rate from 39.6% to 37% and the corporate rate from 35% to 21%. The household experiment applies the 0.026 change to the calibrated effective keep, from 0.62 to 0.646. With the P&C municipal rate fixed at 5.25% under

proration, the P&C relative keep moves from $0.65/0.9475 = 0.686$ to $0.79/0.9475 = 0.834$. For life insurers, equation (56) with the dated corporate rate maps the calibrated relative keeps $0.79 \rightarrow 0.843$ into municipal rates $17.72\% \rightarrow 6.29\%$. Because the inferred municipal keep changes, pre- and post-reform payoffs are expressed in the same pre-TCJA municipal-payoff unit, which keeps the calibrated non-tax preferences unchanged.

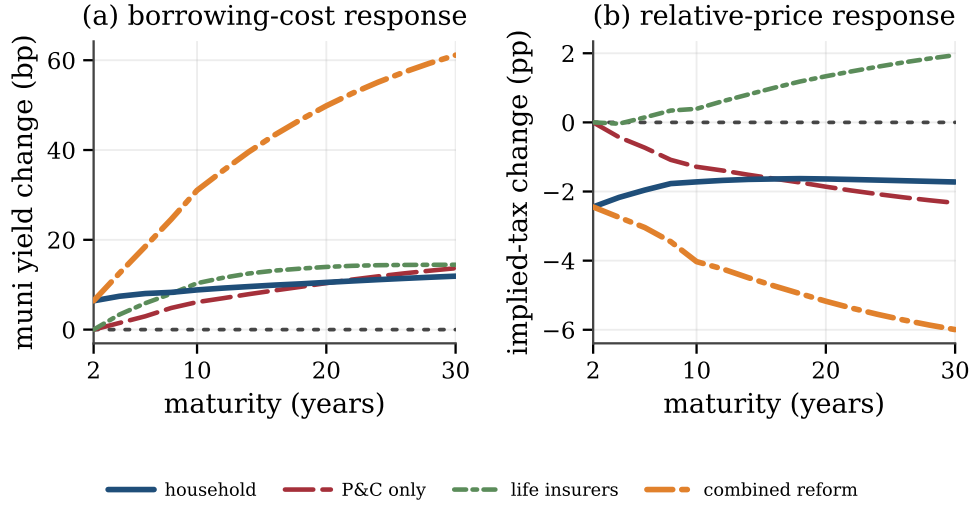


Figure 15: The 2017 TCJA by clientele. Each line is a separate full-equilibrium re-solve in a fixed pre-TCJA municipal-payoff unit. Panel (a) reports the municipal-yield response and Panel (b) the change in $1 - y^M/y^C$. The life-insurer experiment raises both yields; because its corporate-yield response is larger, the yield ratio falls and implied tax rises. The combined line changes all three tax treatments together and need not equal the sum of isolated responses.

Alt text: Isolated institutional tax changes produce different municipal-yield and implied-tax responses across maturities. A municipal yield increase can coexist with a higher implied tax when the jointly determined corporate yield rises proportionally more.

In Figure 15, the household-rate cut raises municipal yields by 6.4 basis points at two years and 11.9 at thirty and moves implied tax by -2.45 and -1.72 percentage points. The P&C change alone leaves two years unchanged, raises the thirty-year municipal yield by 13.7 basis points, and lowers implied tax by 2.33 points. The life-insurer change alone also leaves two years unchanged and raises the thirty-year municipal yield by 14.4 basis points but the corporate yield by 36.1 basis points, so y^M/y^C falls and implied tax rises by 1.95 points. All three together raise municipal yields by 6.4 basis points at two years and 61.1 at thirty and move implied tax by -2.45 and -6.00 percentage points; yields, rate exposures, and participation re-equilibrate together, so the combined response need not equal the sum of the isolated ones.

The P&C-only line is the structural counterpart of the reduced-form TCJA test (§7.3), which measures P&C exposure. It is a response absent at the short end and growing where

P&C insurers bear more duration, a pattern that a representative-tax benchmark applying one wedge uniformly cannot produce. The TCJA test’s point estimates have the same gradient, but their state-level uncertainty is too large to treat the event study as a precise estimate of the counterfactual magnitude.

A.5 Bond-side adjustments and empirical robustness

A.5.1 The five bond-side adjustments

Table 15: What explains the municipal–corporate spread. Contributions in basis points to $\text{Spread}(T)$ of equation (3). The two curve-level columns take the call option and the issuance premium out of the municipal yield itself, giving the residual puzzle plotted in Figure 2; the other three entries are muni-minus-corporate differentials and do not map to a single adjusted curve.

T	$\text{Spread}(T)$	curve-level			differentials			total	expl.	left, split		
		call	prem.	puzzle	cap. gn.	liq.	deft.			left	model	resid.
2	9.9	0.1	3.6	6.2	0.0	−0.4	15.3	14.8	—	−8.7	−2.7	−5.9
6	27.9	0.6	5.8	21.5	9.1	−0.4	11.2	19.9	92%	1.6	11.5	−9.8
10	52.2	4.5	7.9	39.9	9.5	−0.3	11.6	20.8	52%	19.1	22.3	−3.2
14	87.0	18.5	12.3	56.2	7.5	−0.1	11.1	18.4	33%	37.8	38.9	−1.1
18	110.3	23.9	15.3	71.2	5.2	−0.2	11.5	16.5	23%	54.6	56.4	−1.8
22	128.8	26.8	17.9	84.1	4.1	2.8	12.5	19.4	23%	64.7	71.7	−7.0
26	145.0	27.8	20.5	96.7	3.3	1.5	14.8	19.5	20%	77.2	84.6	−7.4
30	159.8	29.0	21.1	109.7	2.6	0.1	15.7	18.4	17%	91.3	94.8	−3.5

Notes. Spread less call and issuance premium is the puzzle; the puzzle less the three differentials is left. The explained share is $100 \times \text{total}/\text{puzzle}$ and is suppressed when the puzzle is below 20 basis points. The final columns satisfy $\text{puzzle} = \text{total} + \text{model} + \text{resid.}$ up to rounding. Thirty-year sensitivity envelopes are: call [22.4, 35.8], issuance premium [18.3, 24.0], capital gains [1.3, 2.8], liquidity [−2.2, 2.3], and default [7.9, 19.7] basis points. Sections A.5.1.2–A.5.1.6 give their construction; the issuance-premium estimate uses the 16% of curve bonds trading within one month of issuance.

A.5.1.1 Object, conventions, and the adjustments at a glance

Each adjustment is a basis-point contribution $\Delta_j(T)$ to the observed spread

$$\text{Spread}(T) = y_m(T) - y_c^H(T), \quad (102)$$

with y_m the all-bond municipal zero curve and y_c^H the corporate zero curve after the household wedge. The wedge combines the federal rate, state treatment, the net-investment-income tax and the post-2018 SALT rule, applied bond by bond before the same staging

as the pretax leg. It averages 41.3% in the main sample. A positive contribution explains part of the observed premium, and a negative one enlarges the puzzle. Spread space is used rather than the ratio $1 - y_m/y_c$, which amplifies a common subtraction: removing d from both legs moves τ by roughly $\tau d/(y_c - d)$ while the difference stays unchanged, and every adjustment is already a difference between two markets (Section A.5.1.5 shows the trap).

The object $\text{Spread}(T)$ is built on the all-bond panel over the full sample, staged maturity $\times$ month $\times$ half-notch rating. It excludes the CDS-matched inner join of the default estimation (Section A.5.1.5), which starts in 2004 and retains 89% of post-2004 bond-months (87% of all). That spread lies below the all-bond series at every tabulated maturity, by 14 to 26 bp (about -20.1 bp at thirty years). Quoting it as the denominator would compare components measured on one population against a total measured on another.

The call, liquidity and issuance-premium bands are sampling bands, drawn from a bootstrap of the calibrated volatility multiplier conditional on the coupon-matched design, the year-clustered standard error of the round-trip-cost differential, and a four-quarter moving-block bootstrap over issue quarters. The capital-gains and default bands are sensitivity envelopes over methods, horizons, sample windows and identifying assumptions. The combined band is the interval sum of the component ranges.

The sum of the five is a first-order approximation to the total, because the adjustments interact. The call pricer's fitted OAS absorbs credit and liquidity compensation, default risk affects callable cash flows and exercise incentives, liquidity can differ between callable and non-callable bonds, and refundings affect resale incidence. Without a joint repricing across frictions, these interactions stay unestimated rather than bounded. The volatility elasticity e_T measures sensitivity to the volatility multiplier alone, and the per-bond OAS refit absorbs parallel price-level differences without identifying them.

The five adjustments at a glance. Each $\Delta_j(T)$ uses the panel and staging of $\text{Spread}(T)$, so that the residual $\text{Spread}(T) - \sum_j \Delta_j(T)$ is the object the model is asked to explain.

1. **Call options (§A.5.1.2).** Each callable bond is repriced at its own fitted OAS with and without its call schedule, and

$$\Delta_{\text{call}}(T) = \overline{y_{\text{callable}}^{\text{zc}} - y_{\text{bullet}}^{\text{zc}}}^{\text{staged}}(T),$$

the staged (rating $\times$ month) mean of that difference, centered at the volatility multiplier $\hat{m}$ calibrated to the coupon-matched callable–noncallable yield separation.

2. **Capital gains (§A.5.1.3).** The quantity $W(h, T)$ is the municipal issuance-yield shift that leaves a seller at horizon h indifferent between the benchmark and actual

tax mappings under the corporate-anchored r^* closure, weighted by the measured first-sale probability.

$$\Delta_{\text{cg}}(T) = \sum_h \text{inc}_h(T) W(h, T), \quad f_\phi(a) = \phi \lambda_a \prod_{j < a} (1 - \lambda_j), \quad \lambda_a = \min(t_a/\phi, 0.99),$$

Here f is computed inside a trading sleeve of share $\phi(T)$ ($\phi = 0.60$ at thirty years), with zero adjustment on the held-to-maturity mass.

3. **Liquidity (§A.5.1.4).** Round-trip costs c , imputed within cusip $\times$ day $\times$ size cells and matched across markets on rating $\times$ institutional size $\times$ maturity bin $\times$ month, are amortized over a holding period,

$$\Delta_{\text{liq}}(T) = \frac{c_m(T) - c_c(T)}{h}, \quad h = 5 \text{ central}, h \in \{2, 10\} \text{ in the band.}$$

4. **Default (§A.5.1.5).** Each leg's default component is the after-tax cost of CDS protection, $L = \text{CDS} \times (1 - \tau_c)$, at the marginal CDS investor's statutory corporate rate τ_c ; the between-market difference, scaled by the loss-deductibility rate, is

$$\Delta_{\text{def}}(T) = (L_m(T) - L_c(T))(1 - \tau_L), \\ \tau_L = 0.20 \text{ central}, \quad \tau_L \in \{0, 0.37\} \text{ in the band.}$$

5. **Issuance premium (§A.5.1.6).** The same bond's offering-minus-secondary yield gap, drift-controlled against seasoned bonds of the same maturity bucket trading in the same week,

$$\Delta_{\text{prem}}(T) = [y_{\text{off}} - \bar{y}_{\text{seasoned}}(w_{\text{issue}})] - [y_{\text{trade}} - \bar{y}_{\text{seasoned}}(w_{\text{trade}})],$$

at short trade age on the curve's own bond population and staging.

A.5.1.2 Adjustment I: embedded call options

Municipal issuers sell long bonds callable, by convention at par ten years after issue and on any coupon date thereafter, and the investor is paid in yield for writing the option. In the panel behind the curve, 49.5% of observations are on callable issues, and 50.8% over the two-to-thirty-year range where the adjustment matters. Investment-grade corporates are bullet or make-whole callable, hence close to option-free, so the municipal-minus-corporate gap compensates for an option the corporate investor never wrote; dropping callable municipals

instead would skew the sample toward short maturities. What is removed, bond by bond, is the bullet value of the same cash flows less their callable value,

$$CV_b = V_b^{\text{bul}} - V_b^{\text{call}} \geq 0; \quad (103)$$

adding CV_b to the observed price and re-inverting gives the option-free yield.

Method. A lattice built from the equilibrium’s municipal discount surface gives the flat option-adjusted spread θ^* at which the callable bond reprices at its observed dirty price, and the same cash flows are repriced at that θ^* with the call removed. Because θ^* is fitted on the callable leg and held fixed on the bullet leg, everything that is not the option, credit and liquidity included, cancels in (103); only the option value comes from the model.

Implementation. Cash flows follow each bond’s coupon, dated date, first-coupon date and recorded call schedule on a semiannual 30/360 grid, with call dates mapped to coupon dates and strikes used as recorded; the pricing target is the dirty offering price. The lattice discounts on the supply-weighted mean of the equilibrium’s 45 state municipal log zero-price surfaces, so the option-adjusted spread is a spread over the model’s mean municipal curve. Before the pricing drift is fitted, the surface’s spot-volatility curve is rescaled to the measured municipal volatility curve σ_T^{data} , the annual standard deviation of twelve-month changes in the staged municipal zero yield, times a multiplier m calibrated below. Without the rescaling, the pricer gives the model-volatility diagnostic of 21.8 basis points at thirty years. The state follows the model’s AR(1), annualised from the two-year period, with a drift fitted node by node so that a half-year step of discounting reproduces the surface. The fitted drift at $z = 0$ is +0.0077, and the fit residual is common to both legs and cancels in (103). Each bond is priced from its issuance-month state by backward induction on a 201-node grid with ex-coupon exercise, and the OAS is solved by bisection to a repricing error below 10^{-9} per 100 face; its median is +39 basis points. The bullet leg is repriced at the same OAS without the call cap, $CV_b = \max(V_b^{\text{bul}} - V_b^{\text{call}}, 0)$ and $P_b^{\text{dec}} = P_b^{\text{dirty}} + CV_b$. Of the 449,412 bonds with a mapped call, 449,181 carry a positive call value, with median \$2.01 per 100 face. Raw and de-called prices are converted to nominal semiannual zero-coupon yields on a common base curve, and $\Delta_{\text{call}}(T)$ is the per-bond difference staged the same way as the observed curve, within maturity $\times$ issue quarter $\times$ rating cells, then across ratings, then across quarters. The pricer excludes 6,986 of 904,093 rows whose price the bisection bracket does not span or whose schedule cannot be mapped.

Identification of the volatility level. Let m scale the measured volatility curve σ_T^{data} . Callable and non-callable bonds are matched within maturity $\times$ issue quarter $\times$ rating $\times$ half-point coupon cells, and their yield difference g_T is the calibration target for call value. Running the pricer at $m = 1.00$ and $m = 1.15$ gives each maturity its local elasticity

$$e_T = \frac{\partial \ln s_{\text{call}}(T)}{\partial \ln m}, \quad (104)$$

so the model's gap at any m is $s_T(m) = s_T(1) m^{e_T}$ locally. This e_T is estimated on the coupon-matched cells; the adjustment's own band uses the corresponding elasticity of the all-bond staged panel. The multiplier is then weighted least squares on integer maturities six through twenty with at least twelve quarterly observations and both gaps positive,

$$\ln g_T = \ln s_T(1) + e_T \ln m + \varepsilon_T, \quad \text{weight } 1/\text{var}(\ln g_T), \quad (105)$$

with solution

$$\ln \hat{m} = \frac{\sum_T w_T e_T \ln(g_T/s_T(1))}{\sum_T w_T e_T^2}. \quad (106)$$

This gives $\hat{m} = 1.089$. Resampling calendar quarters as blocks gives the 95% band $[0.970, 1.206]$, which accounts for within-quarter cross-maturity dependence but not serial dependence across quarters. The band measures sampling variation conditional on the coupon-matched design, not callable-selection or model-specification uncertainty, since matched cells can still differ in issuer, security design and liquidity. $m = 1$ lies inside the band.

Result. At thirty years the adjustment is 29.0 bp, band $[22.4, 35.8]$ bp; under the measured volatility curve ($m = 1$) it would be 24.0 bp and under the model's own volatility 21.8 bp, both retained as diagnostics (Table 16). The elasticity rises from 0.76 at two years to 2.16 at thirty, which is why the band widens with maturity. Beyond about twelve years this is the largest of the five adjustments and, with the issuance premium, one of only two that grows materially with maturity. The central estimate and both band edges are direct full-pricer repricings at $\hat{m}$ and the bootstrap quantiles. The power law is fitted from two runs ($m = 1.00$ and 1.15), and the upper endpoint lies beyond the fitted pair. The direct runs differ from the elasticity prediction by at most a quarter of a basis point, at thirty years and the upper quantile, so nothing rests on the functional form.

Let $y^{M,\text{call}}(T)$ be the call-adjusted municipal yield and $y^{C,\text{AT}}(T)$ the after-tax corporate

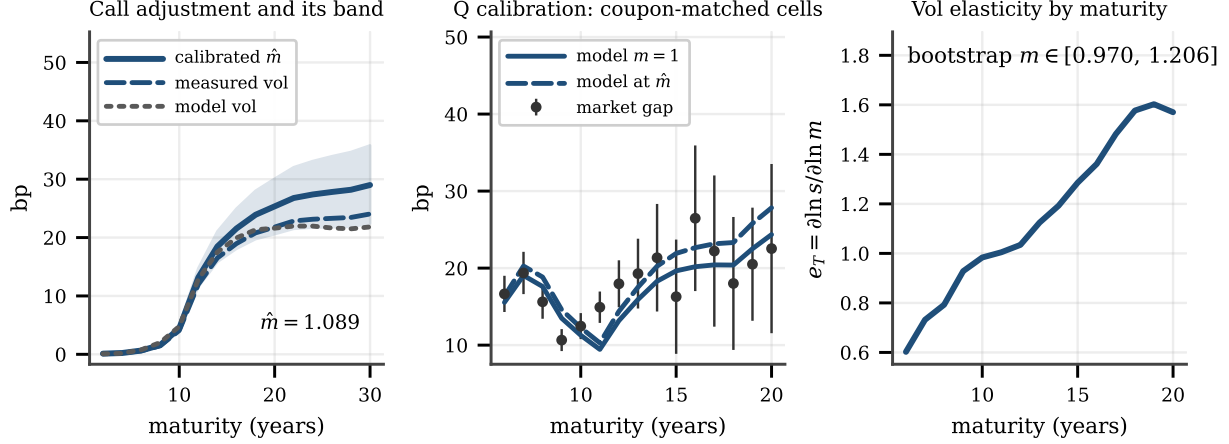


Figure 16: Left: the call adjustment under the measured volatility curve, at the calibrated multiplier $\hat{m}$, and under the model’s own volatility. Middle: callable–noncallable yield separation in coupon-matched cells (± 1.96 standard errors) against the model at $m = 1$ and at $\hat{m}$. Right: the volatility elasticity by maturity, which maps the bootstrap band into the band on the adjustment.

Alt text: The panels show call-adjustment sensitivity to volatility, the coupon-matched calibration comparison, and maturity-specific volatility elasticities. The plotted bands represent the stated conditional sampling or sensitivity exercises.

Table 16: Call adjustment by maturity (bp). “Model vol” uses the equilibrium’s own volatility and “measured” the realized volatility curve, both diagnostics; the central case and the band edges are direct repricings at $\hat{m}$ and the bootstrap quantiles; e_T is the fitted elasticity.

T	model vol	measured ($m=1$)	e_T	central ($\hat{m}$)	band
2	0.1	0.1	0.76	0.1	[0.1, 0.1]
4	0.2	0.2	0.61	0.2	[0.2, 0.2]
6	0.8	0.6	0.90	0.6	[0.6, 0.7]
8	2.0	1.4	1.03	1.6	[1.4, 1.8]
10	4.8	4.1	1.05	4.5	[3.9, 5.0]
12	12.8	12.0	1.11	13.2	[11.6, 14.7]
14	17.5	16.5	1.31	18.5	[15.8, 21.0]
16	19.9	19.0	1.48	21.6	[18.1, 24.9]
18	21.3	20.8	1.61	23.9	[19.7, 28.0]
20	21.6	21.8	1.74	25.3	[20.6, 30.1]
22	22.0	22.8	1.84	26.8	[21.5, 32.0]
24	22.0	23.1	1.94	27.4	[21.7, 33.1]
26	21.6	23.3	2.04	27.8	[21.8, 33.9]
28	21.5	23.4	2.12	28.2	[21.9, 34.6]
30	21.8	24.0	2.16	29.0	[22.4, 35.8]

yield. The call-adjusted spread

$$\Delta^{M-C}(T) \equiv y^{M,\text{call}}(T) - y^{C,\text{AT}}(T)$$

is about 48 bps at $T = 10$, 98 bps at $T = 20$ and 131 bps at $T = 30$, and about 115 bps on average beyond twenty years, from 102 bps at $T = 22$ to 131 bps at $T = 30$. This is computed directly as the observed spread of Table 25 less the call column of Table 16, and not estimated separately. The left panel of Figure 17 plots the smoothed average curves. The rise of the adjustment itself with maturity in Table 16 is a sample-average pattern rather than a theorem, since across heterogeneous coupons, rates and call schedules option value need not be monotone in maturity.

Figure 17 places the de-called full-sample curve beside the non-callable subsample, the non-decalled municipal curve and the corporate pre-tax and after-tax curves, on common-support maturity-by-year-by-month-by-rating cells. Raw and smoothed averages align closely through twenty years, where noncallable observations remain available, as effective average option removal implies. The alignment does not validate each option value, because callable and noncallable bonds can differ within cells.

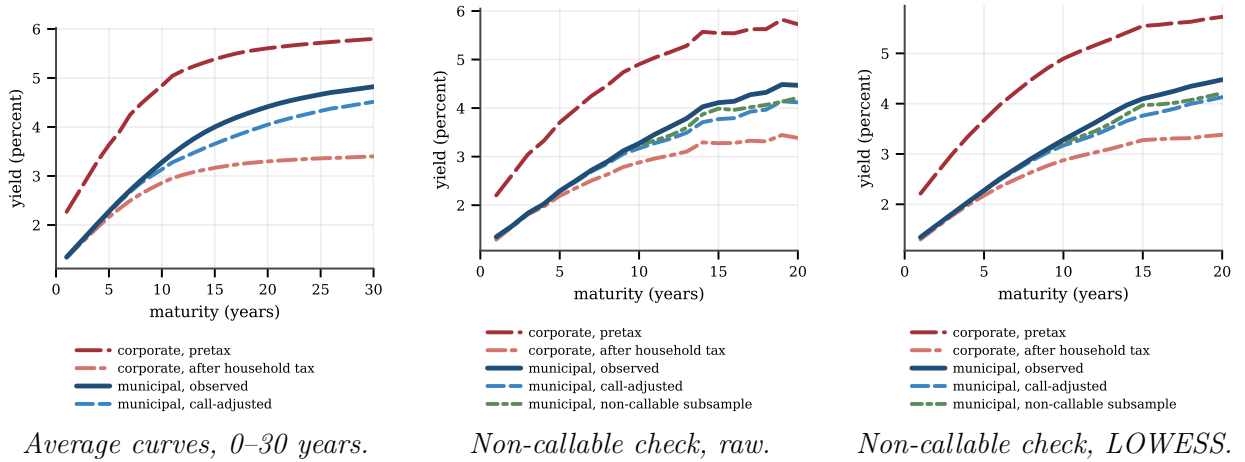


Figure 17: The call adjustment on average yield curves. Left: average municipal and corporate yield curves through thirty years, LOWESS-smoothed, with the municipal curve before and after the call adjustment, which is small through ten years and reaches 29.0 basis points at thirty. Middle and right: the decalling diagnostic against the non-callable subsample through twenty years, raw and LOWESS-smoothed, on common-support maturity-by-year-by-month-by-rating cells.

Alt text: Average municipal and corporate yield curves are shown with the municipal curve before and after removing the call option, which lowers option-free municipal yields farther along the curve. Two further panels place the de-called curve beside the non-callable subsample within common comparison cells; their proximity is a descriptive validation of the call adjustment.

A.5.1.3 Adjustment II: capital-gains taxation

$\text{Spread}(T) = y_m - y_c^H$ already prices the coupon exemption. What remains is the difference in how the two bonds' capital income is taxed. Let A_h be the accreted basis, or adjusted issue price, at year h : the issue price grown at the issuance yield. On a municipal bond, coupon interest is exempt. Original issue discount accretes into A_h as tax-exempt interest and is never taxed. Market discount, the gap between A_h and a lower price paid later in the secondary market, is taxed on disposition, under the default election once at maturity. A discount below the de minimis threshold at n remaining years,

$$\kappa_j = 0.0025 \times 100 \times j, \quad (107)$$

is taxed as capital gain rather than ordinary income. Bond or acquisition premium generates no tax cash flow and only prevents a terminal loss. On the taxable benchmark, coupon interest and OID accretion are ordinary income year by year, and acquisition premium proportionally reduces future OID. Premium above par is amortised against income under §171 (the election is taken), and market discount is taxed on disposition subject to the same de minimis test. Ordinary income is taxed at $\tau_{oi} = 0.37$ and capital gain at $\tau_{cg} = 0.20$. The tax on a realized gain G is

$$\mathcal{T}(G) = \tau_{cg} [(1 - \lambda)G + \lambda \max(G, 0)], \quad \ell = 0, \quad (108)$$

so losses are fully deductible, a maintained assumption that does not affect the reported band. Rates are federal and flat, with no state layer.

In the benchmark world, every municipal dollar is exempt, no capital-income tax is ever due, and all taxable-leg income, including any discount, is ordinary. In the actual world, the statutory rules above apply on both legs. The adjustment is the difference between the issuance-yield shifts δ_B and δ_A that the municipal bond needs in the two worlds. The benchmark world is not a no-tax world, since the corporate leg is taxed in it at ordinary rates, so only the difference between the two shifts isolates the capital-income asymmetry.

Construction. Two archetypes are built at each even maturity T from four to thirty on the equilibrium's supply-weighted municipal curve and its corporate counterpart at the issue state $z = 0$. One is a par bond (PAR), whose basis stays at par; the other is a zero-coupon bond (ZERO), whose basis A_h accretes at its issuance yield. For a sale at horizon $h = 2, 4, \dots, T - 2$, resale states are integrated against the AR(1) distribution started at $z = 0$ with 32 Gauss-Legendre nodes per interval of the model's state grid; 48

nodes change the adjustment by at most 0.02 basis points. A municipal resale price and a municipal required return cannot both be read off the municipal market, so the seasoned corporate price anchors the buyer’s after-tax required return r^* under each tax mapping. At that return, the benchmark resale price is the present value of the remaining exempt flows. Under the actual mapping, market discount $A_h - P$ is taxed once at maturity, which gives $P(\tau) = (G_A - q\tau A_h)/(1 - q\tau)$, with G_A the gross present value and q the terminal discount factor. This is evaluated at τ_{cg} when the discount is below the de minimis threshold and at τ_{oi} otherwise, and $P = G_A$ when the bond sells at or above its basis. A parallel shift δ of the municipal issuance curve is then solved in each tax world so that the municipal seller’s annualised log of probability-weighted mean terminal wealth, which is not $\mathbb{E}[\log W]$, equals the corporate seller’s in that world; only the municipal instrument moves with δ . The conditional adjustment is the difference between the two shifts,

$$W(h, T) = 10^4(\delta_A - \delta_B) \text{ basis points}, \quad (109)$$

averaged across archetype pairs with issuance weights.

Archetypes and issuance weights. The adjustment is computed for four (corporate, municipal) archetype pairs in $\{\text{PAR}, \text{ZERO}\}^2$ at every even maturity. Issued bonds are PAR-like when the offering price lies in $[90, 110]$ or exceeds 110, and ZERO-like below 90 or with a zero coupon. A non-positive offering price marks a missing value rather than a deep discount, because FISD records an unknown price as 0.00, and reading it literally classified 830 fixed-coupon corporates as ZERO-like, more than half of the measured corporate ZERO weight. Amount-weighted over 1998–2021 issues, the ZERO-like share pooled over all maturities is 1.3% for municipals and 2.2% for corporates. The baseline uses maturity-matched shares, computed on issues whose original maturity is within three years of the grid maturity, with the pooled share as fallback when that window carries too little offering amount. At thirty years they are 0.97% and 0.31%, and the pair weight $w(c, m) = \text{share}_c \times \text{share}_m$ puts 98.7% of the matched weight on PAR/PAR (Table 17).

The first-sale density. Let π_a denote the turnover rate at age a , seasoned transfer par over par outstanding, from the full MSRB dealer-to-customer tape (71.5M legs, 2005–2021) pooled over original maturities within ± 4 years of T , excluding when-issued trades and the first quarter after the dated date. The weight is the original buyer’s first sale rather than turnover. The corporate-anchored resale price already prices later generations’ taxes, so turnover weighting counts their dispositions a second time, inflates the number severalfold and can push the “unconditional” value above the conditional one. A share $\phi(T)$ of par

Table 17: The four archetype pairs at $T = 30$: the pooled interacted issuance weight, and each pair’s conditional adjustment across holding horizons. The baseline uses maturity-matched shares, which put PAR/PAR at 98.7% at thirty years.

corp / muni	weight (%)	max over h (bp)	mean over h (bp)
PAR / PAR	96.5	8.52	6.75
ZERO / PAR	2.2	27.23	20.39
PAR / ZERO	1.3	8.69	6.51
ZERO / ZERO	0.0	18.90	13.92

trades and the rest never does. The hazard borne by the trading sleeve is scaled up and capped,

$$\lambda_a = \min(t_a/\phi, 0.99), \quad (110)$$

the cap binding only at thirty years in the central calibration, at the first two ages, where the scaled hazard would otherwise reach 1.00 and 0.99. The first-sale density and the held-to-maturity mass are

$$f_\phi(a) = \phi \lambda_a \prod_{j < a} (1 - \lambda_j), \quad S_\phi(T) = 1 - \phi \left(1 - \prod_{a < T} (1 - \lambda_a)\right), \quad \sum_a f_\phi(a) + S_\phi(T) = 1, \quad (111)$$

and the unconditional adjustment is $\Delta_{\text{cg}}(T) = \sum_h \text{inc}_h(T) W(h, T)$. Here inc_h sums f_ϕ over the ages mapped to node h (the first node absorbs the first year and the last node the final pre-maturity year, a discretization approximation), and the held-to-maturity branch carries a zero adjustment. At thirty years $\phi = 0.60$, so $S_\phi = 0.398$ and the first-sale mass is concentrated at the youngest ages.

The trading sleeve. A hazard common to all holders would imply that nearly every thirty-year dollar is sold before maturity, and aggregate turnover does not identify the share of par that trades. For each CUSIP, cumulative seasoned traded par capped at its own outstanding par bounds from above the par that ever changed hands, repeat trades of the same dollar inflating the numerator. Par-weighted over issues dated from 2005, whose entire observed life sits inside the tape, the ceiling on $P(\text{ever sold})$ is 0.47–0.63 across the tabulated bins, with 13 to 23% of par never trading (Table 18). The parameter ϕ is solved so that the implied fraction ever sold reproduces that ceiling over the same observation window and is then applied over the whole life. Bins within four years of T are pooled and weighted by outstanding par, so at thirty years the target is 0.602 against the 0.617 of that bin alone.

Table 18: The measured ceiling on the fraction of par ever sold, by original maturity. The cap counts repeat trades of the same dollar, so it is an upper bound and the reported band extends downward from it.

orig. maturity	ceiling on $P(\text{ever sold})$	never traded	obs. window (y)	CUSIPs
4	0.466	0.195	3.9	124520
6	0.571	0.155	5.6	121412
8	0.556	0.227	7.0	118265
10	0.625	0.132	8.0	110693
14	0.558	0.171	9.4	73189
18	0.546	0.216	9.7	46382
22	0.586	0.219	10.0	13056
26	0.612	0.193	10.6	6551
30	0.617	0.159	10.3	22151

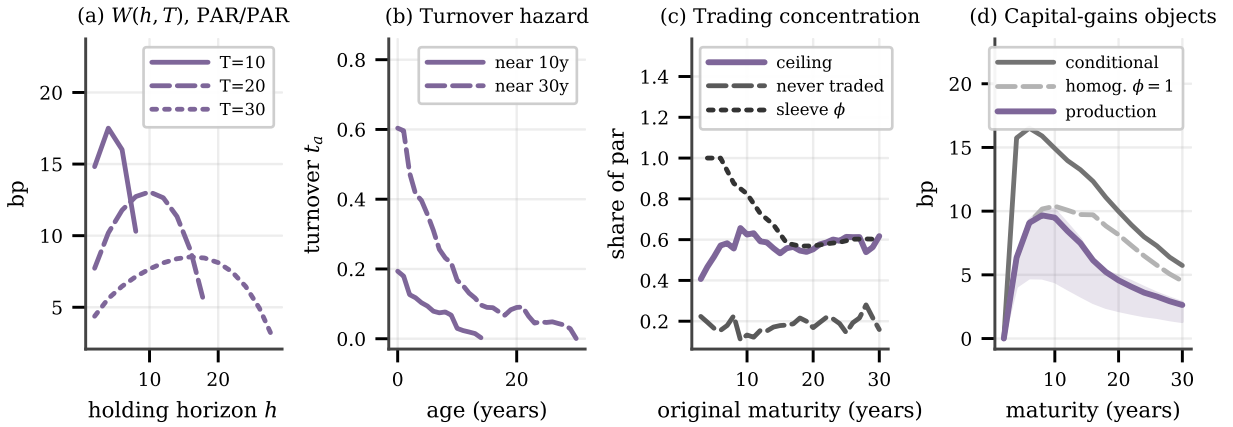


Figure 18: Capital gains. Panel 1: the conditional adjustment $W(h, T)$ by holding horizon, three maturities. Panel 2: the measured turnover hazard by age. Panel 3: the measured ceiling on $P(\text{ever sold})$, the share of par never traded, and the calibrated sleeve ϕ . Panel 4: the conditional-on-sale, homogeneous-hazard unconditional, and reported incidence-weighted series with its band.

Alt text: The panels compare tax-rule adjustments across holding horizons, turnover assumptions, trading bounds, and the incidence-weighted aggregation. They trace how tax treatment and realization timing jointly affect the municipal-corporate yield comparison.

Result. The central case sits at the measured ceiling, which over-counts, so the sleeve dimension can only lower the adjustment; the tape-window and weighting dimensions move it either way, and the band is two-sided. It is the envelope of eighteen variants. These cross three tape windows (the full sample and its early and late halves), three sleeve multipliers (0.5, 0.75 and 1.0 times the measured ceiling), and the two issuance weightings. The sleeve is re-solved on the full window for any multiplier below one, even when incidence is evaluated on a half. The sleeve dimension dominates the width, so the band is chiefly holder-heterogeneity rather than sampling uncertainty. At long maturities the homogeneous value is a diagnostic

excluded from the band, since its implied $P(\text{ever sold})$ exceeds what the data permit; at the short end the calibration returns $\phi = 1$ and the two coincide. At thirty years the adjustment is 2.6 bp, band [1.3, 2.8] bp, against a conditional-on-sale value of 5.7 bp, which averages $W(h, T)$ over the age distribution of observed customer sales rather than first sales, and a homogeneous-hazard value of 4.5 bp, the same sum at $\phi = 1$. The reported series peaks near eight to ten years (Table 19 and Figure 18).

Table 19: Capital gains by maturity (bp). The reported column is incidence-weighted; the first three columns and the homogeneous column are diagnostics.

T	cond. on sale	max over h	ϕ	$P(\text{hold})$	homogeneous	reported	band
2	0.0	0.0	–	1.00	0.0	0.0	[0.0, 0.0]
4	15.7	15.7	1.00	0.60	6.3	6.3	[4.1, 6.4]
6	16.5	17.6	1.00	0.45	9.1	9.1	[4.8, 9.3]
8	15.9	17.4	0.88	0.40	10.2	9.7	[4.7, 10.2]
10	14.9	17.5	0.82	0.37	10.4	9.5	[4.5, 10.1]
12	14.0	16.8	0.73	0.39	10.1	8.4	[3.9, 9.0]
14	13.3	16.1	0.67	0.40	9.7	7.5	[3.3, 7.8]
16	12.3	15.4	0.58	0.44	9.7	6.1	[2.8, 6.1]
18	11.1	14.1	0.57	0.44	8.8	5.2	[2.4, 5.3]
20	10.0	13.2	0.57	0.43	8.1	4.6	[2.2, 4.9]
22	8.9	12.1	0.58	0.42	7.3	4.1	[2.0, 4.5]
24	8.0	11.2	0.59	0.41	6.5	3.6	[1.8, 4.0]
26	7.3	10.5	0.60	0.40	5.8	3.3	[1.6, 3.6]
28	6.4	9.3	0.60	0.40	5.1	2.9	[1.5, 3.2]
30	5.7	8.6	0.60	0.40	4.5	2.6	[1.3, 2.8]

A.5.1.4 Adjustment III: liquidity

The object is the difference in liquidity price discounts between the two markets rather than the municipal market’s own liquidity cost.

Construction. Round-trip costs are imputed within $\text{cusip} \times \text{day} \times \text{size}$ cells containing both customer sides, as $c = (\text{max} - \text{min}) / \text{max}$ of transaction prices, and pre-averaged to bond-months so that heavily traded bonds do not dominate. Only institutional-size trades ($\geq \$250\text{k}$) on seasoned bonds with over a year remaining are kept. Trades whose size TRACE disseminates as a cap, 1MM+ and 5MM+, are read at their lower bounds, which clear the size floor and are used for nothing else. Treating them as missing would discard the largest trades on the corporate side, the only one where the caps apply. The two markets are compared only in common $\text{month} \times \text{rating-rank} \times \text{maturity-bin}$ cells with at least ten bonds per market cell; ratings are keyed at quarter-end and never forward-filled. Cells are

formed on both sides by letter family on the twenty-two notch agency scale, notch one for AAA, notches two through four for AA and five through seven for A, so a cell holds the same letter grade in both markets. The grades come from different agencies over different issuer populations, so the match is by stated grade rather than by measured default risk, though no longer by ordered rank across two different credit ranges. Aggregation is equally weighted at each level, ratings within a month, months within a year, and years.

Conversion to yield. The matched cost difference is amortized in the Amihud–Mendelson manner, $\Delta y = (c_m - c_c)/h$, with $h = 5$ years central and $h \in \{2, 10\}$ in the band, and with no tax-keep factor, since the round-trip cost is already an after-tax wealth loss. Each bin estimate sits at its midpoint, 3, 8, 13, 18, 23 and 28 years, and is interpolated linearly onto the even-maturity grid 2, 4, $\dots$, 30, flat beyond the end knots, so that the reported thirty-year figure is the 26–30 bin estimate itself. A step function across the estimation cells would put spurious jumps at bin edges into the total. A within-market regression of yields on liquidity with controls supplies an alternative cost-to-yield map, a fitted pricing coefficient in place of the $1/h$ amortization. It is a variant rather than an independent check. It is applied to the same matched cell difference, so it rescales that difference by a constant, 0.876 in every bin, and is equivalent to amortizing at a slightly different horizon.

The band combines a method envelope over the horizons and the regression map with ± 1.96 times the year-clustered standard error of the cost difference, which is 0.9–1.8 bp at $h = 5$, as large as the estimates themselves. That standard error is the dispersion of the annual mean differences divided by the square root of the number of years, the sampling error of the estimator’s own equally weighted average over years. Trading frequency, turnover and a quoted-gap proxy estimate different objects on different scales from the imputed round-trip cost, so a spread across them is not an uncertainty interval for any one estimand; they are retained as diagnostics only.

Result. Matched municipal institutional round trips are close to corporate ones, cheaper inside twenty years and dearer beyond: 0.1 bp at thirty years, band $[-2.2, 2.3]$ bp (Table 20, Figure 19), and the component is not statistically distinguishable from zero at any maturity. The 21–25 bin, the one bin where the municipal cost is materially the higher, lifts the long end of the profile toward zero. Its positive point estimate is barely two standard errors from zero and does not survive the horizon variants, so it is not read as a sign reversal. Liquidity does not explain the municipal premium; if anything it slightly deepens it.

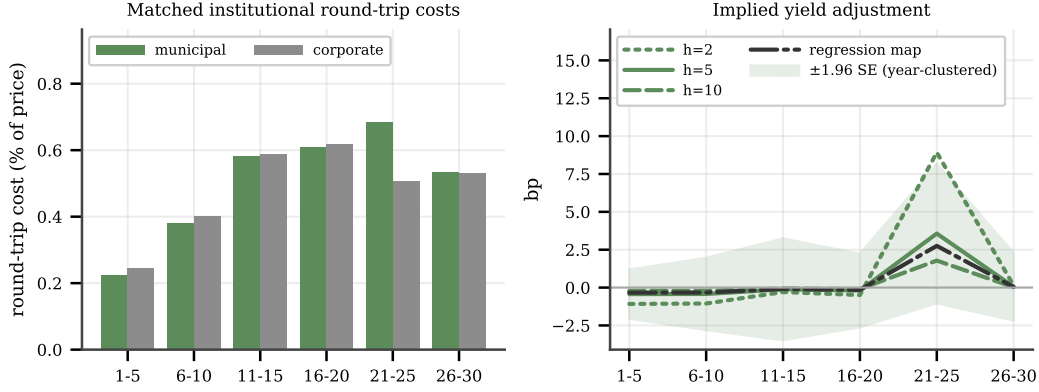


Figure 19: Left: matched institutional round-trip costs by maturity bin, in percent of price, for the two markets. Right: the implied yield adjustment under three amortization horizons and under the regression map, with the ± 1.96 year-clustered standard-error band at $h = 5$ shaded.

Alt text: Transaction-cost comparisons and holding-horizon conversions show the estimated liquidity contribution to the yield gap. The panels distinguish price-cost proxies from annualized yield effects and display their uncertainty.

Table 20: Liquidity by maturity bin. Round-trip costs in percent of price; adjustments in bp under three amortization horizons and the regression map.

bin	muni RTC	corp RTC	difference	$h = 2$	$h = 5$	$h = 10$	regr.
1-5	0.225	0.246	-0.022	-1.08	-0.43	-0.22	-0.33
6-10	0.382	0.403	-0.021	-1.05	-0.42	-0.21	-0.33
11-15	0.582	0.588	-0.006	-0.30	-0.12	-0.06	-0.09
16-20	0.609	0.619	-0.010	-0.49	-0.20	-0.10	-0.15
21-25	0.684	0.506	0.178	8.92	3.57	1.78	2.75
26-30	0.533	0.531	0.003	0.13	0.05	0.03	0.04

Sensitivity to the cell rule. The ten-bond rule is one of several discretionary matching choices, and Table 21 varies it; the size floor and the corporate comparison sample move the differential by more and are not varied here. Relaxing it to a single bond admits thin cells dominated by one or two prints, dragging the 26–30 bin from 0.05 to -6.53 bp; tightening to twenty makes several bins too thin to estimate, moving 11–15 from -0.12 to -2.99 bp. Ten is a compromise between admitting thin cells and discarding usable ones; the estimates keep moving beyond it.

Table 21: Liquidity adjustment at $h = 5$ (bp) under alternative minimum bond counts per market cell.

bin	≥ 1	≥ 5	≥ 10 (used)	≥ 20
1-5	-0.59	-0.50	-0.43	-0.39
6-10	-0.79	-0.49	-0.42	-0.27
11-15	-0.12	0.42	-0.12	-2.99
16-20	-0.09	-0.30	-0.20	-0.31
21-25	2.05	2.20	3.57	1.81
26-30	-6.53	-5.81	0.05	0.30

A.5.1.5 Adjustment IV: default

Method. Following Fleckenstein and Longstaff (RFS 2025, §5.1), each leg’s default component is the after-tax cost of CDS protection,

$$L = \text{CDS} \times (1 - \tau_c), \quad (112)$$

with τ_c the maximum statutory corporate rate (35% through 2017, 21% from 2018), the marginal CDS investor being a corporate entity that deducts its protection payments. The municipal leg matches each state general-obligation issue to its own state’s CDS curve (Markit) in quarter $\times$ rating $\times$ maturity cells; the benchmark leg is one pooled AA corporate curve by quarter $\times$ maturity. The quarterly join keeps 87% of bond-months; annual CDS cells instead move the thirty-year component by under 0.1 bp. Beyond the observed tenors the state curve is extrapolated linearly in the cumulative hazard $H(T) = T \text{ cds}(T)$ and floored at zero before staging; the line crosses zero for two per cent of filled cells. Each leg is staged separately by the equal-weighted three-stage mean (maturity $\times$ year $\times$ month $\times$ rating cells, then rating buckets within a month, then months) before the difference $\Delta_{\text{def}}(T) = (\bar{L}_m - \bar{L}_c)(1 - \tau_L)$ is taken; the two do not commute, and the order matters at the short end.

The baseline credit spread is the linear after-tax object $s(1 - \tau_c)$; the exact conversion, adding the protection PV to the par bond’s price and re-solving its yield, is a band variant. The two differ by about 0.9 bp at the median cell but far more in crisis cells where the after-tax CDS exceeds the bond’s own yield (a 3.7% thirty-year municipal yield cannot contain a 4.6% credit spread). Both legs shrink together, and the thirty-year implied-tax shift falls from 3.16 to 2.35 percentage points (Table 23), inside the band. The linear object is the standard construction and the conservative side under the project rule, and the two diverge where the CDS-cash basis is widest, which the basis scalings in the band already discount.

Subtracting each leg’s own default component inside the ratio $1 - y_m/y_c$ is wrong: mu-

municipal default compensation sits in an exempt coupon, corporate in a taxed one. With τ_L the rate at which credit losses are deductible, the indifference condition

$$y_m - L_m(1 - \tau_L) = y_c(1 - \tau) - L_c(1 - \tau_L) \implies \tau = 1 - \frac{y_m - (L_m - L_c)(1 - \tau_L)}{y_c}, \quad (113)$$

leaves only the difference $L_m - L_c$, scaled by $(1 - \tau_L)$; $\tau_L = 0.20$ centrally and $\tau_L \in \{0, 0.37\}$ in the band. The two legs' credit spreads differ by only a couple of basis points, so the correct adjustment is small. Putting both legs' raw CDS inside the ratio, scaling the differential at the wrong rate, or transplanting the FL construction into the ratio each falls into the ratio-amplification trap of Section A.5.1.1 and gives an adjustment several times too large.

State CDS quotes thin with maturity: 72% of rating-matched cells are exactly observed at 1–10 years and 69% at 11–20 years, but only 15% at 21–30 years, where most cells are filled along the curve (Figure 20, right panel). The long end is an extrapolation whose band, spanning sample variants (quality-capped cells, 2007 onward), CDS-basis scalings of 0.75 and 0.50, the exact conversion, and the τ_L grid, is the informative object; the point estimate alone is not.

Result. At thirty years the component is 15.7 bp, band [7.9, 19.7] bp, about 11.2% of the CDS-matched spread and 9.8% of the all-bond denominator $\text{Spread}(T)$. On the wider sample it is positive at every maturity and roughly flat through the belly before rising at the long end (Table 22).

A.5.1.6 Adjustment V: the issuance premium

The municipal leg of $\text{Spread}(T)$ is a primary price and the corporate leg a fitted seasoned-market curve. The panel is a cross-section of 914,501 distinct bonds, each observed once at issuance, whose zero yield tracks the recorded offering yield at correlations of 0.998, 0.992 and 0.93 at five, ten and thirty years. The small wedge, growing with maturity, is the coupon-to-zero conversion. The issuance premium underwriters need to clear a deal inflates the municipal leg, and $\text{Spread}(T)$, without being a clientele premium.

Measurement. Market drift is differenced out. Over 2000–2020 market drift between offering and trade dwarfs any issuance premium. The premium (offering yield less early secondary yield, positive when the bond is issued cheap) is a difference in differences against seasoned bonds, aged over three years, in the same maturity bucket and week.

$$\text{issuancepremium} = [y_{\text{off}} - \bar{y}_{\text{seasoned}}(w_{\text{issue}})] - [y_{\text{trade}} - \bar{y}_{\text{seasoned}}(w_{\text{trade}})]. \quad (114)$$

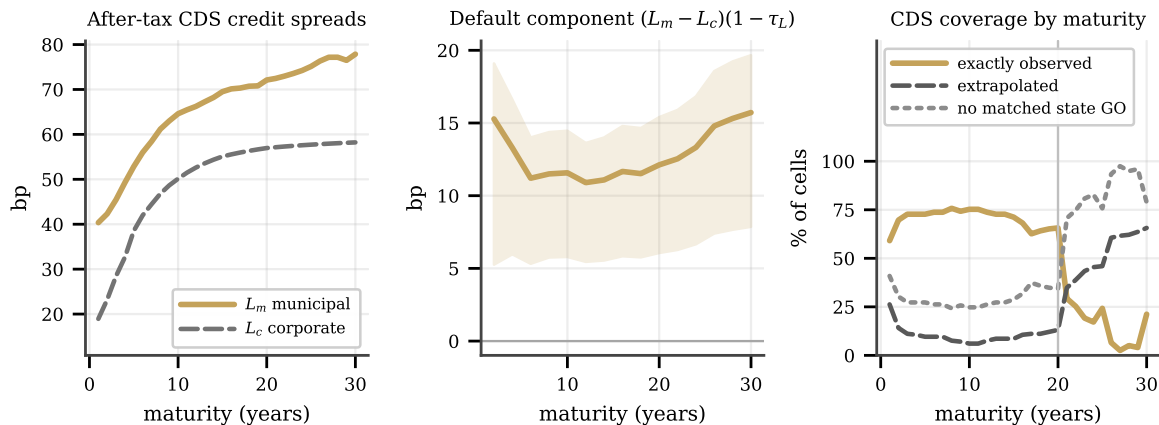


Figure 20: Left: after-tax CDS credit spreads for the two legs. Middle: the default component $(L_m - L_c)(1 - \tau_L)$ with its band. Right: the share of rating-matched CDS cells that is exactly observed, the share filled by extrapolation, and the share with no exactly observed state general obligation; the last overlaps the second and the three are not a partition. Coverage collapses beyond twenty years (vertical rule), which makes the long end an extrapolation.

Alt text: Municipal and tax-adjusted corporate CDS curves produce a maturity-specific default proxy. The figure also shows uncertainty and declining long-maturity quote coverage, including the region requiring extrapolation.

Table 22: Default by maturity. Yields in percent; spread and component in bp. The spread column is the CDS-matched sample, the population the credit legs are measured on.

T	muni yield	corp after-tax	spread (bp)	component (bp)	band
2	1.56	1.61	-4.4	15.28	[5.26, 19.10]
4	1.99	1.97	1.8	13.29	[5.97, 16.62]
6	2.39	2.30	9.3	11.20	[5.32, 14.00]
8	2.79	2.60	18.7	11.50	[5.75, 14.37]
10	3.12	2.84	28.9	11.57	[5.79, 14.47]
12	3.47	3.01	46.1	10.90	[5.45, 13.62]
14	3.73	3.12	61.0	11.08	[5.54, 14.00]
16	3.93	3.19	73.9	11.67	[5.83, 14.79]
18	4.09	3.24	85.0	11.52	[5.76, 14.65]
20	4.26	3.28	98.4	12.12	[6.06, 15.40]
22	4.36	3.30	106.1	12.54	[6.27, 15.90]
24	4.47	3.32	114.7	13.32	[6.66, 16.83]
26	4.59	3.34	124.8	14.80	[7.40, 18.57]
28	4.58	3.36	122.9	15.31	[7.66, 19.23]
30	4.77	3.37	139.7	15.72	[7.86, 19.66]

Table 23: Default sensitivity: shift in the implied tax rate (percentage points) under alternative samples, CDS-basis scalings, and the exact-conversion variant.

T	2007+	basis 0.50	basis 0.75	exact conversion	full	quality-capped
2	5.39	2.22	3.27	4.27	4.28	5.40
6	4.02	1.24	1.84	2.30	2.43	3.10
10	3.48	1.10	1.64	1.94	2.17	3.08
14	2.81	1.04	1.55	1.75	2.06	2.77
18	2.53	1.10	1.64	1.78	2.17	2.63
22	2.46	1.23	1.83	1.92	2.43	2.50
26	2.69	1.49	2.21	2.25	2.93	4.25
30	2.62	1.61	2.39	2.35	3.16	3.81

Rates at either date and the bond’s own credit quality cancel.

The same side of the market. A reoffering is a customer buy, so the comparison uses dealer-to-customer prints, also customer buys; customer sells would put the bid–ask spread inside the estimate and double-count the liquidity component of Section A.5.1.4.

The curve’s own population, dates and staging. Only the panel behind the curve, with its non-missing filters, is used; offering yields merged from all of Mergent would leave only 40% of the estimation bonds inside the target universe. Trade age and the issue week are dated by the offering date, as the curve’s yield is. The dated date differs from it by a median of fourteen days, by more than a week for 83.6% of bonds and by more than a month for 8.1%, and precedes it for 13.7%. A window built on it is really weeks two to six, with its drift benchmark in the wrong week. Averaged within the thirty-day window and staged the same way as the curve (maturity bucket $\times$ issue quarter $\times$ half-notch rating cells, then across cells, then across quarters), the premium gives $\Delta_{\text{prem}}(T)$. An unstaged par-weighted mean falls well below the staged 21.1 bp at thirty years, because the largest-trade decile is a small share of bonds but most of the traded par and is priced more keenly. When-issued prints, the primary distribution itself, are excluded.

Sensitivity to the trade-age window. A discount earned at placement is there on day one, whereas residual rate drift would start near zero and grow with the window. The premium is present immediately (Figure 21, left): 11.0 bp within three days of the offering and 12.3 bp out to ninety, with no trend between. The age bins are not composition-controlled beyond the staging, so only the level is read here, and not differences between bins. Nor does drift manufacture the maturity gradient that feeds $\Delta_{\text{prem}}(T)$. Across the five windows of Table 24, the gradient is present from the first week and does not steepen. The short buckets fall with the window, as drift would have them, but the long buckets, which carry the

adjustment, are stable from a week out, and only the thinnest, three-day column understates the longest bucket.

Table 24: Issuance premium by maturity bucket (bp) under alternative trade-age windows. The baseline estimate uses thirty days.

maturity bucket (yr)	3	7	14	30	60
1	4.5	6.7	5.6	3.6	2.2
5	5.9	10.1	7.6	6.1	5.0
8	7.2	10.5	9.0	7.9	7.4
12	15.5	16.2	13.8	12.8	12.4
17	15.0	17.3	16.7	16.6	16.7
23	18.1	18.7	20.0	21.1	21.0

Notes. Column headings are the trade-age window in days. Each cell is the drift-controlled difference-in-differences premium, staged as the curve is staged, for bonds whose first observed print falls inside the window.

Two further tests. Not primarily a yield-convention artefact. Long municipals are typically issued at a premium with a ten-year par call, so offering yields quoted to maturity against secondary prints at yield-to-worst would give premium bonds a spurious gap growing with maturity, the observed shape. Splitting on offering price speaks against this: at 17–23 and 23–31 years, par and discount bonds, for which the two conventions coincide, show 16.5 and 20.2 bp against premium bonds’ 15.7 and 21.3 bp.

The size gradient, as a diagnostic. At the long end every average-trade-size decile shows a positive premium, from 9.8 to 19.0 bp, the largest-trade decile lowest (Figure 21, right), so no single slice of the size distribution carries the estimate. Dealer markup and placement discount are entangled in the small-trade cells where markups are largest, so the pattern is no proof of mechanism.

Result. The issuance premium rises steeply with maturity, from 4.3 bp at four years to 21.1 bp at thirty (Table 25). Inference is a 10,000-draw moving-block bootstrap over the 64 issue quarters of the estimator’s quarterly staging, conditional on the matched sample and the estimated seasoned benchmark, neither re-drawn. The staged premium is serially correlated at the short end (lag-one autocorrelation 0.35 in the 1–5 and 5–8 year buckets, between -0.02 and 0.18 in the middle buckets, about 0.28 in the longest). Quarters are therefore drawn in circular blocks of four, preserving multiplicity, which materially widens the short-end interval (1–5 years: $[2.58, 4.62]$ under iid quarters against $[2.22, 5.03]$ with blocks) and leaves the long end nearly unchanged. The intervals exclude zero at every maturity.

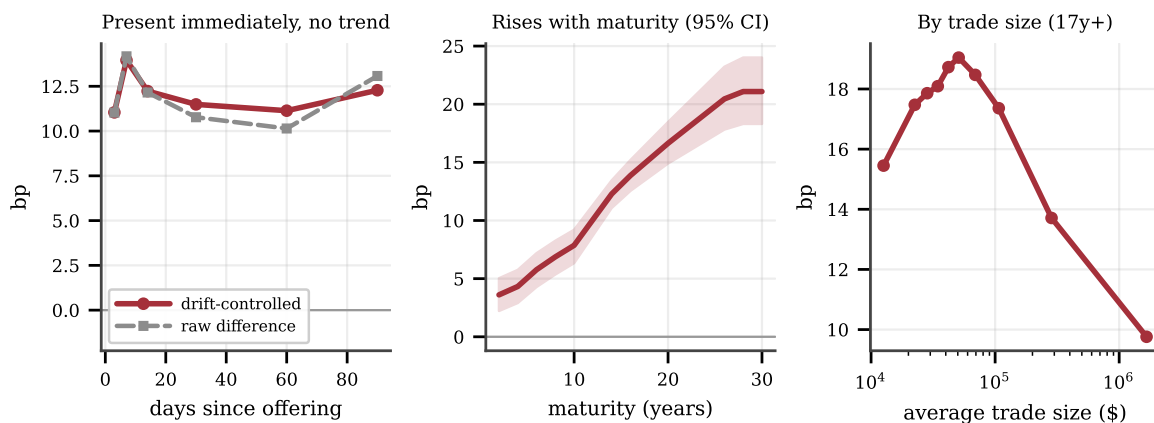


Figure 21: Left: the issuance premium by trade age, drift-controlled and raw, flat from the first days rather than trending. Middle: the interpolated premium by maturity with its bootstrap band. Right: by average trade size at the long end, where every decile shows a substantial issuance premium.

Alt text: The panels compare issuance premiums across post-issuance trade ages, maturities, and trade sizes. They show how the estimated premium depends on matching and aggregation choices.

The measure captures the primary-versus-seasoned mismatch that the curve comparison needs; dealer markups, ongoing distribution, liquidity seasoning and price discovery may also contribute to the post-offering yield change, so it is a proxy rather than a bound on the issuer's placement cost. Only the 16 % of curve bonds with a print inside thirty days of offering contribute; the direction of that selection is not bounded. The benchmark matches week and maturity bucket only, so movements in credit spreads or call exposure between offering and trade survive the same-bond differencing. The trade tape begins in 2005 while the curve spans 2000–2020, so the estimate is made on later issues and transported to the whole curve; it also inherits any staleness in recorded offering yields.

A.5.1.7 The combined effect

Table 25 is the full ledger and Figure 22 plots its components with their bands.

The shape matters more than the level. The explained share falls from 60 % at twelve years to 49 % at twenty and 43 % at thirty, while the residual grows from 29 bp to 63 and 91 bp. The call adjustment and the issuance premium rise with maturity but the spread rises faster; capital gains and liquidity are flat or shrinking, and default rises only modestly. Whatever generates the long-end premium therefore strengthens with maturity and is not an option, a tax-timing wedge, a transaction cost, credit, or a placement discount: the pattern a maturity-segmented clientele produces.

Reading the short end. At two and four years the observed spreads, 9.9 and 16.9 bp, are below the summed adjustments, 18.6 and 23.8 bp, so the residuals are negative; capital gains, zero at two years, are not the cause.

The capital-gains profile peaks near eight to ten years rather than at the long end for two reasons. Amortizing a resale wedge over the remaining life gives, for a sale at $h = 2$, a compensation multiplier $(h + \text{ann}_{T-h})/h$, with ann_n the present-value annuity factor over n years. This multiplier equals 1.96 at $T = 4$ against 10.38 at $T = 30$, so the same wedge demands roughly five times the annual yield at the short end. And the de minimis threshold, a quarter point of par per year of remaining maturity, is 0.50% at resale after two years for an originally four-year par bond against 7.00% for a thirty-year one. Short-end discounts more often breach it and are taxed at ordinary rates. The measured probability of ordinary market discount is 0.30 at $T = 4$ against 0.23 at $T = 30$, the de minimis share moving the opposite way.

The components are basis-point shifts from a common base curve, so their sum is order-invariant. Cross-friction interactions are set to zero; the OAS refit absorbs parallel price-level differences but does not identify them.

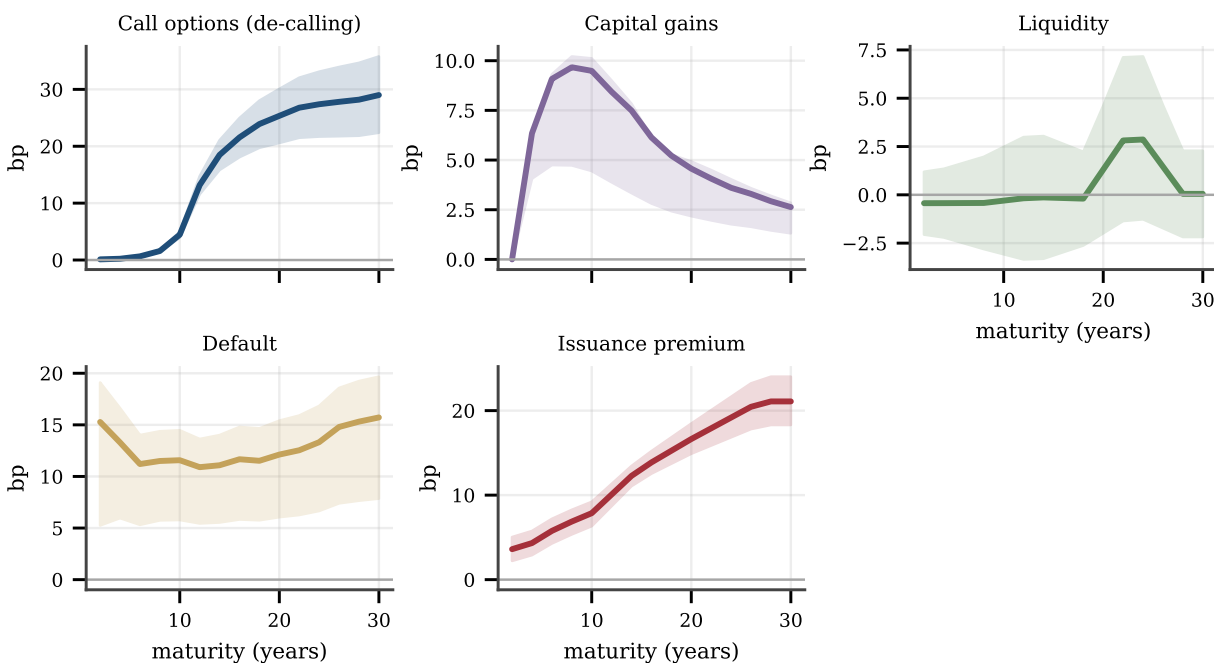


Figure 22: The five components with their bands.

Alt text: Five separate panels or series show the call, capital-gains, liquidity, default, and issuance-premium contributions to the relative-yield gap, together with the stated component uncertainty or scenario bands.

Table 25: Master ledger: the five adjustments by maturity, in bp of $\text{Spread}(T) = y_m - y_c^H$. Brackets are component bands; the total band is their interval sum. The final column is the total as a share of the raw spread $\text{Spread}(T)$, which is the natural scale for a component-by-component listing; it is not the share of the puzzle reported in Table 15, where the call and the issuance premium have first been removed. The share column is suppressed where the observed spread is below 20 bp.

T	call pt band	capital gains pt band	liquidity pt band	default pt band	issuance premium pt band	total pt band	obs. Spread(T)	% expl.
2	0.1 [0.1, 0.1]	0.0 [0.0, 0.0]	-0.4 [-2.1, 1.2]	15.3 [5.3, 19.1]	3.6 [2.2, 5.0]	18.6 [5.5, 25.4]	9.9	—
4	0.2 [0.2, 0.2]	6.3 [4.1, 6.4]	-0.4 [-2.2, 1.3]	13.3 [6.0, 16.6]	4.3 [2.9, 5.7]	23.8 [10.9, 30.3]	16.9	—
6	0.6 [0.6, 0.7]	9.1 [4.8, 9.3]	-0.4 [-2.5, 1.7]	11.2 [5.3, 14.0]	5.8 [4.3, 7.2]	26.3 [12.4, 32.9]	27.9	94
8	1.6 [1.4, 1.8]	9.7 [4.7, 10.2]	-0.4 [-2.8, 2.0]	11.5 [5.7, 14.4]	6.9 [5.4, 8.3]	29.2 [14.4, 36.6]	41.1	71
10	4.5 [3.9, 5.0]	9.5 [4.5, 10.1]	-0.3 [-3.1, 2.5]	11.6 [5.8, 14.5]	7.9 [6.3, 9.2]	33.1 [17.4, 41.2]	52.2	63
12	13.2 [11.6, 14.7]	8.4 [3.9, 9.0]	-0.2 [-3.3, 3.0]	10.9 [5.4, 13.6]	10.1 [8.7, 11.3]	42.4 [26.3, 51.7]	71.2	60
14	18.5 [15.8, 21.0]	7.5 [3.3, 7.8]	-0.1 [-3.3, 3.0]	11.1 [5.5, 14.0]	12.3 [11.1, 13.4]	49.2 [32.5, 59.4]	87.0	57
16	21.6 [18.1, 24.9]	6.1 [2.8, 6.1]	-0.2 [-3.0, 2.6]	11.7 [5.8, 14.8]	13.9 [12.6, 15.2]	53.1 [36.3, 63.7]	99.4	53
18	23.9 [19.7, 28.0]	5.2 [2.4, 5.3]	-0.2 [-2.6, 2.2]	11.5 [5.8, 14.7]	15.3 [13.7, 16.8]	55.7 [39.0, 67.0]	110.3	50
20	25.3 [20.6, 30.1]	4.6 [2.2, 4.9]	1.3 [-2.0, 4.6]	12.1 [6.1, 15.4]	16.6 [14.9, 18.5]	60.0 [41.7, 73.5]	123.1	49
22	26.8 [21.5, 32.0]	4.1 [2.0, 4.5]	2.8 [-1.4, 7.1]	12.5 [6.3, 15.9]	17.9 [15.9, 20.0]	64.1 [44.3, 79.6]	128.8	50
24	27.4 [21.7, 33.1]	3.6 [1.8, 4.0]	2.9 [-1.3, 7.2]	13.3 [6.7, 16.8]	19.2 [16.8, 21.6]	66.4 [45.7, 82.7]	136.9	48
26	27.8 [21.8, 33.9]	3.3 [1.6, 3.6]	1.5 [-1.7, 4.6]	14.8 [7.4, 18.6]	20.5 [17.8, 23.2]	67.8 [46.9, 83.9]	145.0	47
28	28.2 [21.9, 34.6]	2.9 [1.5, 3.2]	0.1 [-2.2, 2.3]	15.3 [7.7, 19.2]	21.1 [18.3, 24.0]	67.6 [47.1, 83.4]	145.6	46
30	29.0 [22.4, 35.8]	2.6 [1.3, 2.8]	0.1 [-2.2, 2.3]	15.7 [7.9, 19.7]	21.1 [18.3, 24.0]	68.5 [47.7, 84.6]	159.8	43

The model against the observed curve. Figure 23 adds the five frictions to the option-free model curve of Figure 6, with the same accounting, and compares the sum with the observed curve. At the separately estimated adjustment values, the root-mean-square residual is 6.55 bp and the largest absolute residual 13.4 bp at twenty-eight years; the residual is -1.3 bp at twelve years, -1.5 at twenty and -3.5 at thirty, against a pre-adjustment gap of 65.0 bp at thirty. The equilibrium targets a different curve, so the figure is a diagnostic of the calibration rather than a test of the mechanism.

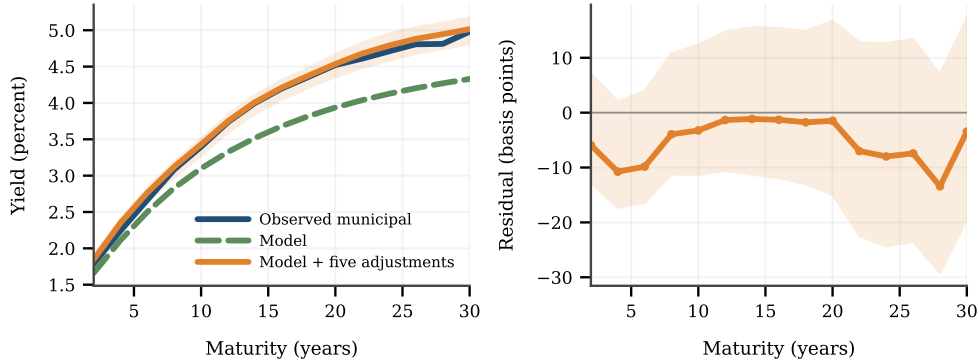


Figure 23: Left: the observed municipal curve, the model’s option-free curve, and the model once the five frictions are added, with the band spanning the low and high friction estimates. Right: the residuals. The band is the range of the separate adjustment scenarios, not a confidence interval; thirteen of fifteen point residuals lie within ± 10 bp.

Alt text: The model-only and model-plus-adjustment curves are compared with observed municipal yields. A second panel shows residuals and the range induced by the low and high adjustment scenarios.

A.6 Empirical designs: construction and robustness

A.6.1 Liquidity adjustment: round-trip-cost construction and regression tables

The round-trip cost of §A.5.1.4 is computed within each (CUSIP, date, par) cell containing both customer sides, from MSRB for munis and TRACE for corporates, as

$$RTC_{i,t} = \frac{P_{\max} - P_{\min}}{P_{\max}}.$$

The extremes need not be a matched dealer round trip and may include intraday price movement.

Tables 26 and 27 regress offering yields on full-life issuer liquidity proxies. The proxies

are not predetermined at issuance, so the coefficients describe association and give neither a causal ex-ante premium nor a formal bound. Trading frequency is the issuer mean, across its bonds, of MSRB trades per day from dated date to maturity. Column (1) indicates issuers above the median of the same measure among plain-vanilla corporate debentures and medium-term notes in TRACE, and Column (2) the top 25% of muni issuers.

Table 26: Issuer trading frequency and offering yield, by maturity bin.

	(1) Above corp median	(2) Top 25% trade
Maturity 1–5y	−0.046*** (0.009)	−0.028*** (0.007)
Maturity 6–10y	−0.007** (0.003)	0.021*** (0.005)
Maturity 11–15y	−0.011*** (0.004)	0.041*** (0.001)
Maturity 16–20y	−0.035*** (0.003)	0.024*** (0.006)
Maturity 21–25y	−0.091*** (0.009)	−0.041** (0.016)
Maturity 26–30y	−0.093*** (0.030)	−0.082*** (0.018)
Observations	785,256	785,256
Maturity × Rating × Offer-year FE	Yes	Yes
SE clustered at maturity-year level	Yes	Yes

Notes. Sample: federally tax-exempt muni issuance rated AA and above, agency score ≤ 3 on the twenty-two-notch scale, 2000–early 2021. Dependent variable: offering yield, in percent (a coefficient of -0.046 is -4.6 basis points). Each row is the bin-specific coefficient on the issuer-level treatment dummy. Column (1): issuer’s average daily MSRB trades per bond (per maturity-day) exceed the median across plain-vanilla corporate debentures in TRACE. Column (2): issuer is in the top 25% of average daily MSRB trades across muni issuers. No bond-count restriction is imposed. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The spread measure is the issuer mean, across its bonds, of bond-level average imputed round-trip cost in price units. Column (3) uses it continuously, interacted with maturity bin; Column (4) indicates the lowest 25% in average spread.

The 26–30-year indicator contrasts in Columns (1), (2), and (4) are about 8–15 basis points in absolute value. Column (3) is continuous, and 27.262 implies 27.3 basis points for a 0.01 spread contrast. Liquidity is priced, but the tables do not support a universal 10–15-basis-point maximum.

Table 27: Issuer transaction-cost spread and offering yield, by maturity bin.

	(3) Issuer-avg spread (cont.)	(4) Top 25% low spread
Maturity 1–5y	13.446*** (0.489)	−0.057*** (0.001)
Maturity 6–10y	11.624*** (0.216)	−0.048*** (0.002)
Maturity 11–15y	11.860*** (0.355)	−0.040*** (0.001)
Maturity 16–20y	13.447*** (0.557)	−0.043*** (0.003)
Maturity 21–25y	20.427*** (1.225)	−0.104*** (0.011)
Maturity 26–30y	27.262*** (4.815)	−0.146*** (0.037)
Observations	779,988	785,256
Maturity $\times$ Rating $\times$ Offer-year FE	Yes	Yes
SE clustered at maturity-year level	Yes	Yes

Notes. Sample: federally tax-exempt muni issuance rated AA and above, agency score ≤ 3 on the twenty-two-notch scale, 2000–early 2021. Dependent variable: offering yield, in percent. The issuer-average spread is the issuer-level mean of bond-level mean imputed round-trip costs, in bond-price units (0.01 corresponds to 1% of price, or 100 basis points). Column (3): coefficient on the continuous issuer-average spread interacted with maturity bin (a coefficient of 13.4 means a 0.01-unit change in spread is associated with a 0.134-percentage-point yield change, i.e. 13.4 bp). Column (4): issuer is in the bottom 25% of average MSRB spread (the most-liquid issuers); no bond-count restriction. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.6.2 Paired issues: Build America Bonds (BABs) vs. tax-exempt

Sample and construction. The sample is Mergent Municipal Bond Database issues (CUSIP level) in the program window 2009Q2–2010Q4, rated AA and above (agency score ≤ 3), excluding AMT bonds. The yield measure is the stated offering yield, times 0.65 for BABs, at integer remaining maturities $m \in \{1, \dots, 30\}$. Figure 3 (§4.5) averages it within maturity-treatment cells. The A—and-above sample gives the same maturity slope, 0.033 against 0.034, and a monotone binned profile (Table 8).

Away from par, $0.65 \times \text{YTM}$ is not an exact after-tax IRR, because taxable BAB OID, tax-exempt OID, acquisition premium, coupon income, and basis can be treated differently. Exact normalization would need bond-level after-tax IRRs. The scaling is retained as a near-par approximation.

Regression specification. Let y_b denote the offering-yield measure for bond b , $EXMP_b$ an indicator for tax-exempt status (BAB the base category), and m_b remaining maturity at issuance. I estimate

$$y_b = \alpha_0 + \alpha_1 EXMP_b + \alpha_2 m_b + \beta (EXMP_b \times m_b) + \gamma y_{m_b, t_b}^C + \lambda_{i(b) \times m_b \times q(b)} + \varepsilon_b, \quad (115)$$

where $\lambda_{i(b) \times m_b \times q(b)}$ is an issuer-by-maturity-by-calendar-quarter fixed effect and standard errors are clustered by state.

Results. Table 8 reports both specifications with and without call matching; the call-matched columns interact the cell with Mergent’s optional-call flag, so within-cell call comparability, maintained in the base columns, is controlled rather than assumed. The tax-exempt gap widens monotonically with maturity in both specifications. Matching attenuates it mildly, and the bin gradient falls from 75.5 to 63.9 basis points, with every interaction still significant at the one per cent level. The attenuation sits where call value lives. The ≥ 25 interaction falls by fifteen per cent, while $[5, 10)$ is unchanged. BABs carry the call flag slightly more often than tax-exempts (49 against 39 per cent).

Interpretation. The estimates reject the constant-wedge benchmark for a maturity-increasing wedge, conditional on within-cell comparability in liquidity, coupon, and OID or premium treatment. They are also conditional on the 0.65 scaling, which imposes a 35% investor rate, a maintained approximation distinct from the program’s 35% issuer subsidy. The paper’s curve uses an effective household rate averaging 41.3%. A lower keep would widen the gap, so the gradient is understated. Four further limits apply. The 2009Q2–2010Q4 window is a short and unusual stretch of the municipal market. The taxable leg is a municipal rather than a corporate credit, which holds issuer and credit fixed without matching the paper’s benchmark. Agreement with the curve is close at both ends and ragged between, the $[15, 20)$ bin matching worst. The differencing is exact only if both issue types carried the same issuance premium, an assumption given their different investor bases.

A.6.3 The muni puzzle on unlimited general-obligation bonds

The default-risk proxy of Appendix A.5.1.5 references *state* credit, so it maps most directly to unlimited-tax general-obligation (GO) bonds. Figure 24 reproduces Figure 2 on that subsample (Mergent security code “K”). After the call and issuance-premium adjustments, the implied tax falls from 41.6% at two years to 27.8% at thirty, 13.8 percentage points

against 16.0 on the full sample. This evidence is complementary robustness, consistent with earlier findings that the long-end residual survives default and liquidity adjustments and high-grade or GO restrictions (Longstaff, 2011; Fleckenstein and Longstaff, 2025; Chalmers, 1998).

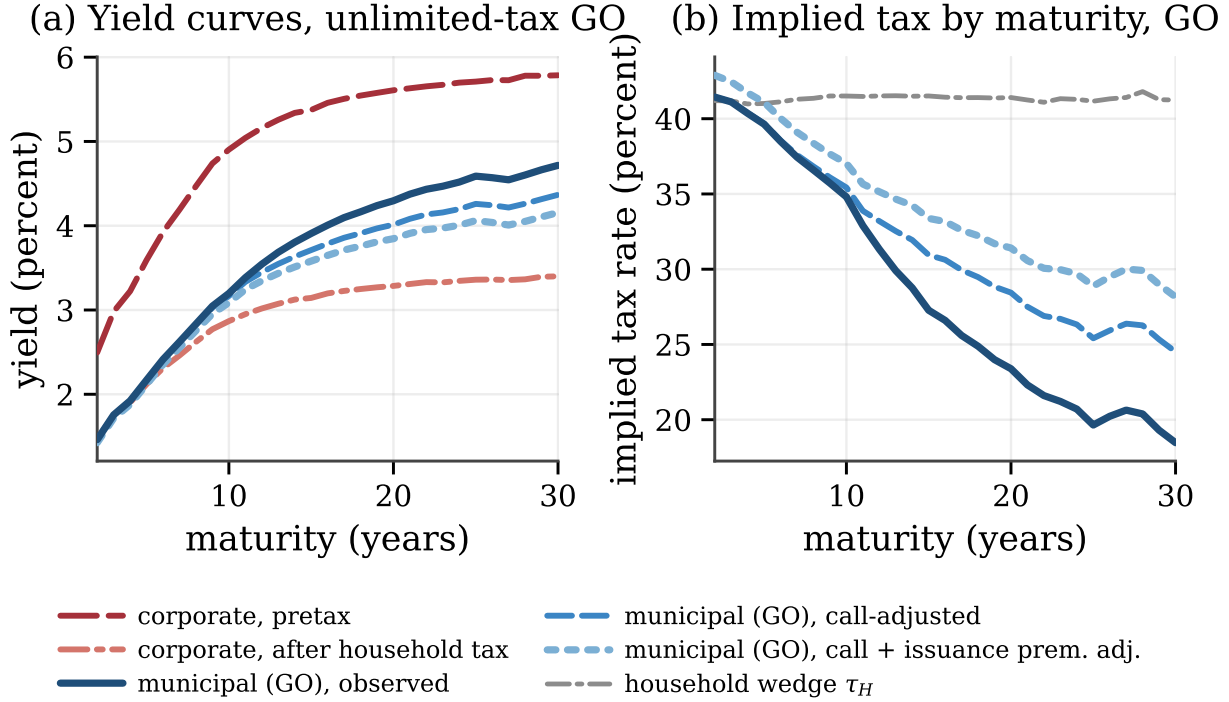


Figure 24: The muni puzzle on the unlimited-tax general-obligation subsample, on the construction of Figure 2: the same after-household-tax benchmark $\xi_{HY_T^C}$, the same staging, and the municipal curve shown as observed, call-adjusted, and call-plus-issuance-premium adjusted. The sample is Mergent security code “K”, whose credit is the issuer’s own full faith and unlimited taxing power rather than a specific project’s revenue, and retains 481,867 bond-months. The call adjustment is this subsample’s own decalled-minus-raw gap. The issuance premium is not separately identified on general-obligation bonds, since the estimator needs offering-week matches the subsample cannot support, so the all-bond series of §A.5.1.6 is carried over. After both adjustments the implied tax still falls 13.8 percentage points between two and thirty years, against 16.0 on the full sample, so the residual puzzle is not a revenue-bond or credit-composition artifact.

Alt text: The curve and implied-tax comparisons are repeated for unlimited-tax general-obligation municipals. The restricted sample retains the widening long-maturity yield gap and declining implied-tax profile.

A.6.4 What the household and bank designs distinguish

The restriction is joint. Household state-tax pass-through declines with maturity while capitalization of the bank-only benefit increases. P1 varies a resident-household benefit and

bank qualification varies only banks' interest-expense deduction, so the designs vary the identity of the tax beneficiary, not merely maturity.

The estimated maturity contrasts reject the representative-household benchmark, unit pass-through with a zero bank-qualified premium at every maturity subject to the tax-ranking condition in (51), and the bank-only annual-carry benchmark, a premium equal to the annual deduction benefit throughout. A non-tax maturity premium smooth at the cutoff cannot generate the discontinuity under the exclusion restriction, and the 2009–10 shift in the first-stage cutoff and the placebo cutoffs tie the treatment to the statute rather than a fixed issuer-size threshold.

The deduction benefit is common across maturities within a bank-year, but not necessarily across the banks buying different maturities. The all-bank normalization supplies a common benchmark, while bank sorting, expected funding costs, or designation-related services can affect capitalization. The selection checks of Appendix A.6.9 test for specific channels directly, and the P1 checks of Appendices A.6.5–A.6.6 vary the source of tax variation.

A richer tax-sensitive clientele model can generate both profiles without liability hedging. This locates the theoretical contribution in Result 1, which supplies a portfolio-based participation motive, and in Proposition 9, which supplies its cross-maturity consequences through shared portfolio risk, including on bonds the institution does not hold. The tax tests examine investor-specific transmission. The nonlinear calibration supplies the magnitude from balance-sheet inputs, and the untargeted profiles test the allocation of tax effects along the curve.

A.6.5 State-tax pass-through: full regression tables

This appendix reports the regression tables behind the P1 test of §7.1. Equation (49) is estimated on one sample, federally tax-exempt bonds rated AA and above, excluding Iowa, Wisconsin, Oklahoma, and Illinois and retaining bonds coded as state-tax exempt or issued where the state rate is zero, for three constructions of the municipal yield. The sample is unweighted, with year-by-month-by-maturity-by-callable effects and inference clustered at the state (47 clusters: 46 states and the District of Columbia). Clustering at state-year instead (986 clusters) treats a state's bonds as independent across years. It gives standard errors of 0.078 (OLS) and 0.096 (IV) in the shortest bin and 0.050 and 0.047 in the longest, with every bin significant at one per cent and a Kleibergen–Paap F of 139.

The state-clustered Kleibergen–Paap F of 10.9 measures the weakest linear combination of the five bracket first stages. The Sanderson–Windmeijer conditional F for each bin inter-

action given the others is 389, 929, 526, 718, and 520 under state clustering and 2,346, 6,314, 4,599, 3,520, and 3,694 under state-year, so no single bin interaction has a weak first stage. A state-level wild-cluster bootstrap on the OLS de-called column gives p -values against zero of < 0.001 , < 0.001 , 0.178, 0.398, and 0.438 across the five bins and 0.020 for the decline $\beta_1 - \beta_5$.

A.6.5.1 Offering yield, OLS and IV

Table 28 uses the vendor-quoted yield at issuance; the shortest bin is consistent with unit pass-through, which the longer bins reject. Because most of the sample is issued at a premium and nearly every long bond is callable, the quote is a yield to worst rather than to maturity, which understates long yields and steepens the profile relative to the zero-coupon columns.

Table 28: Implied-tax pass-through by maturity bin: offering yield.

	(1) OLS	(2) IV
β_1 : StateTax $\times$ [0,5)y	0.915*** (0.170)	0.922*** (0.218)
β_2 : StateTax $\times$ [5,10)y	0.538*** (0.090)	0.518*** (0.116)
β_3 : StateTax $\times$ [10,15)y	0.331*** (0.070)	0.336*** (0.085)
β_4 : StateTax $\times$ [15,20)y	0.193** (0.089)	0.211** (0.091)
β_5 : StateTax $\times$ ≥ 20 y	0.092** (0.045)	0.077 (0.048)
Observations	834,547	834,547
Rating control (mean rating)	Yes	Yes
Weights	None	None
Instrument: 1990 top state income tax rate	No	Yes
SE clustered at state (47 clusters)	Yes	Yes

A.6.5.2 Call-adjusted (decall) ZC yield, OLS and IV

The call-adjusted zero-coupon columns are those of Table 6. The zero-coupon yield is computed from secondary-market prices and re-priced under the calibrated equilibrium of Table 4, so that the yield and the model line come from one economy.

A.6.5.3 Non-call-adjusted (nondecall) ZC yield, OLS and IV

Table 29 leaves the call value in the price. The profile declines as before, so the call adjustment does not create it. The long-end estimates sit slightly above the de-called ones because the retained call value raises the measured municipal yield and lowers the implied tax most where the option is worth most.

Table 29: Implied-tax pass-through by maturity bin: non-call-adjusted (nondecall) ZC yield.

	(5) OLS	(6) IV
β_1 : StateTax $\times$ [0,5)y	0.995*** (0.145)	0.986*** (0.196)
β_2 : StateTax $\times$ [5,10)y	0.569*** (0.094)	0.549*** (0.119)
β_3 : StateTax $\times$ [10,15)y	0.382** (0.179)	0.359** (0.178)
β_4 : StateTax $\times$ [15,20)y	0.288 (0.210)	0.245 (0.198)
β_5 : StateTax $\times$ ≥ 20 y	0.214 (0.169)	0.176 (0.146)
Observations	834,553	834,553
Rating control (mean rating)	Yes	Yes
Weights	None	None
Instrument: 1990 top state income tax rate	No	Yes
SE clustered at state (47 clusters)	Yes	Yes

A.6.6 State-tax pass-through: instrument vintage and weighting

The IV columns instrument the current top state income-tax rate with the state's 1990 rate, a historical rate because it is predetermined and its variation reflects when and how each state built its income tax rather than current fiscal conditions. The further back the vintage, the more credible that exclusion argument, and the weaker the first stage. Table 30 runs the continuous-maturity version of equation (49), in which the maturity gradient of pass-through is a single coefficient, under seven ITEP vintages.

Relevance rises sharply from 1980. The conditional F for the gradient climbs from about 11 before 1980 to 38, 151 and 221 at 1980, 1990 and 2000. The gradient is stable where the instrument is strong, -0.046 , -0.042 and -0.041 of pass-through per year of maturity against -0.042 by OLS, whereas 1945 gives -0.050 and the weak 1950 through 1970 vintages -0.058 to -0.062 . The strong vintages converge on OLS, so the OLS profile is not an

endogeneity artefact. With equal weight per state (last two columns) the climb is steeper. The five largest issuing states, about two-fifths of the observations, are already ranked by their 1945 rates as by their current ones, so the 1945 rate's issuance-weighted correlation with the current rate is about 0.6 against 0.4 equal-weighted. The near vintages buy their strength with credibility. A 2000 rate is contemporaneous with the 2000–2020 sample and leaves no predetermined-history argument. The 1990 rate precedes the sample by a decade and is the earliest vintage whose state-clustered system F in the five-bracket specification clears ten (10.9, against 4.0 at 1980 and below three before that), which is why it is the main instrument.

Table 30: P1 instrument strength by vintage: the continuous-maturity specification.

Instrument vintage	corr. with current rate	unweighted (main)				equal weight per state	
		gradient	s.e.	KP F	SW F , gradient	gradient	KP F
OLS, no instrument	—	−0.0416***	(0.0109)	—	—	−0.0379***	—
1945	0.40	−0.0497**	(0.0200)	5.0	10.8	−0.0350	1.9
1950	0.31	−0.0577***	(0.0218)	5.1	10.8	−0.0816	1.3
1960	0.39	−0.0623***	(0.0209)	6.5	13.5	−0.0986**	1.9
1970	0.51	−0.0603***	(0.0169)	6.1	12.5	−0.0883***	4.3
1980	0.77	−0.0464***	(0.0131)	9.6	38.3	−0.0497***	23.0
1990 (main)	0.82	−0.0422***	(0.0120)	25.4	150.6	−0.0377**	23.4
2000	0.88	−0.0407***	(0.0108)	71.8	221.2	−0.0351**	37.1

Notes. Each row instruments the current top state income-tax rate and its product with years to maturity by the state's top rate in the year shown, from the ITEP historical series, and that rate's product with maturity, on the de-called zero-coupon implied state tax with the fixed effects and rating control of equation (49). The gradient is the coefficient on the maturity interaction, the change in pass-through per year of maturity. The correlation is the cross-state correlation of the vintage with the current rate. Standard errors and every F statistic are clustered at the state (47 clusters). The Kleibergen–Paap statistic is the two-regressor system F ; the Sanderson–Windmeijer statistic is the conditional F for the gradient alone. The first four columns are unweighted, as the main table is; the last two re-estimate each row with weight $1/n_{\text{state}}$, so that each state carries equal weight. Stars from the asymptotic normal p -value for the IV rows and from $t(G - 1)$ for OLS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In the five-bracket specification every vintage gives the same shape (Figure 25). The shortest bracket lies between 0.98 and 1.34, the longest between 0.08 and 0.18, and the decline $\beta_1 - \beta_5$ is rejected at the one per cent level in every column. The vintage changes the first stage; the point estimates stay fixed.

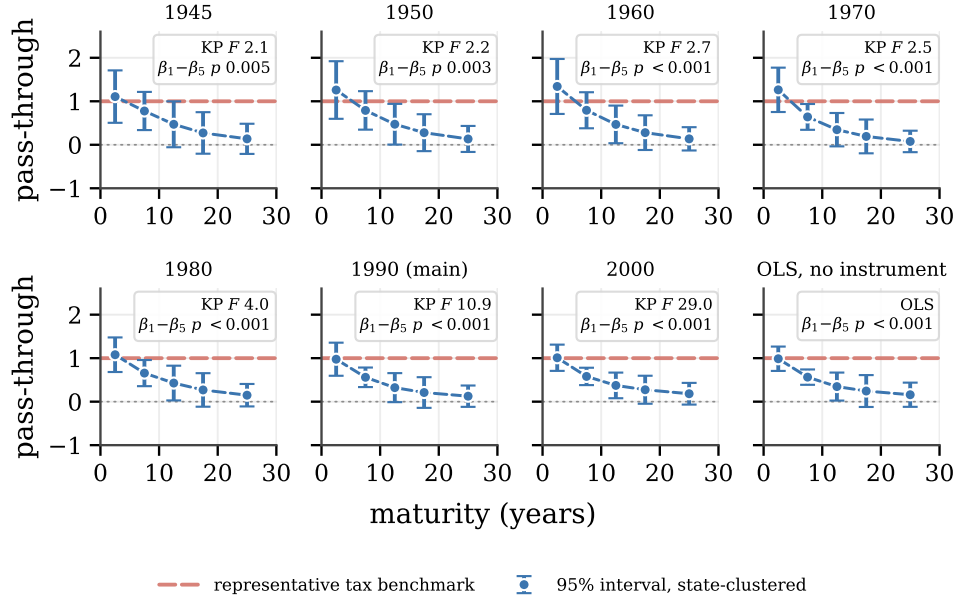


Figure 25: P1 five-bracket pass-through under each instrument vintage, and the OLS profile. Each of the first seven panels instruments the current top state rate, bracket by bracket, with the state's top rate in the year shown, from the ITEP series; the 1990 column is the main specification, and the last panel is the OLS profile with the same weighting and clustering. Points are the five bracket coefficients with 95 percent bands clustered at the state, unweighted, and the dashed line at one is the representative-tax benchmark. Each panel prints the Kleibergen–Paap system F , none for OLS, and the p -value of the decline $\beta_1 - \beta_5$. The 1990 and OLS panels are the estimates of Table 6.

Alt text: Pass-through profiles using different historical state-tax instruments are compared with OLS. The panels show a common declining maturity pattern and differences in precision across instrument vintages.

With the 1990 instrument fixed, equal weight per state, per state-year, or $1/\sqrt{n_{\text{state}}}$ moves the long-end bracket between 0.05 and 0.13 and leaves the decline significant at the one per cent level in every case.

With state fixed effects, the profile is identified from cross-state rate differences interacted with the maturity bins and from within-state rate changes. By OLS on the same 834,553 observations, state-clustered, the shortest-bin coefficient is 0.587 (s.e. 0.177), the twenty-plus coefficient -0.165 (s.e. 0.104), and the short-minus-long difference 0.752 (s.e. 0.150, $p < 0.001$). The bin interactions of the time-invariant 1990 rate are not generally absorbed by state fixed effects, so an interacted IV design is not ruled out.

Re-estimating the five-bracket specification of Table 6 47 times, each without one state, for OLS and for the 1990-rate IV, every one of the 94 profiles declines (Figure 26). The shortest-bin coefficient stays within 0.925–1.071 under OLS and 0.900–1.080 under IV, the longest-bin coefficient within 0.117–0.396 and 0.036–0.289, and the decline $\beta_1 - \beta_5$ within 0.68–0.91 and 0.76–0.96. The five-bracket Kleibergen–Paap F lies between 4.8 and 19.4.

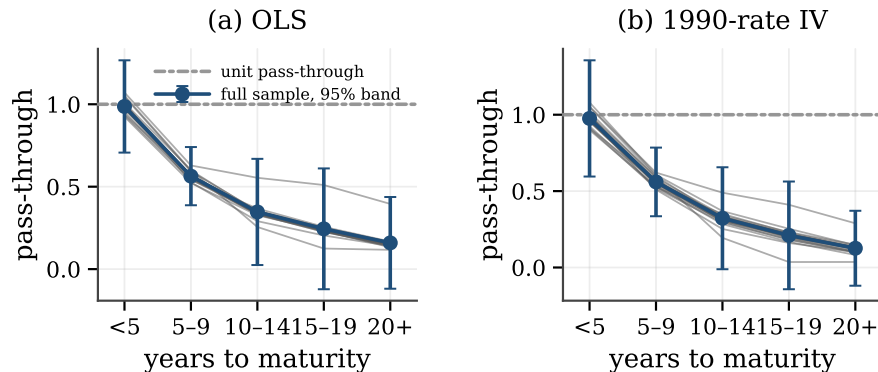


Figure 26: P1 leave-one-state-out. Each grey line is the five-bracket pass-through profile estimated without one of the 47 states, for OLS (panel a) and the 1990-rate IV (panel b); the blue line and band are the full-sample estimates of Table 6 with their 95 per cent state-clustered intervals, and the dashed line is unit pass-through. Decoll zero-coupon outcome, unweighted, standard errors clustered at the state.

Alt text: Forty-seven leave-one-state-out profiles are overlaid on the full-sample OLS and IV estimates. The declining household pass-through profile persists when any single state is omitted.

The implied state tax uses the multiplicative federal-and-state mapping before 2018 and the additive mapping from 2018 (§7.1); Table 31 estimates each period separately. Under IV the shortest-bin pass-through is 0.96 before 2018 and 1.04 from 2018 and the longest-bin pass-through 0.13 and 0.10, so the decline depends on neither the mapping nor the 2018 change in the household’s federal treatment.

Table 31: P1 by period: before and from 2018

	2000–2017		2018–2020	
	OLS	IV	OLS	IV
β_1 (under 5 years)	0.971*** (0.165)	0.958*** (0.210)	1.033*** (0.143)	1.038*** (0.206)
β_2 (5–9 years)	0.517*** (0.110)	0.509*** (0.129)	0.747*** (0.097)	0.790*** (0.137)
β_3 (10–14 years)	0.322* (0.191)	0.281 (0.192)	0.447*** (0.081)	0.519*** (0.120)
β_4 (15–19 years)	0.249 (0.199)	0.193 (0.190)	0.222 (0.144)	0.282* (0.162)
β_5 (20 and over years)	0.190 (0.151)	0.131 (0.131)	0.058 (0.112)	0.103 (0.134)
Decline $\beta_1 - \beta_5$	0.78 (0.18)	0.83 (0.20)	0.97 (0.19)	0.94 (0.25)
p -value, decline	0.000	0.000	0.000	0.001
KP F		10.9		9.5
Observations	688,213	688,213	146,340	146,340
Clusters (states)	47	47	47	47

Notes. The specification of Table 6, decall zero-coupon outcome, unweighted, standard errors clustered at the state, estimated separately on the years before 2018, when the implied state tax uses the multiplicative federal-and-state mapping, and from 2018, when it uses the additive mapping that is exact for a household whose SALT deduction is capped. The IV column instruments each bracket's current rate with the state's 1990 top rate. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.6.7 TCJA insurer-tax shock: static DiD and dynamic event study

Proposition 12 and the nonlinear model, in which the two-year bond is riskless and priced by households alone, predict that an insurer tax change leaves the short end unchanged and has a larger effect farther out. P2 asks whether states with larger predetermined P&C exposure display a larger long-end change. The TCJA, effective 2018Q1, cut the corporate rate from 35% to 21% and the top household ordinary rate from 39.6% to 37%. Under §832 proration the effective tax on P&C insurers' municipal interest stayed at 5.25%, so only the taxable side moved. The P&C keep ξ_{I_1} rose from 0.686 to 0.834. That 0.148 compression of the municipal advantage is the largest borne by any clientele, so P&C exposure is the treatment. Life insurers' keep rose only from roughly 0.79 to 0.843 (Appendix A.9.1.4), households' federal keep from 0.62 to 0.646, and tax-exempt institutions had no advantage to lose.

Quarter and event-time-by-maturity effects absorb the common component of the federal tax changes. Household responses may still differ across states, because the SALT cap

changes the in-state exemption's value and household composition may covary with insurer exposure. The specification controls for state taxes and their maturity interaction and instruments exposure with pre-reform domicile-based capacity. Under the covariance conditions in Appendix A.6.8.1 the omitted SALT channel attenuates the negative insurer gradient, consistent with IV estimates larger in magnitude than OLS.

Specification. On customer-side MSRB transactions over 2016Q3–2019Q1 in non-callable, AA-and-above, non-AMT munis,

$$\begin{aligned} \text{ImpliedTax}_{b,t} = & \beta_0(\text{PostTCJA}_t \times \text{PCShare}_{s(b)}) + \beta_1(\text{PostTCJA}_t \times \text{PCShare}_{s(b)} \times m_b) \\ & + \alpha_{\lfloor m_b \rfloor \times \text{Purchase}} + \alpha_{\text{issuer} \times \text{Purchase}} + \alpha_t + \Gamma' X_{b,t} + \varepsilon_{b,t}, \end{aligned} \quad (116)$$

where $\text{PostTCJA}_t = \mathbf{1}\{t \geq 2017\text{Q4}\}$ starts the post period at the announcement quarter and $\text{Purchase}_{b,t}$ indicates a customer purchase. The controls $X_{b,t}$ are the resident state tax rate and agency rating, each in levels and interacted with maturity; the pre-period life-insurer share interacted with maturity; and the P&C share and its maturity interaction. They also include state-tax interactions with post-reform status and event time, and event time interacted with maturity. The maturity profile therefore shifts freely each quarter, and β_1 is identified from cross-state variation in PCShare within a quarter. $\text{PCShare}_{s(b)}$ is the share of state s 's outstanding municipal debt held by property-and-casualty insurers at end-2015, two years before the reform. The differential effect at maturity m is $\beta_0 + \beta_1 m$, so β_0 is an extrapolation to maturity zero. The IV specifications instrument the P&C share with two domicile-based measures of insurer capacity,

$$\text{IV1}_s = \log\left(\frac{\sum_{i \in D_s} \text{Muni Assets}_i}{\text{Muni Outstanding}_s}\right), \quad \text{IV2}_s = \log\left(\frac{\sum_{i \in D_s^{\text{act}}} \text{Bond Assets}_i}{\text{Muni Outstanding}_s}\right), \quad (117)$$

where D_s is the set of P&C insurers domiciled in state s , D_s^{act} the subset holding any municipal bonds, Muni Assets_i and Bond Assets_i insurer i 's par holdings of municipals and of all bonds from NAIC Schedule D, and $\text{Muni Outstanding}_s$ the state's municipal par outstanding. Relevance comes from insurer home bias. The exclusion restriction, an assumption rather than a consequence of the construction, is that conditional on the fixed effects and controls domicile and capacity affect the post-TCJA maturity gradient only through local P&C exposure. The non-callable restriction keeps the post-TCJA collapse in advance refunding from driving the comparison.

Result. The static maturity gradient β_1 (Table 32) is -0.017 under OLS and -0.038 and -0.040 under IV1 and IV2 (issuer-clustered standard errors 0.005, 0.010, and 0.015). It is negative under all three estimators, as the model predicts, where the representative-tax benchmark predicts $\beta_0 = \beta_1 = 0$. At a one-standard-deviation difference in exposure (mean 0.096, standard deviation 0.039 across states), the post-reform decline in implied tax steepens by about 7 basis points per year of maturity under OLS and 15 under IV. That is roughly 3.0 percentage points between a five- and a twenty-five-year bond. Clustered by state, the treatment level, the OLS level term has a wild-cluster bootstrap p -value of 0.044 and the gradient 0.79, and first-stage F statistics of 1.4 and 1.0 make the IV estimates shape checks (Appendix A.6.8).

The issuer-clustered event study (Figure 27) supports the timing. Eleven of the twelve pre-TCJA coefficients across the three estimators are indistinguishable from zero at the five per cent level, and every column turns negative at 2018Q1 and stays negative, with five of the six post-reform quarters excluding zero under IV1 and four under IV2. On the wider A—and-above sample the gradient is unchanged or larger, OLS from -0.018 to -0.036 and IV1 from -0.037 to -0.039 , and the level term roughly doubles under the instrument (Table 40, whose columns omit the rating control of Table 32).

Table 32: TCJA difference-in-differences: implied-tax response to insurer exposure.

	(1) OLS	(2) IV1	(3) IV2
$Post \times PCShare$	-0.122^{***} (0.029)	-0.129 (0.084)	-0.017 (0.100)
$Post \times PCShare \times m_b$	-0.017^{***} (0.005)	-0.038^{***} (0.010)	-0.040^{***} (0.015)
Observations	1,733,881	1,733,881	1,733,881
Estimator	OLS	IV	IV
Instrument set	None	$IV1_s$	$IV2_s$
$[m_b] \times Purchase$ FE	Yes	Yes	Yes
$Issuer \times Purchase$ FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Rating control ($rating, rating \times m_b$)	Yes	Yes	Yes
SE clustered at issuer (6-digit CUSIP)	Yes	Yes	Yes

Notes. The dependent variable is the implied tax. All columns use the AA-and-above sample and the rating-controlled specification described above. Standard errors in parentheses are clustered by six-digit issuer. State-level inference is reported in Appendix A.6.8. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

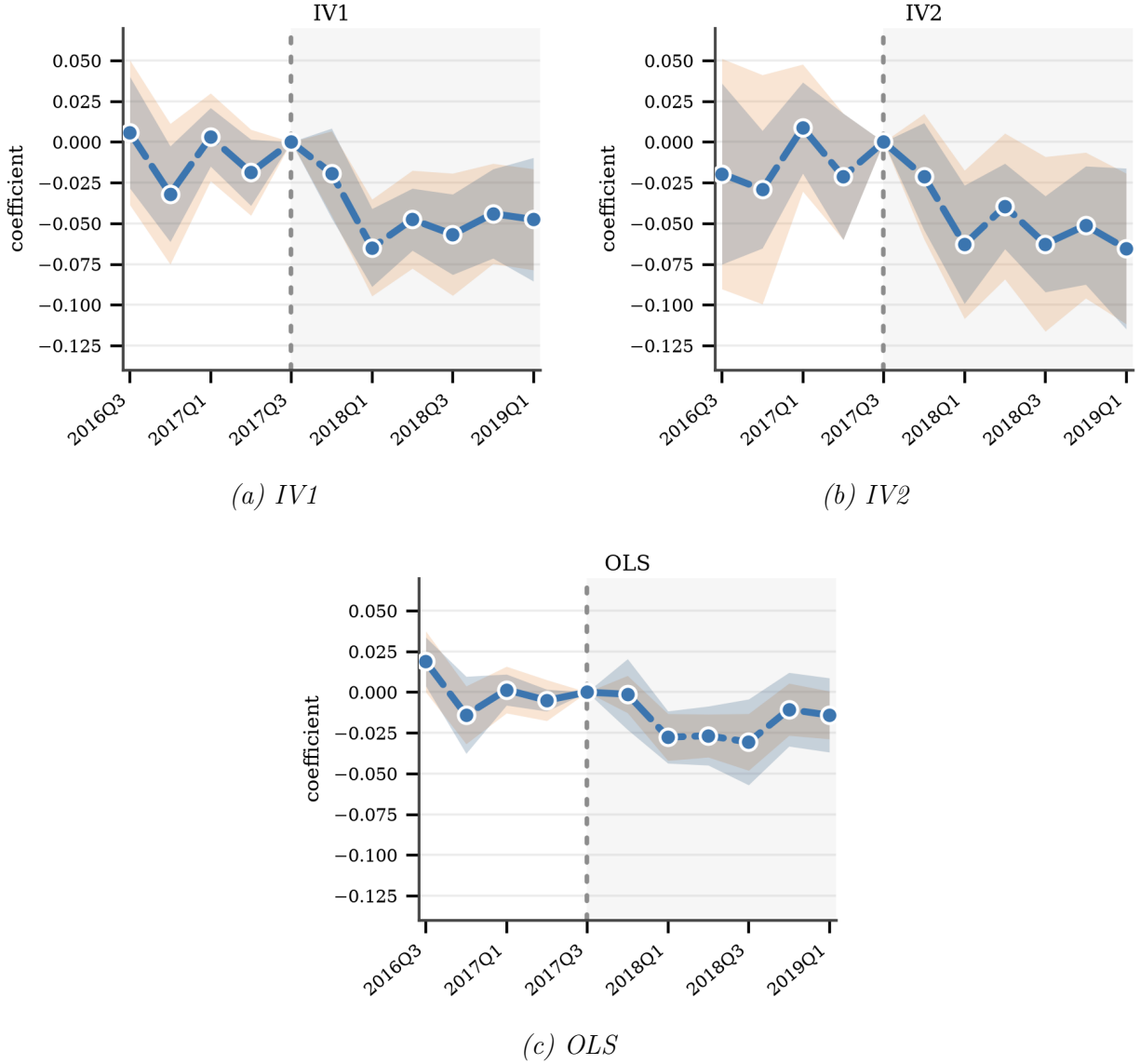


Figure 27: TCJA event-study maturity-gradient estimates under both clustering choices, omitted quarter 2017Q3. The coefficient path is identical across the two, so one path is drawn with two intervals around it: the darker band clusters on integer maturity, about thirty groups, and the lighter band on the six-digit CUSIP issuer, about ten thousand. Issuer clustering widens the instrumented intervals by roughly a fifth and narrows the ordinary least squares intervals by about a third. The number of post-reform quarters whose interval excludes zero is unchanged for OLS and for IV1, at three and five of six, and falls by one for IV2.

Alt text: OLS and two IV event-study profiles are shown with issuer- and maturity-clustered intervals. These alternative bands describe the same coefficient paths and do not replace treatment-level state inference.

A.6.8 OLS attenuation and IV in the TCJA event study

Because the treatment $PCShare_s$ varies by issuing state, the rating-controlled specification is reclustered by state (51 clusters) instead of by six-digit issuer. Point estimates are unchanged. The OLS maturity-gradient coefficient is -0.0170 with a state-clustered standard error of 0.0129 , against 0.005 under issuer clustering, and the post-TCJA level term is -0.122 with a state-clustered standard error of 0.061 . The state-clustered t test gives $p = 0.052$ for the level, and the gradient's t statistic is 1.3 . A wild-cluster bootstrap at the state level (Rademacher weights, 9,999 replications) gives $p = 0.044$ for the level term and $p = 0.79$ for the static gradient. The IV gradients are -0.0378 (IV1) and -0.0397 (IV2), with state-clustered standard errors of 0.0245 and 0.0342 .

Under issuer clustering the Kleibergen–Paap rk Wald F statistics are 25.7 for IV1 and 17.5 for IV2. Clustered by state they fall to 1.4 and 1.0 , so some joint linear combination of the three endogenous terms (one state-level instrument crossed with ym , the post-reform indicator, and their product) is weakly identified. The Sanderson–Windmeijer conditional F statistics are 264 , 321 , and 2156 under issuer clustering and 97 , 77 , and 125 under state clustering, the last for the maturity-gradient term, so no individual endogenous term is weak once the others are partialled out. The reduced form on the triple instrument is correctly signed at both clusterings, -2.75×10^{-4} with $t = -3.70$ by issuer and $t = -1.45$ by state. The IV estimates are therefore read as shape checks.

A.6.8.1 Why OLS attenuates the insurer channel in the presence of SALT-induced household demand

The regressor of interest is $X_{b,t} \equiv PostTCJA_t \times PCShare_{s(b)} \times m_b$, and the insurer-marginal long-end prediction for the implied tax $1 - y_{b,t}^{muni}/y_{T,b,t}^{corp}$ is $\beta_1 < 0$. The SALT cap is a second TCJA channel. It raises the value of the exemption for high-income households in high-tax states, and the added household demand lowers municipal yields mainly at the short end, where households are marginal. In reduced form,

$$\begin{aligned} ImpliedTax_{b,t} = & \beta_1 X_{b,t} + \theta_0 (PostTCJA_t \times Tax_s) \\ & + \theta_1 (PostTCJA_t \times Tax_s \times m_b) + \text{controls} + \text{FE} + \varepsilon_{b,t}, \end{aligned} \quad (118)$$

with Tax_s a state-level proxy for SALT exposure, $\theta_0 > 0$ and $\theta_1 < 0$. A higher state tax raises the household's benefit from in-state municipals but not the insurer's, which is taxed federally through proration and receives no in-state exemption, so households crowd toward the municipals of high-tax states and $\text{Cov}(PCShare_s, Tax_s) < 0$. If the OLS event study omits $W_{b,t} := PostTCJA_t \times Tax_{s(b)} \times m_b$, then with $\tilde{X}_{b,t}$ and $\tilde{W}_{b,t}$ residualized on the

included controls and fixed effects,

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \theta_1 \cdot \frac{\text{Cov}(\tilde{X}_{b,t}, \tilde{W}_{b,t})}{\text{Var}(\tilde{X}_{b,t})}. \quad (119)$$

The fixed effects absorb the variation common to $PostTCJA_t$ and m_b , and under the maintained orthogonality approximation that maturity composition and timing do not vary systematically with $PCShare_s \times Tax_s$, the residualized covariance inherits the negative sign of the state covariance. With $\theta_1 < 0$ the bias term is positive,

$$\text{plim } \tilde{\beta}_1 > \beta_1, \quad (120)$$

so OLS attenuates the insurer slope. IV removes the bias when the domicile instruments are relevant, affect pricing only through insurer share, and are orthogonal in their instrumented component to the SALT channel, which is why IV estimates can exceed OLS in magnitude. The implication is testable. The post-TCJA slope change should be more negative in high-tax states conditional on insurer exposure, and controlling for $W_{b,t}$ should move the OLS estimate toward IV.

A.6.9 The bank-qualified discontinuity: construction, robustness, and placebos

A.6.9.1 Derivation of the designation premium

Consider matched bonds with the same supply $q_k > 0$, outside opportunities and non-tax characteristics fixed, household demand $d_k^H(y)$ at yield y , and bank demand $d_k^B(y + u)$ when designation adds an annual benefit u , with funding costs, taxes and portfolio-risk compensation inside the schedules. Assume a clearing solution exists for every $u \in [0, X]$, the schedules are continuous and nondecreasing with finitely many continuously differentiable regimes, and total demand has strictly positive slope at clearing:

$$d_k^H(y_k(u)) + d_k^B(y_k(u) + u) = q_k, \quad y_k^N = y_k(0), \quad y_k^Q = y_k(X). \quad (121)$$

A strictly concave quadratic portfolio objective conditional on the remaining portfolio delivers these properties, with demand $x_{ik} = \max\{0, (y + \mathbf{1}\{i = B\}u - r_{ik})/a_{ik}\}$ for net reservation return r_{ik} and curvature $a_{ik} > 0$; the zero-holding region permits changes in participation.

Differentiating (121) on a smooth segment gives, at the clearing net returns,

$$-\frac{dy_k(u)}{du} = \frac{\partial d_k^B / \partial y}{\partial d_k^H / \partial y + \partial d_k^B / \partial y} \equiv w_{Bk}(u) \in [0, 1], \quad (122)$$

and integrating across participation boundaries by continuity gives

$$\pi_k = y_k(0) - y_k(X) = \int_0^X w_{Bk}(u) du = \theta_k X, \quad \theta_k = \frac{1}{X} \int_0^X w_{Bk}(u) du. \quad (123)$$

With banks outside the municipal market throughout, $\partial d_k^B / \partial y = 0$ and $\pi_k = 0$, the household benchmark; with banks alone, $\partial d_k^H / \partial y = 0$ and $\pi_k = X$, the bank-only benchmark; with both responding the premium lies between. The capitalization share θ_k is the part of the bank-only benefit reflected in the yield, a quantity distinct from the fraction of bonds banks hold. Household participation at short maturities and stronger bank participation farther out give the increasing profile tested in the paper. The comparison is local to matched bonds rather than a re-solution of every outside market. The flat- X prediction is in annual yield-equivalent units under the carry comparison, and the per-dollar-of-basis benefit maps to an offering-yield difference through comparable cash flows and basis relative to price.

A.6.9.2 Scaling by the expected benefit

The priced share depends on the benefit chosen. The issue year's deduction $X_t = 0.8 \tau_{B,t} C_t$ falls from 104 basis points in 2000 to about 9 over 2013–17, while the premium does not. In year-level terciles of X_t , the 5–20 year premium is 20.8 to 21.1 basis points where the year's deduction averages 8.5, 12.3 to 23.2 where it averages 26.6, and 6.5 to 13.2 where it averages 70.1, with the long-minus-short contrast significant in each. Bond-by-bond ratios to the current-year deduction exceed two in the near-zero-rate years and fall below a quarter in the high-benefit years, 2000–02 and 2005–08, so a pooled ratio is whatever weighting those years receive. The scaled estimates therefore use the benefit the bank expects over the bond's remaining life.

Construction. For a bond issued in month t with m years to run, $X_{tm} = 0.8 \tau_{B,t} \hat{c}(t, m)$, with $\hat{c}(t, m)$ the average funding cost the bank expects over the next $\min(m, 10)$ years, in basis points. First, the expected average short rate over the bond's life, r_{tm} , is the risk-neutral component of the m -year treasury yield in the issue month. This is the fitted yield less a term-structure model's term premium, at maturity $\min(m, 10)$ and held at its ten-year value beyond ten years, where the models stop. Second, it is mapped to bank funding costs by the annual relation between the municipal-securities-weighted cost of funds and the one-

year expected rate over 2000–2021, $c_t = a + b r_{t,1}$ with $a = 0.38$ and $b = 0.55$ ($R^2 = 0.75$, 22 years) under Kim–Wright. The mapping is linear, so applied to the average expected rate it averages expected funding costs year by year over $\min(m, 10)$ years, and its percentage-point fit is multiplied by 100. Third, the issue year’s deduction rate $0.8 \tau_{B,t}$ converts the funding cost into the annual benefit per dollar of basis. X_{tm} averages 33.9, 41.5, 50.5, 56.8 and 53.3 basis points across the five maturity bins, has median 45, and never falls below 8.

Removing the term premium. The raw m -year zero yield is the expected average short rate plus a term premium that rises with maturity. Over 2000–2021 the Kim–Wright estimates average 0.11, 0.22 and 0.51 percent at one, five and ten years, and the Adrian–Crump–Moench estimates are 0.02, 0.43 and 0.81. In the denominator it would give long bonds a larger benefit unrelated to expected funding costs and mechanically lower their shares. Because term premia are estimated, the benefit is built from both models, the Federal Reserve’s three-factor Kim–Wright and the five-factor Adrian–Crump–Moench. Over the 264 months of 2000–2021 their risk-neutral yields have correlations of 0.99 at one year, 0.90 at five and 0.78 at ten, and mean absolute differences of 0.19, 0.56 and 0.63 percentage points. The gap widens with the horizon as the estimated premia diverge (Figure 28, panel a). Under Adrian–Crump–Moench the pass-through is $c_t = 0.41 + 0.49 r_{t,1}$ ($R^2 = 0.66$) and the benefits are 0.8–8.6 basis points lower across the bins.

Results under both measures. Dividing each bond’s yield by its X_{tm} before the 2SLS of equation (54) gives the shares of Table 7, 0.17, 0.31, 0.41, 0.52 and 0.47, contrast 0.34 (standard error 0.11); the Adrian–Crump–Moench benefit gives 0.17, 0.35, 0.46, 0.55 and 0.48, contrast 0.39 (Figure 28, panel b). Within rate regimes the contrast is 0.57 in the low-benefit years, 0.37 in the middle and 0.27 in the high-benefit years, where the expected benefit is nearly flat in maturity (62.8 to 71.2 basis points) and the curve was often inverted. The gradient is therefore not the maturity structure of the benefit. Without 2020–21, when bank demand for municipals was strongest, the shares are 0.18, 0.25, 0.35, 0.39 and 0.36, contrast 0.21.

Running variable. Issuer-year issuance sums par over every tax-exempt maturity in the Mergent issuance file sharing a six-character CUSIP issuer prefix and an offering calendar year, before any sample restriction. The statutory quantity aggregates a legal issuer with its subordinate entities, excludes refundings of designated bonds, which are deemed designated (Section 265(b)(3)(D)), excludes private-activity bonds other than 501(c)(3) bonds, and counts tax-exempt bank loans and leases, which mostly carry no CUSIP. The measured

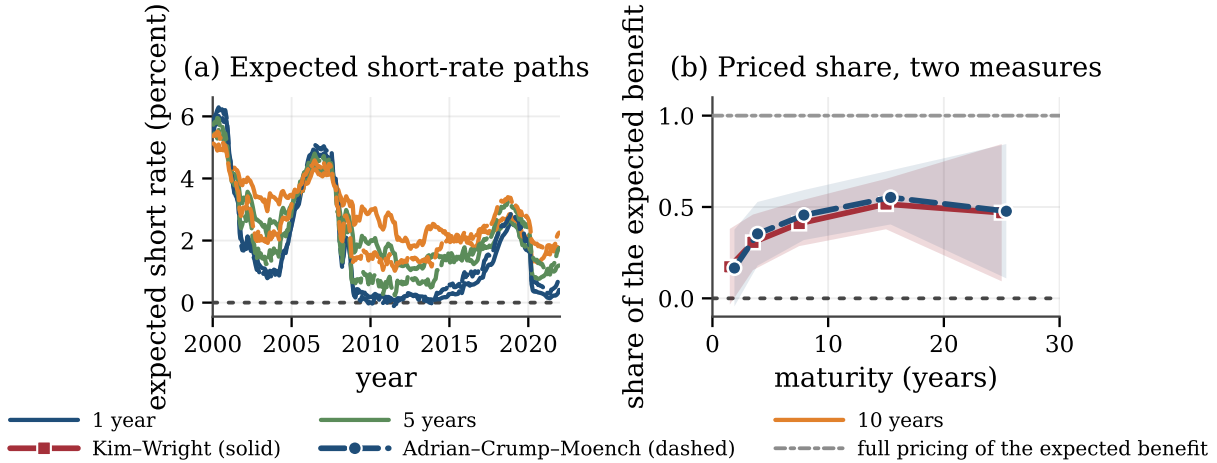


Figure 28: Two measures of the expected bank benefit. Panel (a) plots the risk-neutral treasury yields at one, five and ten years, the expected average short rate over each horizon, from the Kim–Wright model (solid) and the Adrian–Crump–Moench model (dashed), monthly averages, 2000–2021. Panel (b) reports the priced share of the bank benefit by maturity bin when the benefit is built from each model: 2SLS estimates of equation (54) with the matched fixed effects, issuer-clustered 95 percent bands, and the horizontal line at one marking full pricing of the expected benefit.

Alt text: The first panel shows two closely tracking sets of expected short-rate paths at three horizons over two decades. The second panel shows two priced-share profiles by maturity that rise from near zero at the shortest maturities to about one half from five years out, with overlapping uncertainty bands, both below the full-pricing line.

running variable is a noisy version of the statutory one, so measured eligibility instruments designation in a fuzzy discontinuity, with a first stage of 0.42 to 0.45. Just over the cap, 21% of bonds are designated. Issuer-years with a single sale above \$10.5 million designate 4% of the time, and issuer-years with several sales designate 45%. This is because the statute tests issuance reasonably anticipated at each sale, so a March designation stands when an unanticipated October sale takes the year over the cap. Refunding bonds are 46% of bonds just over the cap, and netting refunding par out of the running variable cuts over-cap designation to 9%. The designated share declines smoothly with distance above the cap, 24% within \$1 million and 7% beyond \$8 million. That pattern is the signature of measurement noise rather than of a rule that stops applying.

Balance at the cap. Local linear regressions of covariates on eligibility by bin, in the baseline window with the set-A fixed effects, show callability and the number of maturities in the issue balanced in every bin. Ratings, available for 61% of bonds, differ only at 1–2 years, by 0.17 of a notch ($t = 2.0$). Coupons are lower under the cap in bins 2–5 by 0.09 to 0.14 points; the coupon is set at pricing and partly an outcome, and estimates with and

without the coupon control agree. Log par, lower under the cap at 1–2 years (-0.26) and higher at 10–20 years ($+0.12$), is a control throughout.

The window. Table 33 varies the donut and the bandwidth. The half-million donut holds the issuers that manage annual issuance to stay under the cap, 2,081 issuer-years below against 572 above within \$0.25 million. The 1,034 issuer-years (12,076 bonds) exactly at the cap of their year are eligible under the statute, and 79.6% of their bonds are designated, the rate below the cap, against 27% immediately above. A strict inequality would place them on the wrong side and lower the no-donut first stage from 0.39 to 0.30 while barely moving the 2SLS estimates, since dividing by the first stage undoes the attenuation; the text uses $E = \mathbf{1}\{R \leq 0\}$ throughout. Across donuts from none to \$1.5 million and bandwidths from \$2 to \$8 million the 10–20 minus 1–2 contrast lies between 13.5 and 18.4 basis points and the 1–2 year premium never departs from zero. The window of [Dagostino \(2025\)](#)’s spread comparison, issuance of \$5 to \$15 million with no exclusion, gives the most precise contrast in the table. Dropping the region her bunching estimation treats as behaviorally contaminated, issuance of \$8.5 to \$13.5 million, compares issuers at least \$1.5 million below the cap with issuers at least \$3.5 million above it. The first stage rises to 0.47 to 0.54, and the premium attenuates at 5–20 years because the local linear function extrapolates to the cap from further away. The contrast is 9.8 (4.2).

Table 33: Bank-qualified discontinuity: sensitivity to the donut, the bandwidth, and the window.

	first stage	1–2y	2–5y	5–10y	10–20y	20y+	10–20 minus 1–2
<i>Donut half-width, bandwidth \$4 million</i>							
none	0.39	1.9 (1.8)	7.5 (1.8)	13.9 (1.8)	17.7 (1.8)	17.7 (4.5)	15.7 (2.0)
\$0.25 million	0.40	1.6 (2.0)	7.6 (2.0)	13.4 (2.0)	16.7 (2.2)	16.3 (5.5)	15.1 (2.4)
\$0.5 million (baseline)	0.42	1.8 (2.2)	8.7 (2.2)	13.8 (2.2)	18.4 (2.3)	18.2 (6.3)	16.5 (2.7)
\$1 million	0.44	1.2 (2.9)	8.2 (2.7)	13.3 (2.8)	19.7 (3.1)	20.6 (8.0)	18.4 (3.5)
\$1.5 million	0.44	2.0 (4.4)	7.3 (3.8)	11.8 (4.0)	15.5 (4.2)	19.5 (10.2)	13.5 (5.2)
<i>Bandwidth, donut \$0.5 million</i>							
\$2 million	0.38	0.8 (4.7)	7.9 (4.9)	13.3 (4.8)	16.9 (5.0)	21.0 (14.1)	16.1 (4.8)
\$3 million	0.40	2.7 (3.0)	8.9 (3.0)	15.2 (2.9)	19.2 (3.1)	19.9 (7.6)	16.5 (3.3)
\$5 million	0.42	−0.9 (1.9)	5.7 (1.8)	11.4 (1.8)	14.9 (1.9)	15.9 (4.9)	15.8 (2.3)
\$6 million	0.43	−0.5 (1.7)	5.4 (1.5)	10.8 (1.6)	13.2 (1.6)	11.4 (4.1)	13.6 (2.1)
\$8 million	0.45	0.0 (1.5)	6.3 (1.2)	10.1 (1.3)	13.5 (1.3)	12.3 (3.5)	13.5 (1.8)
<i>Other windows</i>							
\$5–15 million, no donut	0.40	0.1 (1.6)	5.6 (1.5)	12.1 (1.6)	15.5 (1.5)	16.7 (3.8)	15.4 (1.8)
\$8.5–13.5 million dropped, $ R \leq 8$	0.47	0.2 (3.5)	6.0 (3.0)	7.6 (3.1)	10.0 (3.4)	−1.8 (9.5)	9.8 (4.2)

Notes. 2SLS premium of designation in basis points by maturity bin, pooled 2000–2021 at the cap of the year, matched fixed effects, issuer-clustered standard errors in parentheses; the first-stage column is the 1–2 year bin. The \$5–15 million window is the one used in the spread comparison of [Dagostino \(2025\)](#); the last row drops the region her bunching estimation excludes.

Construction, and the relation to [Dagostino \(2025\)](#). That paper builds the same issuer-year issuance from Ipreo MuniIC, finds the same bunching under the cap and the same first stage at \$30 million in 2009–10, and compares tax-adjusted spreads over maturity-matched treasuries by \$1 million issuance bin for general-obligation bonds of 2000–2008. Its specification uses size, maturity, insurance, rating and use-of-proceeds controls, year and issuer fixed effects and year clustering, and it finds a jump of about 10 basis points at the cap. Its Call Report evidence, 4 to 5% of bank assets in designated bonds and under half a percent in non-designated ones through 2010, is the regime in which banks hold only designated bonds. Table [34](#) re-estimates the baseline under each of its construction choices and under a running variable net of refunding par. Every row gives nothing at 1–2 years and a contrast of 13.5 to 22.3 basis points. Within issuer the profile is steeper, with a small reversed sign at the short end, so identification from issuers that cross the cap over time gives the claim in its sharpest form. The refunding-net running variable, closest to the statute, shifts the window to a different set of issuer-years, which lowers the 5–10 year point and leaves the 10–20 year premium within one standard error of the baseline.

Table 34: Bank-qualified discontinuity: construction checks.

	1–2y	2–5y	5–10y	10–20y	20y+	10–20 minus 1–2
Baseline (Table 7)	1.8 (2.2)	8.7 (2.2)	13.8 (2.2)	18.4 (2.3)	18.2 (6.3)	16.5 (2.7)
Bond insurance and enhancement $\times$ year	0.9 (2.2)	8.0 (2.2)	13.0 (2.2)	17.7 (2.3)	17.3 (6.2)	16.8 (2.6)
General-obligation bonds only	1.7 (2.3)	7.6 (2.2)	13.3 (2.3)	15.2 (2.5)	15.8 (7.3)	13.5 (2.9)
Issuer fixed effects	−5.0 (2.6)	1.8 (2.5)	7.5 (2.6)	17.3 (2.8)	16.8 (4.1)	22.3 (2.4)
Tax-adjusted spread over treasuries	5.3 (3.3)	12.6 (3.2)	21.4 (3.2)	26.9 (3.5)	22.8 (8.6)	21.6 (3.8)
Clustered by year (22 clusters)	0.9 (2.9)	8.0 (2.6)	13.0 (3.6)	17.7 (3.4)	17.3 (7.6)	16.8 (3.7)
Running variable net of refundings	0.8 (3.1)	4.9 (3.1)	8.1 (3.1)	15.7 (3.6)	11.0 (8.2)	14.8 (3.7)
and refunding bonds excluded	0.9 (3.1)	6.3 (3.2)	8.9 (3.2)	16.0 (3.6)	10.9 (7.5)	15.1 (3.6)

Notes. 2SLS premium of designation in basis points, pooled 2000–2021 at the cap of the year, window \$0.5 to \$4 million, matched fixed effects, issuer-clustered standard errors unless stated. Insurance and enhancement: bond insurance $\times$ year and credit enhancement $\times$ year added (47% of window bonds are insured). General obligation: unlimited-tax, limited and double-barreled general-obligation bonds, 70% of the window. Issuer fixed effects: identification from issuers that cross the cap across years. Tax-adjusted spread: $100[y/(1 - \tau_{B,t}) - y^{\text{Tsy}}]$, the outcome of [Dagostino \(2025\)](#). Refunding-net: the running variable excludes refunding par; the last row also drops refunding bonds from the estimation sample. First stages are 0.40 to 0.51 in every bin of every row.

Figure 29 runs her estimator by maturity. The outcome is \$1 million issuance-bin dummies over her window, the last eligible bin as reference, with maturity bucket $\times$ year, state $\times$ year, rating $\times$ year and use-of-proceeds controls, issuer-clustered. The bins below the cap are flat and the bins above step up, her Figure 3 shape. Pooled across maturities the step is 4.2 basis points in the offering yield and 5.6 in the tax-adjusted spread. Split by maturity, it is 1.1 and 1.4 at 1–5 years and 7.8 and 10.2 at 10–20 years. In her own setting, general-obligation bonds of 2000–2008 with all maturities pooled and the tax-adjusted spread as outcome, the reduced-form jump at the cap is 9.7 basis points (standard error 3.2). This uses size, maturity, rating, use of proceeds, insurance and state fixed effects, issuer-clustered, against her approximately 10. With her issuer fixed effects and clustering by year, nine clusters, it is 3.1 (4.7), because pre-2009 within-issuer variation across the cap is thin.

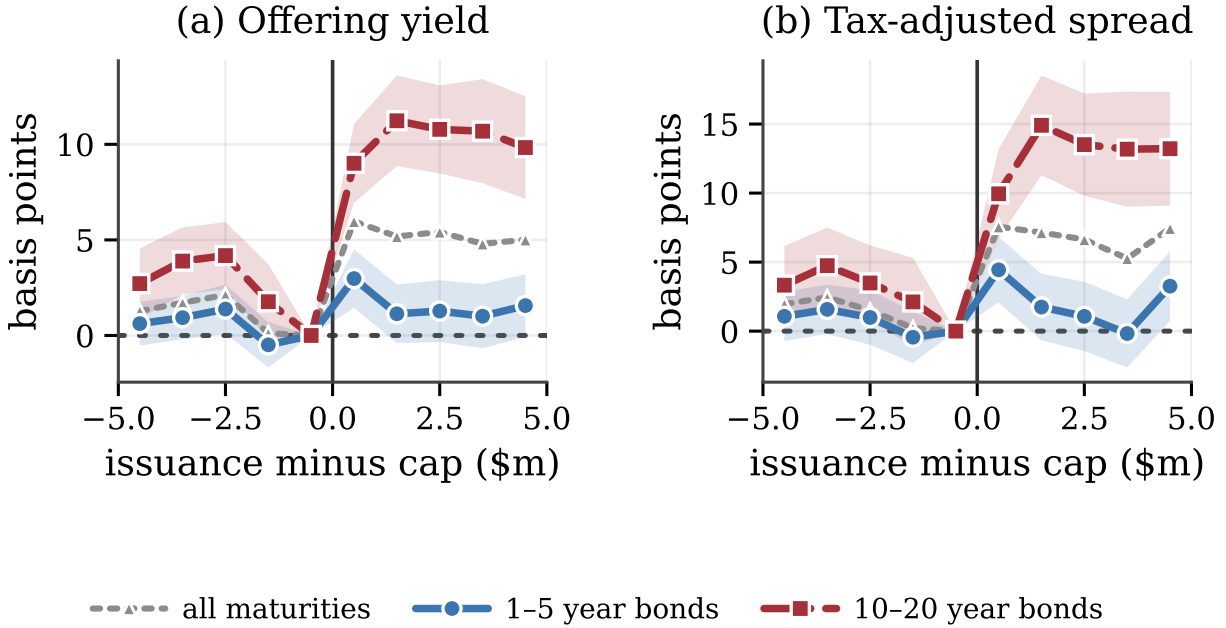


Figure 29: The binned estimator of [Dagostino \(2025\)](#) in this sample, by maturity. Coefficients on \$1 million issuance-bin dummies relative to the last eligible bin, for all maturities, for 1–5 year bonds and for 10–20 year bonds, with maturity bucket $\times$ year, state $\times$ year, rating $\times$ year and use-of-proceeds fixed effects; issuer-clustered 95 percent intervals. Panel (a) uses the offering yield, panel (b) the tax-adjusted spread over maturity-matched treasuries.

Alt text: Binned outcomes are plotted against issuer-year issuance around the qualification threshold. The panels compare full, short-, and long-maturity samples for offering yields and tax-adjusted spreads.

Placebo cutoffs and the ARRA cap. At a \$5 million placebo cutoff the first stage is 0.00 to 0.05 across bins, and the yield effect is within two basis points of zero in every bin. At \$15 million the first stage is -0.01 to 0.02 , and the yield effect is zero except in one 20+ year cell. At \$20 million a residual first stage of 0.04 to 0.10 , one fifth of the true one, comes with a yield effect of 2 to 4 basis points at 5–20 years, one quarter of the true one. This happens because the CUSIP prefix does not always match the statutory aggregation of related issuers. The cap moved with the statute. At \$30 million the first stage is 0.20 , 0.22 , 0.25 , 0.18 and 0.29 across the five bins in 2009–10, 0.01 to 0.05 in 2004–08 and -0.09 to 0.02 in 2011–15. At \$10 million it is 0.00 to 0.01 in bins 1–4 during 2009–10. In this auxiliary specification, with maturity-bin-by-year and state-by-year effects, the reduced-form yield discontinuities at \$30 million in 2009–10 are estimated on 16 thousand crisis-year bonds of issuers between \$22 and \$38 million. They are 6 to 10 basis points at 1–10 years, with t statistics below 1.5 in absolute value, approximately zero beyond ten years, and without a systematic maturity pattern.

By period. Figure 30 draws the premium by maturity for 2000–04, 2005–08, 2009–10, 2011–13, 2014–17 and 2018–19. Before 2009, with bank funding costs of 2.3 to 2.4% and X_t of 65 to 66 basis points, the point estimates are about -0.1 to 0.3 of X_t across bins. The 2009–10 panel is 2SLS at the \$30 million ARRA cap with the full matched fixed effects and a \$0.5 to \$8 million window. Its first stage is 0.16 to 0.26 , against 0.42 to 0.45 at \$10 million, because the temporary cap was less used, and its Kleibergen–Paap F is 1.5. With funding costs of 1.09% and an ordinary-rule X_t of 30 basis points, the premium is 11 (24), 22 (23), 17 (23), 50 (28) and 60 (29) basis points, rising with maturity as in every other period. The long-minus-short difference is 43 (20), significant at five per cent. In 2011–17, with period-average funding costs of 0.31 to 0.51% and X_t of 9 to 14 basis points, the premium is 2.4 to 2.5 basis points at 1–2 years and 13.5 to 28.4 at 5–20 years, at or above X_t . After the 2018 rate cut it is 6.9 (4.3), 6.4 (5.0), 7.7 (5.7), 14.9 (7.7) and 3.7 (24.2). The difference-in-discontinuities design, the post-2018 interaction of these profiles with designation on 2014–2019, gives 3.1 (6.5), -9.1 (6.7), -19.2 (7.6), 3.6 (10.0) and -1.3 (28.7) basis points, with first stages unchanged. Annual pooled reduced forms are largest in 2011–13, 16 to 22 basis points, and near zero in 2004–05 and 2017. This is more consistent with changes in bank participation in this segment than with the formula benefit, which was largest when funding costs were highest. Subchapter S banks, taxed at shareholder rates and so less affected by TCJA, hold 6 to 14% of bank municipal holdings over the sample and 11 to 14% from 2007 onward.

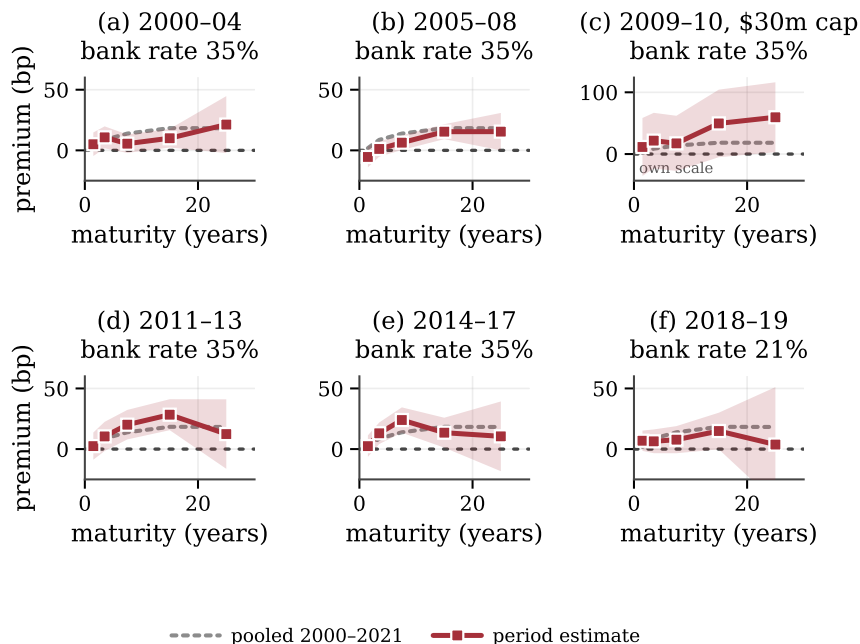


Figure 30: The bank-qualified premium by maturity and period, 2SLS in basis points with the matched fixed effects, one panel per period, with issuer-clustered 95 percent bands; the dotted line is the pooled 2000–2021 estimate of Table 7. Panel (c) is estimated at the \$30 million cap in force in 2009–10 under the ARRA, on a \$0.5 to \$8 million window, and is drawn on its own scale because its bands are several times wider; every other panel uses the \$10 million cap and the \$0.5 to \$4 million window. The bank rate was 35% through 2017 and 21% from 2018.

Alt text: Bank-qualified premium profiles are compared across issuance periods. The temporary higher qualification cap in 2009–2010 is treated separately, and the figure uses the indicated panel scales.

A.6.9.3 Call Report items and the funding-cost measure

The annual series uses December reports from FDIC BankFind for 2000–2021. Interest expense is year-to-date item EINTEXP (Schedule RI, item 2.e, RIAD4073, for commercial-bank Call Report filers), including interest on deposits and other borrowing. The denominator is the average-total-assets series ASSET5 instead of year-end assets or interest-bearing deposits. The ratio is thus interest expense per dollar of assets, which differs from both a deposit rate and the cost of maturity-matched debt. The weights are municipal securities, item SCMUNI (Schedule RC-B, item 3, columns A and D, 8496 and 8499 with the reporting prefix), held-to-maturity at amortized cost plus available-for-sale at fair value. These are all municipal securities, bank-qualified or not, and exclude municipal loans. No bank-size cutoff is imposed.⁴

⁴Source definitions: the FDIC’s [Financial API definitions](#) and the FFIEC [December 2017 Call Report](#), Schedules RI and RC-B; BankFind harmonizes the historical series, and the schedule references

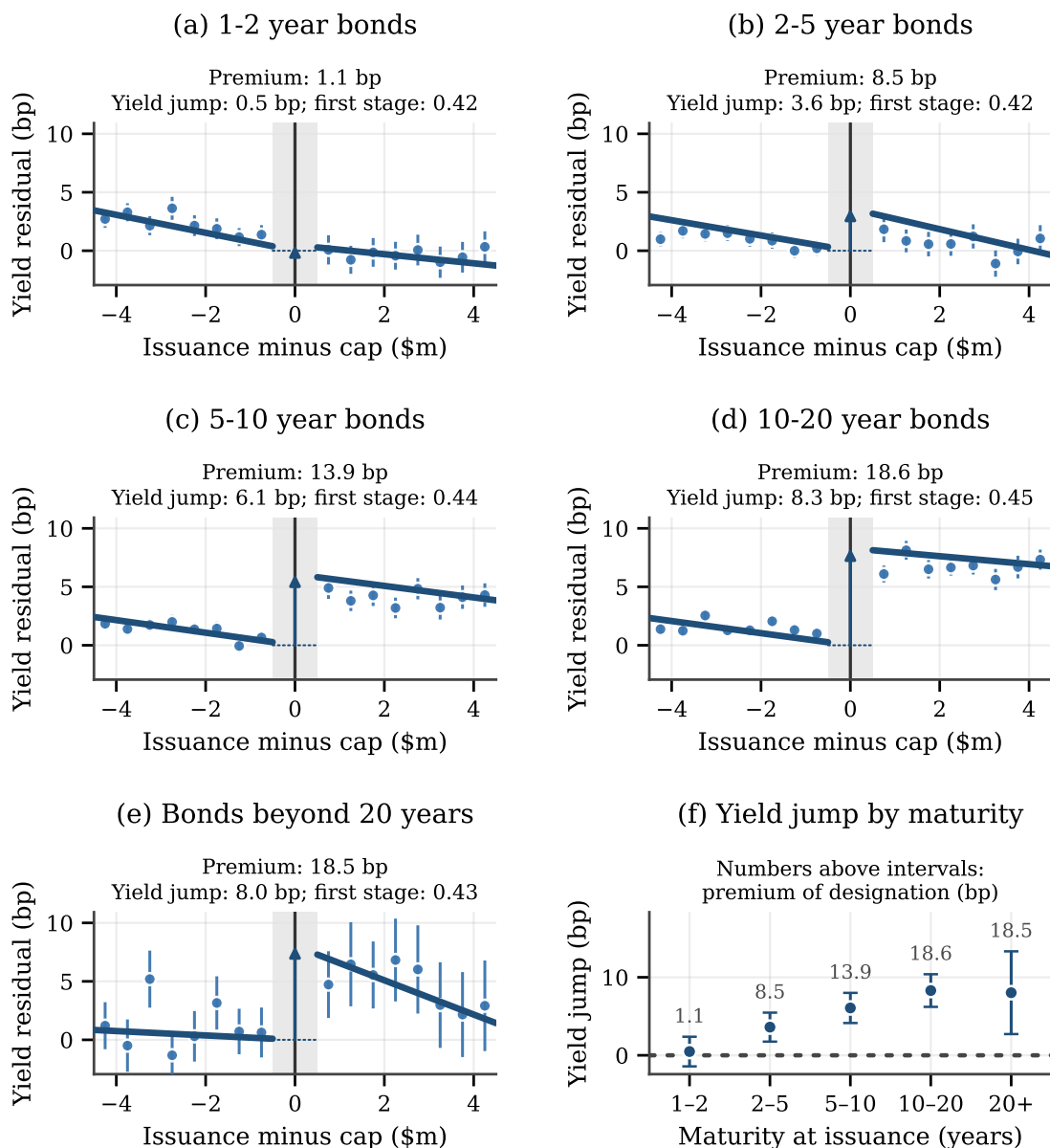


Figure 31: The discontinuity in the offering yield, by maturity bin at issuance. Panels (a)–(e) show residualized offering yields in \$0.5 million issuance bins, relative to the eligible side’s fitted value at the cap. Lines are the estimated local linear functions, the shaded region is the donut, and bars are 95 percent intervals. Arrows mark the ineligible-minus-eligible yield jump. Each panel’s header reports the designation premium, calculated as this jump divided by its first stage. Panel (f) collects the five jumps and intervals, with the corresponding premia above them. These are the bin-by-bin counterparts to the joint estimates in Table 7. Panels (a)–(e) share a scale based on the central 96 percent of binned means; a few sparse cells lie outside it.

Alt text: Offering-yield discontinuities are displayed separately by maturity bin. The accompanying estimates distinguish the reduced-form yield jump from the premium scaled by the qualification first stage.

With M_{bt} these holdings and $\mathcal{B}_t$ the retained banks with positive holdings,

$$c_t = \frac{\sum_{b \in \mathcal{B}_t} M_{bt} (I_{bt} / \bar{A}_{bt})}{\sum_{b \in \mathcal{B}_t} M_{bt}}, \quad X_t = 0.8 \tau_{B,t} c_t. \quad (124)$$

Retained banks have positive average assets, nonmissing interest expense, and an expense-to-assets ratio at least zero and below 15%. The period averages of c_t are 2.34% in 2000–04, 2.31% in 2005–08, 0.51% in 2011–13, 0.31% in 2014–17, and 0.66% in 2018–19; with the federal corporate rate the annual benefits are approximately 66, 65, 14, 9, and 11 basis points.

The statute allocates by average adjusted tax bases and otherwise-deductible interest expense, which the Call Report does not identify, and the IRS permits book-asset averaging conventions. X_t is therefore an observable ordinary-rule C-corporation benchmark instead of any bank’s tax-return calculation, the same across maturity bins.⁵ Subchapter S banks face shareholder-level taxation but remain in the weights. In 2009–10 the two-percent-of-assets exception of Section 265(b)(7) could also reduce the disallowance on non-designated bonds and so the value of designation; the \$30 million cap alone does not capture it, and X_t is the ordinary-rule scale for those years.

Bank holdings of municipal securities rose from \$97 billion in 2000 to \$370 billion in 2017 and fell after the 2018 rate cut. Designation, refunding status, security type, bond insurance, and credit enhancement come from the Mergent bond master files.

A.6.10 P3: muni-yield response to treasury yield, construction details

Transaction sample. The sample is the cleaned MSRB customer-side transaction file, restricted to federally tax-exempt, non-callable issues with a mean rating of AA and above (agency score ≤ 3), the cut of the other transaction-level tests; non-callability keeps call-option value out of the maturity gradient. Trades less than 92 days after issue, where secondary-market yields are anchored to the offering yield, and less than 365 days before maturity, where the treasury short-rate substitute behaves erratically, are dropped.

Maturity-exact treasury interpolation. Let $d_{b,t}$ be bond b ’s remaining days to maturity on trade date t and $m_b(t) = d_{b,t}/365$. The daily zero-coupon curve of [Gürkaynak et al. \(2007\)](#)

identify the corresponding commercial-bank items.

⁵See the IRS, [Bank Qualified Bonds: Section 265](#), pp. 13-11–13-12, for the allocation and 20% disallowance; the annual per-dollar-of-basis benefit is distinct from the present value of deductions over a bond’s remaining life.

Table 35: P3: muni yield response to treasury yield by remaining maturity.

	(1) Continuous $\Delta y_{b,t}^{\text{muni}}$	(2) Bin $\Delta y_{b,t}^{\text{muni}}$
$\Delta y_{m_b,t}^{\text{Tsy}}$	0.346*** (0.016)	0.351*** (0.013)
$\Delta y_{m_b,t}^{\text{Tsy}} \times m_b$	0.007*** (0.002)	
$\Delta y_{m_b,t}^{\text{Tsy}} \times Long_b$		0.075*** (0.012)
$Long_b$		0.016*** (0.002)
m_b	0.002*** (0.000)	
Observations	952,874	952,874
Weighting	Candidate-pair	Candidate-pair
SE cluster	Trade date	Trade date

Notes. Sample: MSRB transactions in federally tax-exempt, non-callable bonds rated AA and above (agency score ≤ 3); consecutive trades within 90 days on which the maturity-matched treasury moved at least 10 basis points in absolute value. Observations are weighted by the inverse number of candidate pairs per CUSIP passing the date-gap and rate-move filters, counted before excluding missing or trimmed outcomes, so surviving total weights need not be equal across bonds; the dependent variable is trimmed below the first and above the ninety-ninth percentiles. The treasury yield is interpolated on the daily GSW curve at the bond's exact remaining maturity on each trade date. $Long_b = \mathbf{1}\{m_b \geq 5\}$. The first-difference design absorbs time-invariant bond characteristics. Standard errors are clustered by trade date (3,864 clusters); Table 38 reports state clustering. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

is observed at integer maturities $j \in \{1, 2, \dots, 30\}$; the matched yield interpolates linearly between the bracketing integers,

$$y_{m_b(t),t}^{\text{Tsy}} = (1 - w) y_{\lfloor m_b(t) \rfloor, t}^{\text{Tsy}} + w y_{\lceil m_b(t) \rceil, t}^{\text{Tsy}},$$

$$w = \begin{cases} 0, & \lfloor m_b(t) \rfloor = \lceil m_b(t) \rceil, \\ \frac{m_b(t) - \lfloor m_b(t) \rfloor}{\lceil m_b(t) \rceil - \lfloor m_b(t) \rfloor}, & \text{otherwise.} \end{cases}$$

Remaining maturities outside the grid are clamped to its nearest endpoint.

First-difference construction. Within each (CUSIP, dealer-side) cell, consecutive transactions give

$$\begin{aligned}\Delta y_{b,t}^{\text{muni}} &= y_{b,t}^{\text{muni}} - y_{b,t-1}^{\text{muni}}, \\ \Delta y_{m_b,t}^{\text{Tsy}} &= y_{m_b(t),t}^{\text{Tsy}} - y_{m_b(t-1),t-1}^{\text{Tsy}}, \\ \Delta t &= \text{elapsed days between the two trades.}\end{aligned}$$

Stratifying by dealer side compares purchases with purchases and sales with sales, limiting bid-ask, markup and intraday-price contamination. I keep pairs with $0 < \Delta t \leq 90$ days and $|\Delta y_{m_b,t}^{\text{Tsy}}| \geq 10$ basis points, weight by the inverse candidate-pair count per CUSIP, and trim the dependent variable below the first and above the ninety-ninth percentiles.

Specifications. Column (1) of Table 35 estimates equation (55), the counterpart of the slope-amplification prediction of Proposition 3, without fixed effects. Column (2) estimates $\Delta y_{b,t}^{\text{muni}} = \beta_0 \Delta y_{m_b,t}^{\text{Tsy}} + \beta_L \text{Long}_b + \beta_1 (\Delta y_{m_b,t}^{\text{Tsy}} \times \text{Long}_b) + \varepsilon_{b,t}$ with $\text{Long}_b = \mathbf{1}\{m_b \geq 5\}$ and no continuous maturity main effect; the short response is β_0 and the long response $\beta_0 + \beta_1$.

Holding the benchmark maturity fixed. Because the matched maturity falls as the bond ages, the benchmark difference contains a roll-down term, $-(\Delta t/365) \partial_m y_{m,t}^{\text{Tsy}} + O((\Delta t/365)^2)$. The daily curve lets one measure it directly. Holding the maturity at the first trade gives the roll-down-free regressor $\Delta y_{b,t}^{\text{Tsy, fixed}} = y_{m_b(t-1),t}^{\text{Tsy}} - y_{m_b(t-1),t-1}^{\text{Tsy}}$; its difference from the reported regressor is the slide, $y_{m_b(t),t}^{\text{Tsy}} - y_{m_b(t-1),t}^{\text{Tsy}}$. On an upward-sloping curve the slide is negative. It lowers the measured benchmark change by 2.6 basis points on average in the maturity bins below five years and by 1.6 in those above, against a mean absolute move of about 26 basis points. It does not cancel against the muni leg’s own ageing, because the two curves differ in slope. Removing it (Table 36) lowers the response in both bins, by more in the short bin, so the gradient rises. Controlling the elapsed gap and its maturity interaction raises it again; the control absorbs the outcome’s own ageing to first order, since a bond is never observed twice at the same remaining maturity. The headline specification is the conservative one.

Sensitivity to the filter thresholds. Table 37 re-estimates the maturity interaction across cutoff and window. The point estimates are positive throughout; every binned cell is significant at one per cent, the continuous interaction at five per cent in every cell and at one per cent in twenty-seven of the thirty.

Table 36: P3 robustness: the treasury benchmark’s roll-down. Row one is the specification of Table 35, with the benchmark matched at each trade’s remaining maturity; row two holds the benchmark at the first trade’s remaining maturity; row three adds the elapsed gap and its maturity interaction as controls. All rows use the 952,874 pairs, inverse candidate-pair weights and trade-date clustering of Table 35, estimated by the continuous specification of equation (55) and its binned counterpart.

Treasury benchmark	Bin specification		Interaction	
	< 5y	≥ 5y	Bin	Continuous
Benchmark matched at each trade’s maturity (Table 35)	0.351	0.426	0.0749	0.0070
Benchmark held at the first trade’s maturity	0.340	0.423	0.0825	0.0084
and controlling the gap and gap × maturity	0.334	0.419	0.0850	0.0086

Table 37: P3 sensitivity: maturity-interaction coefficient across the treasury-move cutoff and trade-gap window (inverse candidate-pair weights).

<i>Panel A. Continuous interaction $\Delta y_{m_b,t}^{\text{Tsy}} \times m_b$</i>					
Cutoff \ window	15d	30d	60d	90d	120d
≥ 0.05	0.0053***	0.0057***	0.0051**	0.0071***	0.0090***
≥ 0.075	0.0061***	0.0057***	0.0047**	0.0068***	0.0087***
≥ 0.10	0.0065***	0.0060***	0.0049**	0.0070***	0.0087***
≥ 0.15	0.0080***	0.0071***	0.0059***	0.0077***	0.0091***
≥ 0.20	0.0088***	0.0082***	0.0067***	0.0081***	0.0094***
≥ 0.25	0.0104***	0.0083***	0.0074***	0.0087***	0.0097***
<i>Panel B. Binned interaction $\Delta y_{m_b,t}^{\text{Tsy}} \times Long_b$</i>					
Cutoff \ window	15d	30d	60d	90d	120d
≥ 0.05	0.049***	0.061***	0.070***	0.074***	0.088***
≥ 0.075	0.056***	0.063***	0.068***	0.074***	0.086***
≥ 0.10	0.062***	0.066***	0.069***	0.075***	0.086***
≥ 0.15	0.079***	0.077***	0.079***	0.083***	0.092***
≥ 0.20	0.094***	0.090***	0.085***	0.088***	0.097***
≥ 0.25	0.111***	0.100***	0.096***	0.096***	0.102***

Notes. Each cell re-estimates the headline specification on the pairs whose maturity-matched treasury move is at least the cutoff (in percentage points) and whose trades are at most the window apart, weighted by inverse candidate-pair counts within CUSIP, with standard errors clustered by trade date as in Table 35. Under state clustering (54 clusters) the point estimates are unchanged and the significance level changes only in Panel A: the continuous interaction falls from one to five per cent at the thirty-day window for cutoffs 0.05, 0.075 and 0.10 and at the sixty-day window for cutoff 0.15, and from five to ten per cent at the sixty-day window for cutoffs 0.05, 0.075 and 0.10. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Clustering at the state. Table 38 repeats Table 35 with standard errors clustered at the state (54 clusters); both interactions remain significant at one per cent. In the grid, state

clustering changes the level only in the Panel A cells listed in the note to Table 37, and the continuous interaction stays significant at five per cent in twenty-seven of the thirty.

Table 38: P3 with standard errors clustered at the state.

	(1) Continuous $\Delta y_{b,t}^{\text{muni}}$	(2) Bin $\Delta y_{b,t}^{\text{muni}}$
$\Delta y_{m_b,t}^{\text{Tsy}}$	0.346*** (0.012)	0.351*** (0.006)
$\Delta y_{m_b,t}^{\text{Tsy}} \times m_b$	0.007*** (0.002)	
$\Delta y_{m_b,t}^{\text{Tsy}} \times Long_b$		0.075*** (0.009)
$Long_b$		0.016*** (0.002)
m_b	0.002*** (0.000)	
Observations	952,874	952,874
Weighting	Candidate-pair	Candidate-pair
SE cluster	State	State

Notes. The two columns of Table 35, on the same sample, weights and trimming, with standard errors clustered at the state (54 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.7 Robustness to the rating cut

The two rating cuts are defined on the average agency score on the twenty-two-notch scale (1 = AAA, 3 = AA, 7 = A–). The curve construction of §3 and §6 uses a score of seven or better, A– and above, the credit band of the A-or-above corporate curve it is differenced against; the transaction-level tests of §7 use three or better, AA and above.

A.7.1 Curve construction at AA and above

Restricting the municipal leg to AA and above lowers its credit exposure relative to the A-or-above corporate benchmark, so persistence of the maturity pattern is a check against an explanation based on lower-rated municipal bonds. The decline is smaller but does not go away: between two and thirty years the observed implied tax falls 23.4 percentage points at score ≤ 7 and 21.6 at score ≤ 3 , and 16.0 and 13.5 points after the call and issuance-premium

adjustments. The level is uniformly higher on the AA sample, and the model plus the five adjustments still tracks the observed curve (Figures 32 and 33).

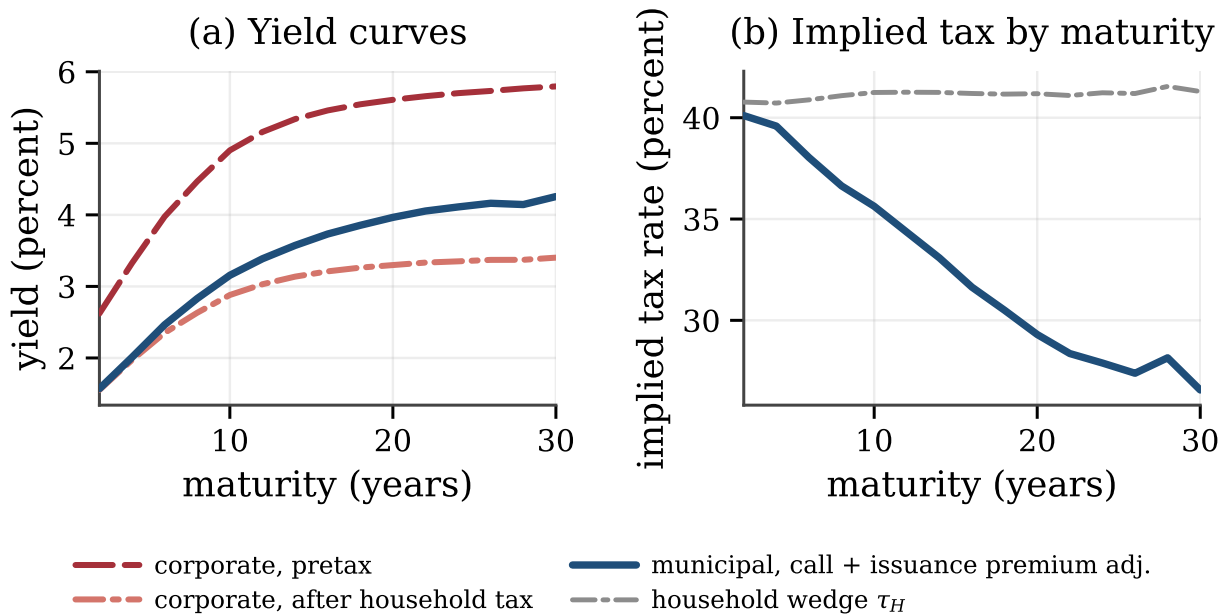


Figure 32: The muni puzzle at score ≤ 3 (AA and above), constructed as Figure 2, which uses score ≤ 7 . Panel A: the pretax corporate curve, the after-household-tax benchmark $\xi_H y_T^C$, and the municipal curve with the embedded call and the estimated issuance premium removed. Panel B: the corresponding implied tax, after both curve-level adjustments, against the household wedge. The adjusted implied tax falls from 40.1% at two years to 26.6% at thirty, against 38.7% to 22.7% at score ≤ 7 .

Alt text: For the AA-and-above sample, municipal yields increasingly exceed the household after-tax corporate benchmark with maturity. The companion panel shows the corresponding declining implied-tax profile.

A.7.2 Empirical tests at A– and above

Tables 39–41 re-estimate the household-tax, insurer-TCJA and rate-response tests at both cuts, as Table 8 does for the BAB test, with the dependent variable, controls, fixed effects and clustering held fixed within each comparison.

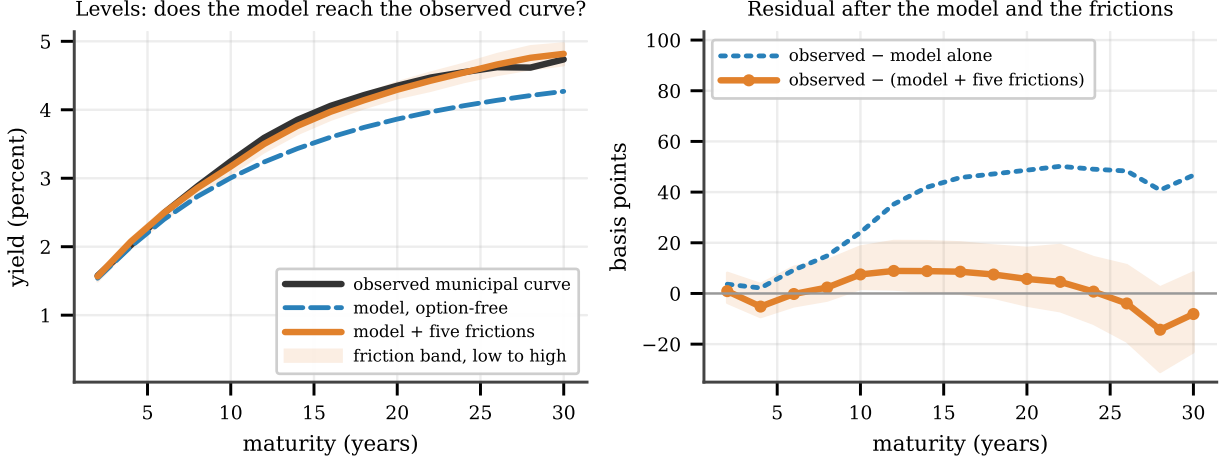


Figure 33: Model fit and residual at score ≤ 3 (AA and above), the counterpart of Figure 23. Left: the observed municipal curve, the model's option-free curve, and the model with the five adjustments added, the band spanning the low and high adjustment estimates. Right: the residuals. As at score ≤ 7 , the model plus the five adjustments tracks the observed curve and the residual is flat within about ten basis points across almost all of the curve.

Alt text: For the AA-and-above sample, the figure compares observed yields with the calibrated model and the model plus bond-side adjustments. The residual panel shows the remaining difference and adjustment range.

Table 39: P1: state-tax pass-through by maturity, by rating cut

	OLS		IV	
	score ≤ 3	score ≤ 7	score ≤ 3	score ≤ 7
β_1 : StateTax $\times$ [0,5)y	0.9866*** (0.1428)	0.8518*** (0.1952)	0.9757*** (0.1941)	0.8948*** (0.2627)
β_2 : StateTax $\times$ [5,10)y	0.5639*** (0.0901)	0.4792*** (0.1370)	0.5601*** (0.1146)	0.5011*** (0.1710)
β_3 : StateTax $\times$ [10,15)y	0.3468** (0.1643)	0.2146 (0.1770)	0.3224* (0.1703)	0.2112 (0.1820)
β_4 : StateTax $\times$ [15,20)y	0.2444 (0.1869)	0.1124 (0.1893)	0.2102 (0.1799)	0.1077 (0.1800)
β_5 : StateTax $\times$ ≥ 20 y	0.1592 (0.1419)	0.0449 (0.1397)	0.1260 (0.1251)	0.0410 (0.1336)
Observations	834,553	1,661,208	834,553	1,661,208

Notes. The de-called zero-coupon implied state tax, re-priced under the calibrated equilibrium, as in Table 6: unweighted, instrumented with the state's 1990 top rate interacted with the maturity bins, standard errors clustered at the state (47 clusters). The declining profile, which identifies the mechanism, is present on both samples, and the decline $\beta_1 - \beta_5$ is rejected at one per cent under both estimators on both; on the wider sample the profile falls further and its three longest brackets are individually indistinguishable from zero under either cut.

Table 40: Complementary insurer-TCJA test, by rating cut

	OLS		IV1		IV2	
	≤ 3	≤ 7	≤ 3	≤ 7	≤ 3	≤ 7
$Post \times PCShare$	-0.118*** (0.029)	-0.161*** (0.035)	-0.124 (0.085)	-0.255** (0.110)	-0.015 (0.102)	-0.175 (0.131)
$Post \times PCShare \times m_b$	-0.018*** (0.005)	-0.036*** (0.005)	-0.037*** (0.011)	-0.039*** (0.009)	-0.038** (0.015)	-0.037*** (0.012)
Observations	1,733,881	3,043,035	1,733,881	3,043,035	1,733,881	3,043,035

Notes. Standard errors clustered at the issuer (six-digit CUSIP), as in the event study of Figure 27; Appendix A.6.8 reports state-level inference. Both rating-cut columns omit the rating control, so the score ≤ 3 estimates are not the rating-controlled baseline of Table 32. The maturity gradient, the term the model predicts, is negative on both samples and significant under issuer clustering; in OLS it roughly doubles on the wider sample, under IV it is similar.

Table 41: P3: transaction-level treasury pass-through, by rating cut

	(1) continuous		(2) bin $<5y$ vs $\geq 5y$	
	score ≤ 3	score ≤ 7	score ≤ 3	score ≤ 7
$\Delta y_{m_b,t}^{Tsy}$	0.346*** (0.012)	0.281*** (0.011)		
$\Delta y_{m_b,t}^{Tsy} \times \text{maturity}$	0.007*** (0.002)	0.013*** (0.002)		
$\Delta y_{m_b,t}^{Tsy} \times (\geq 5y)$			0.075*** (0.009)	0.113*** (0.006)
Observations	952,874	2,023,637	952,874	2,023,637

Notes. First-differenced trade-to-trade specification of Table 35, weighted by inverse candidate-pair counts within CUSIP and clustered by state as in Table 38. The interaction is the maturity gradient in how municipal yields absorb treasury moves. It is positive and significant at one per cent on both samples and larger on the wider one: the continuous one roughly doubles and the binned one rises by about half.

A.8 The single-tax-rate benchmark of Green (1993)

Among single-tax-rate explanations, Green (1993) is the closest to this paper: it locates the problem in the benchmark, not the municipal bond.

Market discount is taxed at disposition; the gain is not recognized as it accrues over the holding period. A portfolio long two five-percent market-discount bonds and short one ten-percent par bond of the same maturity is therefore a synthetic discount bond taxed once, at maturity. Its value to an investor with rate τ is $P_T = (1 - \tau)d_T / (1 - \tau d_T)$, more than a fully taxed claim. Dealers and tax-exempt institutions arbitrage freely within the taxable

market, so par taxables are priced consistently with such packages, his equation (8), and inherit the advantage. The after-tax benchmark is then no longer the par taxable yield times $1 - \tau$. The taxable par curve is pulled down at the long end, where coupons carry more of the value, and the municipal-to-taxable yield ratio rises with maturity under one tax rate, entirely on the taxable side.

Deductions from borrowings and short sales cannot exceed the investment income they support. An individual therefore cannot short tax-disadvantaged against tax-advantaged taxables without bound and is not marginal on all bonds at once: she is marginal on municipals and on the tax-advantaged taxable portfolio, to which dealers, who face no such cap, tie par taxables. Unlike Fama (1977), one representative investor prices the municipal curve at every maturity.

Green's investor, pricing municipals against the synthetic package, requires a higher municipal yield than one holding them against fully taxed coupons. Households run no such strategy: deep market-discount taxables are scarce, original-issue discount is accrued by statute, and the short leg consumes the interest-deduction cap. They accept the traditional benchmark, $C_T(1 - \tau)$ in his notation and $\xi_H y_T^C$ here, and wherever they have capacity they set the price. Green's investor is inframarginal there, and restoring his relation would require the blocked short of the tax-exempt bond (Appendix A.9.4). His relation is a ceiling, not an equality that pins the curve. Nor can it explain the long end. His own 1987 to 1990 estimates put tax-exempt yields above even his benchmark, by roughly twenty basis points of taxable-equivalent yield at thirty years and growing in maturity. The marginal holder there values the exemption less than his strategy investor, as the participation mechanism implies once shorter maturities absorb the higher-wedge clienteles' capacity.

Green's objection is that the bank story cannot apply after the mid-1980s reforms, yet the term premium persists, and the short-end implied tax averages below the corporate rate and is variable; hence one tax rate must explain the curvature. This is decisive against a clientele sorted by deductibility and abolished by statute, but it leaves intact a clientele sorted by the tradeoff between tax benefit and duration exposure. That story predicts that the pattern survives the 1986 reform, with households at the short end and lower-wedge institutions at the long end, and a marginal wedge that tracks supply and clientele capacity instead of any statutory rate. The ownership evidence of Section 4.2 shows the required maturity gradient in holdings.

Tax-timing strategies are compatible with the participation mechanism. Added to an investor's taxable opportunity set, with their tax cash flows valued explicitly, they change her relative valuation without making her marginal at every maturity. The paper measures the puzzle against $\xi_H y_T^C$ and quantifies tax timing separately. The distinctive question is

whose tax benefit is capitalized at which maturity, and P1 (the household exemption) and P2 (insurer and bank benefits) answer it through investor-specific comparison across clienteles.

A.9 Institutional detail: tax treatment, frictions, and accrual rules

This appendix gives the institutional detail behind §4. It covers the federal tax treatment of municipal and taxable interest by investor type, including the proration and interest-expense rules for insurers and banks. It also covers the incidence of the state exemption, the frictions that limit cross-market arbitrage, and the accrual rules that map prices into after-tax returns.

A.9.1 Investor-by-investor federal tax treatment since 2000

The value of the federal exemption depends on the investor's marginal rate on taxable interest, on any partial taxation of its municipal income (insurer proration), and on whether its borrowing is deductible. The wedge between taxable and tax-exempt cash flows therefore differs across investor types and shifts at the major federal reforms.

Notation: net-of-tax rates. The absolute coupon keeps map pre-tax into after-tax cash flows:

$$\xi_i^C \equiv 1 - \tau_i^C, \quad \xi_i^M \equiv 1 - \tau_i^M,$$

where τ_i^C is investor i 's effective marginal rate on taxable coupons and τ_i^M its rate on municipal coupons, strictly positive for prorated institutions. The pricing conditions use the keep ratio $\xi_i = (1 - \tau_i^C)/(1 - \tau_i^M) = \xi_i^C/\xi_i^M$; for prorated insurers ($\tau_i^M > 0$) it is this ratio, not the difference $\xi_i^M - \xi_i^C$, that is calibrated. Here τ_i^C, τ_i^M are rates on the two coupon streams; the model's single pricing object $\tau_i = 1 - \xi_i$ is one minus the keep ratio built from them.

A.9.1.1 Taxable households

Municipal coupons are exempt and taxable coupons are taxed as ordinary income, so in the notation of §5.2 a federal-only-resident household has

$$\xi_h^{M,\text{fed}} \approx 1, \quad \xi_h^{C,\text{fed}} \approx 1 - \tau_h^{\text{ord}},$$

where τ_h^{ord} is its ordinary-income rate. The relevant rate is the top or near-top household rate among active municipal investors, and the household advantage increases in income, which is why state-level household tax variation (Section A.9.3) yields sharp cross-sectional predictions when households are marginal. Figure 1 plots tax-exempt interest reported on individual income-tax returns by adjusted-gross-income bracket: high-income households are the main recipients, but substantial amounts appear in lower brackets.

The top household rate over 2000–2020, and the NIIT. Since 2013 the 3.8% Net Investment Income Tax (NIIT) applies to taxable interest but not to tax-exempt municipal interest, widening the household advantage beyond the statutory ordinary rate. The top ordinary rate followed the 2001 EGTRRA phase-down (39.6% $\rightarrow$ 35% by 2003, held through 2012; the top long-term capital gains rate fell to 15% in 2003), the 2012 American Taxpayer Relief Act’s return to 39.6%, and TCJA’s 37% in 2018. Including the NIIT, the average effective top rate on taxable interest over 2000–2020 is about 38.4% (a keep of roughly 0.62), the calibrated ξ_H^{fed} . The corporate rate has a single break, 35% through 2017 and 21% from 2018, the sole source of the insurer keep changes used in the TCJA counterfactual (Appendix A.4) and the TCJA test (§7.3).

A.9.1.2 Banks: interest-expense disallowance and bank-qualified bonds

Banks are corporate taxpayers, but interest expense allocable to carrying tax-exempt obligations is not deductible, which penalizes levered municipal carry trades and limits banks as arbitrageurs between the taxable and tax-exempt markets. TEFRA imposed a 15% disallowance, raised to 20% by the 1984 Act; the 1986 Act then denied financial institutions the deduction for interest allocable to tax-exempt obligations acquired after August 7, 1986, and bank municipal holdings fell. The bank-qualified exception is defined in Sections 265(b)(3) and 291: an 80% deduction for designated obligations of qualifying small issuers under the \$10 million annual issuance limit, expanded for 2009–2010 issues (see §7.2). It leaves banks important holders in that segment. Outside it, and at long maturities where balance-sheet costs are larger, banks are imperfect arbitrageurs of taxable–tax-exempt parity. Banks motivate the bank-qualified empirical design but are not a separate calibrated clientele; I_2 denotes life insurers.

A.9.1.3 Property & casualty insurers: federal proration and the effective wedge

P&C insurers pay the corporate rate τ_c on taxable coupons, and proration under §832(b)(5) subjects a fraction of their tax-exempt interest to tax:

$$\tau_{\text{pc}}^M \approx \kappa_{\text{pc}} \tau_c,$$

where κ_{pc} is the applicable proration percentage. Before TCJA the percentage was fixed at 15%, an effective municipal tax of $0.15 \times 0.35 = 0.0525$. TCJA replaced it with 5.25% divided by the highest corporate rate, $5.25/21 = 25\%$ at the 21% rate, so that a corporate rate change moves the percentage and preserves the 5.25% effective tax; a counterfactual holding the percentage fixed would be a different tax rule.

Effective muni advantage relative to taxable bonds. The P&C insurer's net-of-tax coupon rates are

$$\xi_{\text{pc}}^C = 1 - \tau_c, \quad \xi_{\text{pc}}^M \approx 1 - 0.0525 = 0.9475,$$

and parity, abstracting from risk and liquidity, requires

$$y^M \cdot \xi_{\text{pc}}^M \approx y^{\text{tax}} \cdot \xi_{\text{pc}}^C \quad \Rightarrow \quad \frac{y^M}{y^{\text{tax}}} \approx \frac{1 - \tau_c}{0.9475}.$$

At $\tau_c = 35\%$ the ratio is $0.65/0.9475 \approx 0.686$, an implied wedge of about $1 - 0.686 \approx 31.4\%$; at $\tau_c = 21\%$ it is $0.79/0.9475 \approx 0.834$, a wedge of about $1 - 0.834 \approx 16.6\%$. A corporate tax cut therefore sharply reduces the insurer's municipal advantage while leaving its effective muni tax unchanged, an investor-specific shock that the mechanism transmits most strongly to the maturities where insurers are marginal.

The two-sided keep. The keep ξ_i of (1) in §4.1 is the ratio of after-tax retention on the two coupon streams: the textbook $1 - \tau_i^C$ when munis are fully exempt ($\tau_i^M = 0$), and above it when the muni coupon is itself partially taxed (insurer proration, or a resident's out-of-state munis). For the P&C insurer, $\tau_{\text{pc}}^C = 0.35$ and $\tau_{\text{pc}}^M = 0.0525$ give $\xi_{I_1} = (1 - 0.35)/(1 - 0.0525) = 0.686$; Table 2 reports the wedges $1 - \xi_i$ for each clientele.

Tax status of muni-active insurers. The wedge presumes a marginal insurer that expects to face the corporate tax; the flat statutory schedule at the scale of large insurers, loss carryforwards, and the AM Best association between past tax payments and muni holding (Table 42) support that calibration. Holdings are a noisy read on tax status: a zero-current-tax investor has no tax-based muni advantage but can still hold munis for duration matching, diversification, or expected-return compensation. The calibrated keep $\xi_{I_1} = 0.686$ is therefore a maintained marginal-clientele calibration informed by the tax-status evidence.

Table 42: Tax-paying status by tax-exempt muni holding (AM Best, life and health insurers).

Current tax-exempt holding	Taxed at $t - 1$	Share of prior years taxed	N
Holds none	65.1%	67.1%	14,317
Holds any (> 0)	72.0%	75.2%	8,366

Note. Insurer-years for life and health insurers, where AM Best reports firm-level tax status. Municipal holders are more likely than non-holders to have paid corporate tax in the prior year, and paid tax in a larger share of all prior years, consistent with taxpayers self-selecting into muni holding.

A.9.1.4 Life insurers and other institutions

Life insurers face proration under §812: net investment income, including tax-exempt interest, is split into a “company share,” whose tax-exempt interest is excluded, and a “policyholder share” taxed through a reduced reserve deduction. TCJA fixed the company share at 70% (policyholder share 30%), so the effective tax on a life insurer’s muni interest is approximately $\kappa_{\text{life}} \tau_c$ with $\kappa_{\text{life}} \approx 30\%$, about $30\% \times 21\% \approx 6.3\%$ after the reform: $\xi_{\text{life}}^M \approx 1 - 0.063 = 0.937$, $\xi_{\text{life}}^C = 1 - \tau_c = 0.79$, and $\xi_{I_2} \approx 0.79/0.937 \approx 0.843$ post-TCJA, slightly lower-tax than the P&C insurer (0.834). This is the post-reform value used in the TCJA counterfactual (Appendix A.4); the pre-reform calibration sets $\xi_{I_2} = 0.79$. Tax-exempt institutions (many pensions and endowments) and tax-deferred household accounts (IRAs/401(k)s) gain little from the coupon exemption at the account level and are not natural marginal price setters.

A.9.2 Relative tax keeps, statutory liabilities, and fiscal scope

The normalized keeps used for portfolio choice do not identify tax receipts. For a P&C insurer at the pre-TCJA top corporate rate, the absolute corporate and municipal keeps are 0.65 and 0.9475, with ratio 0.6860. Applying $1 - 0.6860$ to corporate interest while treating municipal interest as untaxed would misstate both tax bases. Restoring the absolute keeps preserves portfolio choices only under the preference-unit transformation of Section 5.2. For life insurers, pre-2018 proration depended on company-specific investment income and policy obligations, while Section 812 fixes the post-2017 policyholder share at 30%. The fiscal ledger therefore combines the dated corporate rate with the calibrated relative keep through equation (56). Before TCJA, $\xi_{I_2} = 0.79$ and the 35% corporate rate imply an effective municipal rate of 17.72%. After TCJA, $\xi_{I_2} = 0.843$ and the 21% corporate rate

imply 6.29%, consistent with the statutory 30% proration share up to rounding.⁶

The cash term in the portfolio objective is an opportunity-return benchmark, neither a modeled cash balance nor evidence of deductible borrowing. The reduced-form problem imposes no wealth or cash-balance constraint, so a dollar ledger that introduces cash or borrowing must specify the income, expense, and disallowance rules separately. Household state taxes belong to the residence jurisdiction, with any federal deduction computed under the relevant year’s rules: itemized state taxes on federally exempt interest can reduce federal tax, which one relative keep applied to combined tax cannot capture.⁷ The yield-based flow $y_T Q_T$ is a replacement-financing proxy, not the contractual coupon or original issue discount of an outstanding bond. A zero accrues original issue discount on its original constant-yield basis path, with tax-year allocations independent of the model’s two-year grid. A change in its market yield resets neither that path nor the issuer’s promised redemption. Dollar fiscal projections therefore require an issue-cohort and refinancing interpretation as well as absolute tax rates; a state’s own fiscal incidence uses that state’s revenue, and an economy-wide calculation adds issuer deductions and other counterparties. The ledgers of Sections 8.1 and 8.2 apply these statutory rates to the model’s equilibrium flows as incidence scenarios in the model’s units. Their accounting identity verifies arithmetic consistency within the maintained flow convention.

A.9.3 State tax exemption and who receives it

Many states exempt residents’ interest on in-state municipal bonds while taxing interest on out-of-state municipal bonds; this in-state advantage is central to the cross-sectional identification, and its incidence differs across investors.

A.9.3.1 Resident households in taxable accounts

For a taxable household in state s with state marginal rate τ_s ,

$$\xi_h^C(s) \approx (1 - \tau_h^{\text{ord}})(1 - \tau_s), \quad \xi_h^{M, \text{in-state}}(s) \approx 1 \quad (\text{absent AMT/PAB exposure}),$$

⁶See the IRS, [2017 Form 1120-L instructions](#), Schedule F, and the Joint Committee on Taxation, [General Explanation of Public Law 115-97](#), pp. 237–240.

⁷IRS, [Publication 550 \(2017\)](#), p. 36. Later-year SALT limits, AMT status, and NIIT require the corresponding taxpayer-specific treatment.

whereas out-of-state municipal interest is generally state-taxable:

$$\xi_h^{M,\text{out-of-state}}(s) \approx (1 - \tau_s).$$

The wedge is concentrated among taxable households and largest in high-tax states.

A.9.3.2 Households via municipal mutual funds and ETFs

Federally exempt interest passes through to shareholders as exempt-interest dividends, but a resident of state s may owe state tax on the share attributable to out-of-state issuers. State-tax incentives therefore reach households investing through funds, attenuated according to fund composition.

A.9.3.3 Institutions: the state exemption is a resident-household wedge

The household advantage is keyed to the muni's issuer state matching the holder's residence state and scales with the resident's state rate. Institutions do not face this wedge. Tax-deferred accounts, pensions, and endowments owe no state income tax on the interest. Many states tax P&C insurers primarily through premium taxes instead of corporate income taxes, so federal proration is first-order. Banks and corporations owe state corporate tax apportioned across their business footprint, and their exemptions vary with the state base regardless of the issuing state. The institutional channel is therefore structurally different from the household channel. It does not track the issuer-state variation P1 exploits, so state tax wedges are modeled as a household object. For insurers and banks, the federal wedges (proration; interest-expense disallowance and bank-qualified status) drive clienteles at longer maturities.

A.9.4 Shorting and financing frictions and their tax treatment

Tax parity requires cheap cross-market trades (levered long munis against short taxables) until after-tax risk-adjusted returns align; several frictions block them, especially at long maturities.

A.9.4.1 Shorting munis

Municipal securities lending is thin relative to treasuries and large on-the-run corporates, and fragmented across issuers and CUSIPs. This reduces borrow supply and raises search and balance-sheet costs. A short must also compensate the lender for coupons with substi-

tute payments whose tax character is unfavorable relative to the tax-exempt coupon, which discourages lending and raises the cost of maintaining a muni short.

A.9.4.2 Balance sheet, haircuts, and the maturity gradient

Leverage constraints, haircuts, and regulatory or internal balance-sheet costs rise with maturity, because long-duration positions have larger mark-to-market volatility and require more capital; call features and episodic liquidity add hedging and financing costs. Relatively cheap arbitrage can discipline short-end pricing while the long end remains segmented.

A.9.4.3 The tax treatment of borrowing to invest

For banks, the disallowance of interest expense allocable to tax-exempt obligations is a first-order barrier to levered muni demand outside the bank-qualified segment; other intermediaries face analogous wedges through funding haircuts, repo terms, regulatory constraints, and internal capital charges.

Modeling implication. These are two distinct frictions: the institutional rules raise the cost of financing levered long positions, whereas the model's $x \geq 0$ constraints rule out shorts. The model represents limited arbitrage through no-short constraints, which prevent offsetting long-short positions, and a cost of portfolio risk, which limits desired long duration; it imposes no separate borrowing limit.

A.10 Data sources and construction

A.10.1 Yield construction details

A.10.1.1 Bond-level yields

The municipal curve is built from primary-market offering yields in the Mergent issuance panel (§4.4 gives the reasons); the curve figures use the zero-coupon conversion of step 6 in §A.5.1.2. MSRB secondary-market prices serve only the tests that need them. The issuance premium test uses them (trades in the weeks after the offering against the offering itself, §A.5.1.6), as does P1 (call-adjusted zero-coupon yields, with offering yields as a robustness column in Appendix A.6.5). The TCJA design (customer trades) and the treasury-response design (consecutive trades of the same non-callable bond) use them as well.

A.10.1.2 Call adjustment

No short-rate model on the treasury curve is used. Each callable bond is repriced on the equilibrium municipal discount surface at its own fitted option-adjusted spread, with and without its recorded call schedule. The curve-level adjustment is the staged mean of the per-bond zero-coupon yield difference (§A.5.1.2); the non-callable subsample validates it (Figure 17).

A.10.1.3 Staging and filters

The curve is an equal-weighted three-stage mean (maturity bucket $\times$ issue quarter $\times$ half-notch rating cell, then rating cells within a quarter, then quarters); §A.5.1.6 shows why an unstaged par-weighted mean misstates the long end. The sample is federally tax-exempt, non-AMT bonds with average agency score seven or better (A— and above) over 2000–2020; the AA-and-above cut (score three or better) is the robustness sample of Appendix A.7.

A.10.1.4 Benchmark curves

The taxable benchmark is the monthly A-or-above corporate zero-coupon curve (the published High Quality Market curve, §3.1), the after-household-tax benchmark is $\xi_{HY_T}^C$, and the GSW treasury zero-coupon curve supplies the maturity-matched treasury yield of the P3 design.

A.10.2 Insurer exposure construction details

A.10.2.1 Holdings aggregation from NAIC Schedule D

Par holdings of municipals mapped to issuing state s in the master file are summed across P&C insurers at the CUSIP-year level, then to the state-year level.

A.10.2.2 Outstanding supply and exposure shares

Outstanding par by state-year comes from the issuance data; $PCShare_{s,t}$ is insurer par holdings divided by outstanding par, and P2 fixes it at $PCShare_{s,2015}$, the end-of-2015 level, so that exposure predates the estimation window.

A.10.2.3 Maturity-bucket exposure measures

Variants restrict holdings, outstanding par or both to maturity buckets ($T > 10$, $T > 15$, long-end quantiles) for the maturity-specific predictions.

A.10.2.4 Municipal ownership by broad investor group

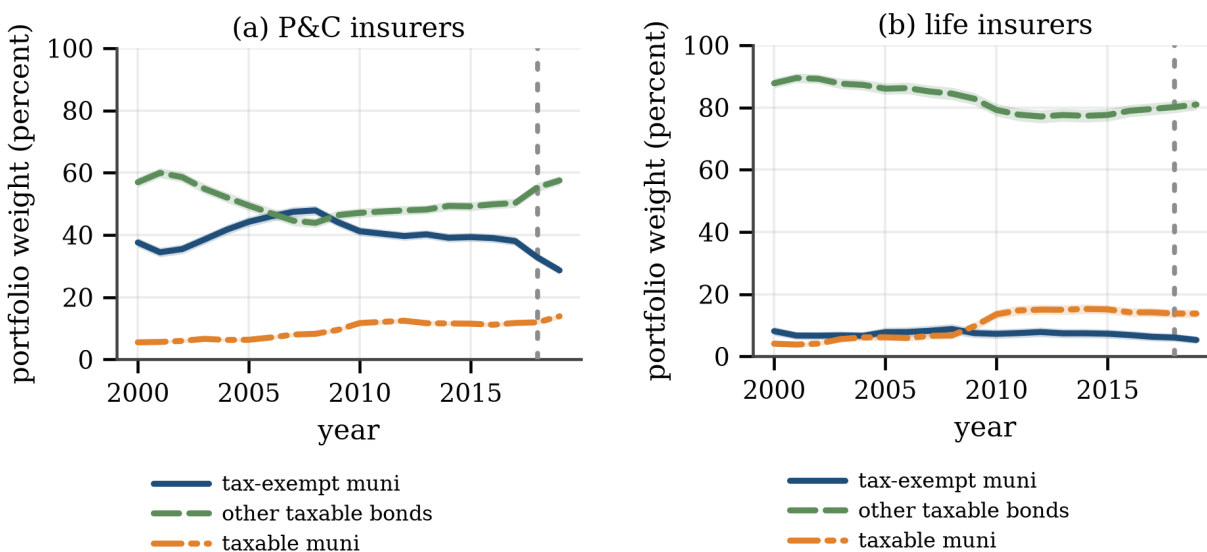
Figure 1 plots municipal holdings by broad investor group over 2004–2020, the composition from which §4.2 reads the clientele ranking.

A.11 Investor holdings by tax status

The two figures are the holdings counterpart to Table 2: each clientele’s muni share lines up with its federal tax wedge.

A.11.1 Insurer muni portfolio shares over time

In Figure 34 the P&C muni share exceeds the life share throughout, consistent with the persistent ownership composition the identification relies on (§4.2), and falls after the 2018Q1 TCJA corporate-rate cut, the quantity counterpart of the long-end repricing tested in P2 (§7.3).



Panel A: P&C insurer muni share over time.

Panel B: Life insurer muni share over time.

Figure 34: Insurer portfolio composition over time. Corporate-and-municipal bond portfolios are split into tax-exempt munis, taxable munis, and corporates for P&C insurers (Panel A) and life insurers (Panel B), 2000–2019, from NAIC Schedule D. P&C insurers hold a much larger muni share than life insurers throughout, and the P&C share falls after the 2018Q1 TCJA shock.

Alt text: Two time-series panels show municipal bonds as a share of P&C and life-insurer bond portfolios. P&C insurers hold the larger municipal share, which declines following the TCJA.

A.11.2 Pension-FOIA muni holdings by tax status

Figure 35 shows the pension sample, which lacks a federal tax advantage, holding essentially no federally tax-exempt munis. It is descriptive support for a small tax-motivated pension role. It does not by itself identify the active set from tax status: the model permits $\xi_i \approx 1$ investors to hold tax-exempt munis for duration, diversification, or expected-return compensation.

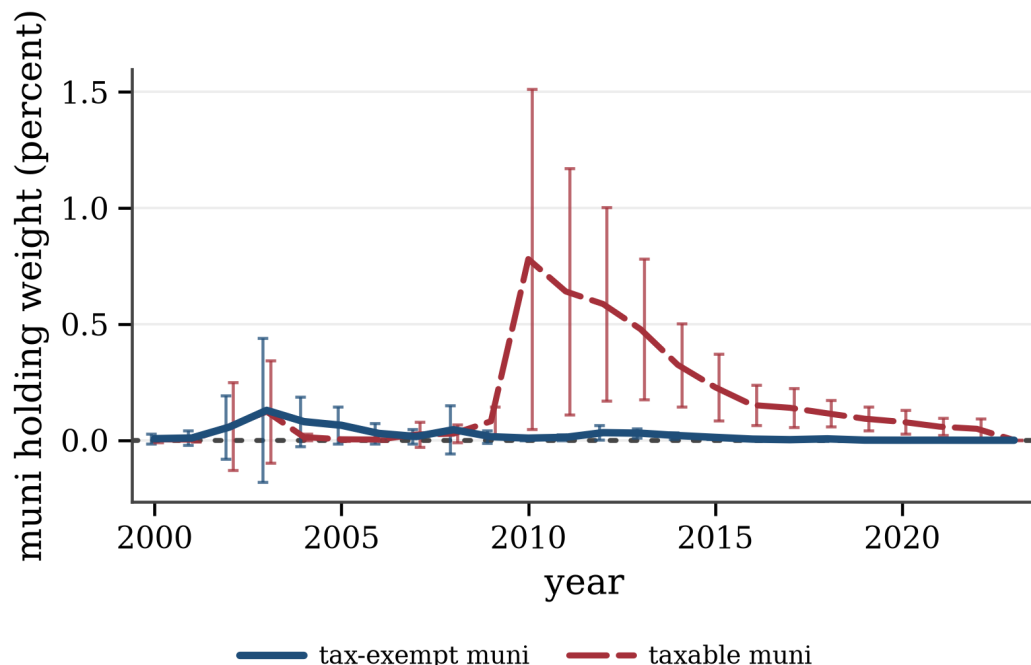


Figure 35: Pension-FOIA panel: share of muni holdings by federal tax status as a fraction of total bond portfolio, by year. State and local public pension funds hold essentially zero federally tax-exempt municipals across the sample; the muni exposure they report is overwhelmingly in taxable munis (Build America Bonds and other taxable issuance), where the federal exemption is irrelevant. Sample: large U.S. state and local public pension systems, fiscal years 2000–2023, with coverage broadening to the full set of 30 plans from 2015, from the custom CUSIP-level FOIA panel of §3.1.

Alt text: Pension municipal holdings are separated by tax status. Tax-exempt municipal exposure is small, while taxable municipal bonds account for most of the reported municipal exposure.