Journal of Financial Economics

Robust difference-in-differences analysis when there is a term structure☆

Kjell G. Nyborg [a,b,c,*], Jiri Woschitz [d]

[a] Department of Finance, University of Zurich, Switzerland
[b] Swiss Finance Institute, Switzerland
[c] CEPR, United Kingdom
[d] Department of Finance, BI Norwegian Business School, Norway

ARTICLE INFO

JEL classification:
G12
C20
E43
E47

Keywords:
Fixed-income pricing
Term structure
Difference-in-differences
Treatment effects
False effects
Garbled measurement
Matching

ABSTRACT

For variables with a term structure, the standard difference-in-differences (DiD) model is predisposed toward misspecification, even under random assignment, because of heterogeneity over the maturity spectrum and imperfect matching between treated and control units. Estimated treatment effects that are false, biased, or hard to interpret become a concern. Neither unit fixed effects nor standard term-structure controls resolve the problem. Solutions that overcome imperfect matching involve estimating the term structure of hypothesized treatment, which is also what is economically interesting (regardless of matching efficiency). These issues are not unique to DiD analysis, but are generic to group-assignment settings.

1. Introduction

Difference-in-differences (DiD) methodology is widely used in finance to analyze fixed-income pricing data. Often, a security's price is expressed inversely in terms of its yield (or a spread) and DiD analysis is applied by running a classical DiD regression of the form

$$yield_{it} = \alpha_i + \delta_t + \beta_{DiD}\ \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \mathbf{\Gamma}'\mathbf{Z}_{it} + \varepsilon_{it}, \quad (1)$$

where $yield_{it}$ is security i's yield-to-maturity at time t and the right-hand side of the equation represents the typical DiD structure: α_i and δ_t correspond to security- and time-fixed effects and $\mathbb{1}_{Treated,i}$ and $\mathbb{1}_{Post,t}$ to treatment and post-event indicator variables, respectively, $\mathbf{Z}_{it}$ is a vector of control variables, $\mathbf{\Gamma}$ is a vector of coefficients, β_{DiD} is the treatment effect, and ε_{it} is an error term. DiD methodology is designed to deal with endogeneity, i.e., to measure the causal impact of a treatment on an outcome variable (yield in this case) by comparing treated to non-treated control units (in this case fixed-income securities) over the treatment event. However, for variables, such as yield, that exhibit a term structure, Specification (1) has drawbacks for both economic and econometric reasons. First, it is not designed to capture term effects. But if a variable exhibits a term structure, there is little reason to believe that treatment effects should be homogeneous over the maturity spectrum, that is, have no term structure. Typically, in applications, we should be interested in estimating the term structure of the treatment effect. Second, because term structures move around from date to

☆ We would like to thank the editor, Toni Whited, and an anonymous referee for their constructive feedback, which helped improve the paper. We gratefully acknowledge financial support from the Swiss National Science Foundation, Switzerland (project #100018_172679 "Trading and Financing during Market Stress"). We would like to thank participants of the brown-bag seminar at BI Norwegian Business School in 2022, the 2023 Financial Econometrics Conference to mark Stephen Taylor's Retirement (Lancaster), the 2023 FMCG Conference at Deakin University, the 2023 EFA annual meeting (Amsterdam), the 2023 FMA annual meeting (Chicago), and the Eurofidai-Essec Finance Meeting in 2023 (Paris), and, especially, Massimiliano Bondatti, Matthias Fleckenstein, Thomas Poulsen, and Karamfil Todorov, for constructive comments and suggestions. The paper was a semifinalist for best paper in Investments at the FMA 2023 (Chicago).

Corresponding author at: Department of Finance, University of Zurich, Switzerland.

Received 24 July 2024; Received in revised form 16 April 2025; Accepted 18 April 2025

Available online 17 May 2025

date, if treated and control units are imperfectly matched on residual maturity, as is almost always the case in the literature (see below), the classical DiD specification is misspecified even if a term structure control is included in $\mathbf{Z}_{it}$. In this paper, we study the term-effect problem and propose methods for dealing with it. While our exposition is cast in terms of fixed-income securities and yields, the issues we discuss are relevant whenever the outcome variable exhibits a term structure, with some examples being interest rate spreads (e.g., credit spreads), loan rates, option-implied volatilities, futures prices, and risk premia.

Fixed-income securities, in practice and in theory, are priced against the term structure of interest rates, typically of the issuer (or a set of homogeneous issuers) and using its individual maturities. Because term structures move from day to day, the relative yields of bonds in different parts of the maturity spectrum also change irrespective of hypothesized treatment. This matters in a DiD analysis because treated and control bonds are rarely, if ever, matched on residual maturity and coupon structure, typically resulting in nonzero correlation between duration and the treatment indicator variable. For variables that exhibit term structures, the classical DiD model is misspecified because it assumes that treatment unrelated effects for each unit are fixed, when they actually depend on residual maturity and underlying changes in the term structure. In short, under Specification (1), the zero correlation condition for consistent OLS estimators is likely to be violated (Roberts and Whited, 2013). Long event windows exacerbate the problem because the residual maturity of an individual unit decreases as time progresses, which gives rise to a curve roll effect that is heterogeneous over the maturity spectrum. Even in the unlikely case that treated and control units are perfectly matched, Specification (1) remains unsatisfactory because it only measures the average treatment effect of treated units in the sample rather than the term structure of the treatment effect, which should be the main object of interest for variables that exhibit term structures. Estimated treatment effects that are false, biased, or hard to interpret are a concern under the classical DiD specification.

How big are term effects in practice? To get a sense of this, Table 1 provides some summary statistics in the government-bond space on monthly term-spread changes for twelve countries (data from Bloomberg). The government-bond term structure is one of the most accepted pricing kernels in finance. Over the period January 3, 2000 to December 14, 2022, the standard deviation of the monthly change in the ten-year minus three-month government-bond term spread varies from 10 basis points (bps) for Japan to 283 bps for Greece (Panel A).[1] The median standard deviation is 31 bps. The mean absolute monthly change varies from 7 bps (Japan) to 97 bps (Greece) and is 23 bps for the US. These numbers show that changes in the relative yields of bonds over the maturity spectrum depend on the term structure one is dealing with (e.g., Japanese or Greek government bonds) and can be substantial. Magnitudes can potentially dwarf true treatment effects, which one may reasonably expect to be in single digit basis points in many applications. For example, Swanson (2011) estimates that the effects of the Fed's Operation Twist in 1961 were +11, +6, −3, −9, and −8 bps at maturities of three months, one year, and two, five, and ten years, respectively. Overcoming the term-effect problem by matching treated and control bonds on residual maturity is difficult in practice because issuers typically have only a few outstanding securities and the range of maturities can be large. For instance, on a selected date (January 1, 2023), for six of the countries in Table 1, there are less than 22 maturity dates that are shared by two or more bonds with good prices (Internet Appendix, Table A.1). Even if feasible, matching typically requires discarding much of the data. "Close matching" can be an alternative, but the term structure can exhibit significant variation even within relatively short ranges. For example, breaking the maturity spectrum up into (mostly) 2–3 year ranges as in Panel B, we see that the standard deviation of the monthly range spread across the selected countries (Greece excluded) goes from 9 bps (7y-5y) to 33 bps (2y-3m). A third of the countries have standard deviations in double digit basis points for all ranges. Spot rates at different maturities can also move in opposite directions (Panel C).

The term-effect problem has received little attention in the literature. We have manually searched the top three finance journals for papers using DiD methodology on outcome variables with term structures (Table 2). The twenty-one papers we have found study a variety of securities and loans. All use versions of Eq. (1). Modifications include an i subscript on the post-event dummy for staggered DiDs; transposition into differences; and fixed effects at the firm or issuer level, especially when treatment is defined at those levels rather than at the individual security level. In the last case, the unit of analysis is typically new issues or loans, with different pre- and post-event units. Notably, *none of the papers estimate the term structure of the hypothesized treatment effect.* Only two papers attempt pairwise matching of treated and control units on residual maturity (and other pertinent characteristics), but the descriptive statistics show significant differences (Panel A). In Pelizzon et al. (2024), the difference is two years; and in Choi et al. (2020) it is one year.[2] Eight of the remaining papers contain information on residual maturities, with three of these reporting similar average residual maturities for treated and control units. But this does not imply pairwise matching, and statistics on higher moments that are needed to ascertain matching-closeness are hard to find. Ten of the papers try to control for term effects by working with spreads over maturity-matched government bonds or money market rates. But such spreads also exhibit term structures.[3] Fifteen papers include a maturity control in $\mathbf{Z}$. But this does not resolve the misspecification problem of the classical DiD approach. In fact, as we show, *it can make it worse* (Section 4.4). Common tactics that do not resolve the problem include (i) aggregating to higher levels (firm, country, bank-firm) (ii) using log transformations of residual maturity, or (iii) working with estimates of expected rates of return. Most of the papers use long event windows, multiple months or years, which exacerbate the problem. The impact of term effects is especially difficult to assess when the full sample period is simply divided into pre- and post-event periods (Panel D). Overall, the overview in Table 2 suggests that the term-effect problem is not well appreciated in the literature.

In this paper, we use simulation to study the term-effect problem. Focusing on the single event date case with treatment defined at the bond level, we estimate Specification (1) on simulated zero coupon bonds under two types of yield-curve effects, namely, treatment-unrelated idiosyncratic effects and true treatment-related systematic effects. Idiosyncratic yield-curve effects are unrelated to hypothesized treatment; they move yields of all bonds, independent of assignment, but heterogeneously over the term structure. True treatment effects are also heterogeneous over the maturity spectrum. To study the effects of different degrees of congruence between the maturities of treated and control bonds, we draw residual maturities for treated and control bonds from different distributions ordered by first order stochastic dominance. In line with what we believe is most interesting in practice, in our simulations, the magnitude of the idiosyncratic effects (at their peaks) is larger than the true treatment effects. To isolate the term

[1] We use the terminology government bond for central-government securities of any maturity, e.g., for the US, T-bills, T-notes, and T-bonds.

[2] Matching approaches are sometimes also used in the non-DiD fixed-income literature (Fleckenstein et al., 2014; Fleckenstein and Longstaff, 2020).

[3] A large literature shows that there is a term structure in yield spreads, typically calculated relative to Treasuries or reference rates such as Libor. See, e.g., Sarig and Warga (1989), Helwege and Turner (1999), Elton et al. (2001), Huang and Huang (2012), Bao and Hou (2017). Maturity matching is by itself typically imprecise because exact matches are rare and because of the complexities associated with coupon bonds.

Table 1

Time-variation in the term spread in practice (monthly).

Panel A provides the standard deviation and summary statistics on the absolute value of the monthly change in the term spread (10y−3m) for a selected group of countries over the period from January 3, 2000 to December 14, 2022 using zero-coupon yields. For the same countries, Panel B shows the standard deviation of the monthly term-spread change for other term-spread definitions (as indicated in the table). For each of the maturity ranges in Panels A and B, Panel C shows the percentage of months the longest and shortest rates moves in opposite directions (twists). Euro area countries are marked with a star.

Data source: Bloomberg.

Panel A: Monthly changes in 10y-3m term spread (in bps)

Country	St. dev.	Absolute value				N
		Mean	Median	Min	Max	
Japan	10	7	5	0	49	274
Germany*	22	16	12	0	149	274
France*	23	16	11	0	118	239
Netherlands*	24	17	12	0	151	261
United Kingdom	27	19	15	0	143	274
United States	30	22	18	0	103	274
Spain*	31	20	13	0	153	271
Italy*	32	21	15	0	219	274
Ireland *	34	21	14	0	208	274
China	35	25	18	0	166	220
Portugal*	46	27	14	0	285	274
Greece*	274	94	17	0	1,810	274
All countries	85	26	13	0	1,810	3,183
Without Greece	30	19	13	0	285	2,909

Panel B: St. dev. in monthly term spread changes for different maturities (in bps)

Country	2y-3m	5y-2y	7y-5y	10y-7y	15y-10y	N
Japan	6	5	3	4	4	274
Germany*	17	9	5	5	6	274
France*	16	9	5	4	5	239
Netherlands*	18	10	5	4	5	261
United Kingdom	22	10	6	6	6	274
United States	21	12	6	5	6	274
Spain*	30	14	7	7	7	271
Italy*	26	12	7	10	9	274
Ireland *	56	28	13	11	13	274
China	22	13	10	11	15	220
Portugal*	68	32	21	24	15	274
Greece*	263	107	54	58	51	274
All countries	83	35	18	20	17	3,183
Without Greece	33	16	9	10	9	2,909

Panel C: Percentage of long- and short-rate moves in opposite direction (twists) across months

Country	2y-3m	5y-2y	7y-5y	10y-7y	15y-10y	10y-3m	N
Japan	38.7	17.5	11.7	7.7	10.2	47.8	274
Germany*	31.4	9.5	5.5	5.1	8.8	35.4	274
France*	26.8	11.3	5.0	5.0	4.2	33.9	239
Netherlands*	31.0	13.8	5.0	4.6	6.9	38.3	261
United Kingdom	34.7	14.2	5.1	7.7	9.5	41.6	274
United States	38.7	12.0	5.5	6.2	7.7	49.3	274
Spain*	31.4	15.9	7.4	7.0	7.7	37.3	271
Italy*	29.6	11.7	7.7	4.7	9.9	35.4	274
Ireland *	31.8	12.0	5.8	6.6	9.1	37.2	274
China	27.7	24.5	10.9	13.2	19.1	44.5	220
Portugal*	31.8	12.0	8.4	9.5	9.5	40.9	274
Greece*	11.7	8.8	6.2	5.8	12.0	22.3	274
All countries	30.5	13.4	7.0	6.8	9.5	38.6	3,183
Without Greece	32.3	13.9	7.0	6.9	9.2	40.2	2,909

effect, yields on individual bonds are simulated without error and all effects apply without error.

We first show that in the absence of a true treatment effect, the classical DiD specification measures a false treatment effect—"false" because, in fact, there is no treatment effect. Unless treated and control bond residual maturities have the same unconditional distribution, DiD estimates are biased and the incidence of Type I errors is excessive relative to a well specified model (see below). False positives can go in either direction. They can be economically large even when unconditional maturity distributions are the same (treatment assignment is random), which is not the case under a well specified model.

In the unrealistic scenario that there are only true treatment effects, Specification (1) is still problematic because it returns an average effect across treated-bond maturities. That this is problematic is especially apparent if the true effect twists the treated curve up at one end and down at the other end. In the more realistic scenario that there are both idiosyncratic and true treatment effects, we show that the classical DiD specification produces the sum of the results under the two corner scenarios. The upshot is garbled results and inference, which can be severe. Even under random assignment, it is possible to measure a significant *negative* treatment effect when the average true treatment effect on treated sample bonds is *positive*, or vice versa. This is the case even though, in our simulations, yields on individual bonds are measured without error and all effects apply without error. The term-effect problem becomes worse the bigger is the difference between the distributions of treated and control bonds over residual maturity.

These results imply that the standard DiD approach is unreliable when the variable of interest exhibits a term structure. False positives

Table 2
Maturity control in the literature.
This table shows a collection of recent top finance publications that apply DiD methodology to variables with term structures. All papers use a version of Specification (1). Modifications include an *i* subscript on the post-event dummy for staggered DiDs; transposition into differences; and fixed effects at the firm or issuer level, especially when treatment is defined at those levels. In the last case, the unit of analysis is typically new issues or loans, with different pre- and post-event units. No paper estimates the term structure of the treatment effect. The list is created by manually searching the Journal of Finance (JF), the Journal of Financial Economics (JFE), and the Review of Financial Studies (RFS) using relevant combinations of key words. N.a. means that the respective information is not available. **Z** refers to the vector of controls in Specification (1) (see text). D, M, Q, and Y refer to daily, monthly, quarterly, and yearly, respectively.

Pairwise maturity matching attempted								
Publication	Unit of treatment	Unit of analysis	Maturity congruence	Maturity control in **Z**	Dependent variable	Event window Pre	Event window Post	Time unit
Panel A: Staggered treatments (one unit at a time).								
Choi et al. (2020, JFE)	Bond (corporate)	Same as for treatment	No[1]	No	Yield less maturity matched Treasury (in changes)	[−4,−1]	[0,1]	Q
[1] Treated bond maturities are one year longer on average. Matching on issuer, credit rating, seniority, option features, and residual time to maturity.								
Pelizzon et al. (2024, JFE)	Bond (corporate)	Same as for treatment	No[2]	No	Yield less maturity matched German bund curve	[−30,−1]	[0,30]	D
[2] Treated bond maturities are two years longer on average. Matching on credit rating, issue size, and residual time to maturity.								
No pairwise maturity matching								
			Maturity statistics					
Panel B: One treatment event for all units.								
Ayotte et al. (2011, RFS)	Corporate securitizer	New ABS (AAA rated)	Mean treated maturities half of control units	"Average life" and "average life squared"	ABS yield less maturity matched swap rate	[−6,0)	[0,6]	M
Hasan et al. (2014, JFE)	Firm	New loan	N.a.	Log residual maturity	Loan yield less LIBOR (in logs; not maturity matched)	[−3,−1]	[1,3]	Y
Cornaggia et al. (2018, RFS)	Bond (municipal)	Same as for treatment	N.a.	Duration	Yield or yield less duration matched Treasury[3]	[t_1-30, t_1]	[t_2, t_2+30]	D[4]
[3] Either using each bond's maturity or, for callable bonds, the call dates. [4] Exact treatment timing is unit dependent. All treatments take place between t_1 and t_2. The interim period between t_1 and t_2 has a length of between 31 and 52 calendar days.								
Dannhauser (2017, JFE)	Bond (corporate)	Same as for treatment	Available only for different sample period	No	Yield less maturity matched swap rate	[−6,0)	[0,6]	M
Painter (2020, JFE)	County	New bond (municipal)	Mean treated maturities 1 year longer than controls	Log maturity	Annualized issuance cost[5]	[−6,0)	[0,6]	M[6]
[5] Sum of the initial bond yield and the annualized gross spread. [6] Or longer windows: Either -/+ 12 months or full sample (-35/+124 months).								
Benetton et al. (2021, JFE)	Bank	New loan (agg. to bank-firm level)	N.a.	No	Loan rate	[−2,−1]	[0,3]	Q[7]
[7] The specification includes one treated × post interaction separately for each quarter in the post-treatment period.								
Allen et al. (2023, JFE)	Firm	New loan	N.a.	Maturity	Loan yield less maturity matched Treasury	[−18,−1]	[1,18]	M
Panel C: Multiple treatment events (multiple units per event)								
Santos et al. (2008, JF)	Firm	New loan	Mean treated maturities 1 year shorter than controls	Loan maturity	Loan yield less LIBOR (not maturity matched)	[1987:1, t_1)	(t_2,2002:4]	Q[8]
[8] The treatment periods [t_1, t_2] are the three periods [1989:2,1989:3], [1990:3,1991:2], and [2000:4,2002:1], which are in the specification at the same time.								
Rodano et al. (2016, JFE)	Firm	New loan	N.a.	Loan maturity[9]	Loan interest rate	[−3,−1]	[0,10]	Q[10]
[9] Loan maturity is measured with indicator variables for <1, 1-5, and > 5 years. [10] In specification, post-period split into subperiods [0,1], [2,3], and [4,10].								
Todorov (2020, JFE)	Bond (corporate)	Same as for treatment	Mean (median) treated maturities 1 year (shorter) longer than controls[11]	No	Yield	[−10,0)	[0,12]	W[12]
[11] Additionally: matching on two maturity buckets (above and below median maturity) but no further information on maturity congruence. [12] In specification, post-period split into subperiods [0,5] and [6,12].								
Panel D: Staggered treatment (one unit at a time).								
Chava et al. (2009, RFS)	Firm	New loan	N.a.	No	Loan yield less LIBOR (in changes and logs; not maturity matched)	[1990,0)	[0,2004]	Y[13]
[13] Most recent loan at least 30 days before event announcement and all loans between 30 days after announcement and completion.								
Qiu et al. (2009, JFE)	Firm (via state of incorp.)	Bond (agg. to firm level)	N.a.	Duration and convexity	Yield less maturity matched Treasury	[1976,0)	[0,1995]	Y
Titman et al. (2010, RFS)	Mortgage originator	New mortgage	N.a.	Mortgage residual maturity	Mortgage rate less maturity matched Treasury	[1/1996,0)	[0,12/2002]	M[14]
[14] Mortgage issue classified as treated in month *m* when originator is treated in previous 6 (or 3) months.								
Chan et al. (2013, JFE)	Firm	New loan	Mean is similar. Higher order moments n.a.[15]	Loan maturity	Loan yield less LIBOR (not maturity matched)	[2000,0)	[0,2009]	Y
[15] Propensity-score matching of treated and control firms only on firm (not loan) characteristics.								

(*continued on next page*)

Table 2 (*continued*).

Publication	Unit of		Maturity	Maturity	Dependent	Event window		
	treatment	analysis	statistics	control in **Z**	variable	Pre	Post	Time unit
Adelino et al. (2016, RFS)	Bank (via sovereign)	New loan	N.a.	No	Loan yield less LIBOR (not maturity matched)	[1989,0)	[0,2012]	Y[16]
[16] Loan issue classified as treated if bank-sovereign treated in previous 6 months.								
Amiram et al. (2017, JFE)	Firm	New loan	Mean and standard deviation are similar. Higher order moments n.a.[17]	Loan maturity	Loan yield less LIBOR (in logs; not maturity matched)	[1993,0)	[0,2014]	Y[18]
[17] Robustness using propensity-score matching: Match treated and control firms on firm (not loan) characteristics. No further information on maturity-matching congruence.								
[18] Loan issue classified as treated if firm treated in previous 6 months.								
Almeida et al. (2017, JF)	Firm (via sovereign)	Bond	N.a.[19][20]	Bond maturity	Yield (in changes)	[−3,−1]	[0, t]	M[21]
[19] Matching of treated and control firms only on firm (not bond) characteristics.								
[20] Provide distributions of ratio of long-term leverage separately for treated and control firms but no information on bond maturities. [21] t is 1, 3, 4, 5, or 6 months.								
Gao et al. (2020, JFE)	County	Bond (municipal)	Mean is similar. Higher order moments n.a.	Maturity and its inverse	Yield less maturity matched, coupon-equivalent Treasury	[1999,0)	(0,2015]	Y[22]
[22] Bond classified as treated three years after treatment event in respective county.								
Favara et al. (2021, JFE)	Firm (via state of incorp.)	New loan (or bond)	N.a.	Loan (or bond) maturity	1) Loan yield less LIBOR (not maturity matched) 2) Issue yield less maturity matched Treasury	[1992,0)	[0,2010]	Y

will typically be robust to different control vectors, **Z**, and implementation methods because the result is driven by a combination of underlying changes in the term structure and nonzero correlation between residual maturity and treatment assignment. Including a control for residual maturity in **Z** can even make things worse.

While it is important to control for term effects when using DiD analysis on variables that exhibit term structures, a simple adjustment to Specification (1) that replaces the bond fixed effects (the misspecification element) with maturity or functions of it, does not solve the problem. As an example, we analyze a specification that substitutes the bond fixed effects with a model of the yield curve consistent with that used to simulate the true underlying curves. We show that the resulting DiD estimator is practically the same as when using Specification (1). The problem with the specifications is that they impose, either explicitly through the parametric term structure or implicitly through the bond fixed effects, parallel yield-curve shifts between control and treated bonds, pre- and post-treatment. In contrast, in the simulated data, as in reality, the underlying treatment and idiosyncratic effects are heterogeneous over the term structure.

The challenge is to find procedures that allow for heterogeneous true treatment effects over the maturity spectrum while controlling for ubiquitous treatment-unrelated shifts in the yield curve. We discuss two solutions. The first is tailored to zero coupon bonds and originates with Nyborg and Woschitz (2021). With zeros, we can replace the bond and time fixed effects in Specification (1) with separate parameterized curves for control and treated bonds both pre- and post-treatment. The treatment effect is estimated as a "Delta curve", namely, the incremental spot-curve difference between treated and control bonds over the event. Conveniently, the DiD Delta curve can be estimated by running a single regression using standard software. In this fully flexible yield-curve DiD approach, true maturity-dependent treatment effects are identified and separated from idiosyncratic yield-curve effects. Thus, the misspecification is resolved through the estimation of the term structure of the treatment effect, which is also what is economically interesting. Imperfect matching on residual maturity is not an issue. Since the specification uses the full panel structure of the data, it permits clustering standard errors at the bond level, as recommended by Bertrand et al. (2004) and Petersen (2009). It is also more parsimonious than the classical DiD specification since there are no bond fixed effects.

The second procedure, "semi-synthetic matching", works for coupon bonds as well as zeros. This involves matching each treated bond with a synthetic control having the same coupons and residual maturity. Matching can also be done on a potentially wide range of other attributes. Synthetic control curves have been used in the literature previously in non-DiD settings (Ang et al., 2010). In a first stage, individual DiDs are calculated for treated bonds relative to their synthetic controls. In a second stage, these can be averaged, or what is more interesting, examined over the maturity spectrum. If the sample consists of zeros only, curve fitting in the second stage results in the same DiD Delta curve as under the one-step, fully flexible approach (when functional forms are consistent).

The term-effect problem is also relevant with respect to trend plots, which are typically used as a diagnostic tool to assess the exclusion restriction (Roberts and Whited, 2013). Unless treated and control units are perfectly matched on residual maturity (and coupon structure), standard trend plots based on averages are problematic. A simple solution would be to average yields of bonds within selected maturity buckets and then average across maturity buckets within selected time increments. How useful this is depends on the closeness of maturity matching within buckets. We discuss trend plots further below.

Finally, we show that the issues above are not generic to DiD analysis, but can arise in group-assignment settings more generally. Using a straight cross-sectional setup, we show that failure to properly take heterogeneous term effects into account leads to the same kind of false and garbled effects as in the DiD setup.

Remaining structure: Section 2 reviews additional related literature. Section 3 describes the data simulation. Sections 4 and 6 study false and garbled treatment effects, respectively, while Section 7 combines these problems. Sections 5 and 9 discuss two solutions. Section 8 considers alternative curve specifications and provides an example using real data. Section 10 covers analogous issues in non-DiD settings. Section 11 concludes.

2. Other literature

As outlined above, the term-effect problem relates to both heterogeneous idiosyncratic movements in the variable of interest over the maturity spectrum as well as heterogeneous treatment effects. In the literature, the latter is often dealt with by estimating the heterogeneous treatment effects over the distribution of the dependent variable or by using fixed effects on the discrete right-hand side units present in the data.[4] However, with variables that have term structures, the

[4] See, e.g., Heckman et al. (1997), Bitler et al. (2006), Callaway et al. (2018), Callaway and Li (2019), de Chaisemartin and D'Haultfoeuille (2020).

first approach is uninformative because the shape of the term structure fluctuates over time. The second approach is also problematic because residual maturity is a continuous habitat variable. An intuitive workaround is to divide the data into maturity buckets (Bao et al., 2018; Todorov, 2020). However, we show that this simply pushes the problem to the maturity-bucket level. Moreover, the paucity of bonds in practice limits the fineness of the feasible grid over which maturity buckets can be formed.

Our paper relates to a substantial literature that focuses on estimating unobserved yield-curve parameters (Fama and Bliss, 1987; Nelson and Siegel, 1987; Svensson, 1994; Diebold and Li, 2006; Gürkaynak et al., 2007; Liu and Wu, 2021; Filipovic et al., 2023a,b). Since these parameters can be interpreted as unobserved factors and DiD analysis is a special fixed-effects case, our paper could be viewed from the perspective of the literature on confounding factor structures in fixed-effects settings (Bai, 2009). Specifically, Gobillon and Magnac (2016) and Xu (2017) show that nonzero correlations between factors and treatment can give rise to biased DiD estimators (see also Huang and Östberg, 2023). Nyborg and Woschitz (2021) raise these issues in a term structure setting in the context of measuring the effects of central bank collateral policy. In this paper, we show how the fundamental problems of false and garbled treatment effects arise naturally in fixed-income settings. The DiD Delta curve approach resolves these problems by modeling the term structure as part of the DiD estimator, while the semi-synthetic matching approach removes the term structure before applying DiD. The latter has parallels to the synthetic control literature (Abadie and Gardeazabal, 2003; Abadie et al., 2010; Abadie, 2021) and is a natural approach in a staggered setup, as in Ben-Michael et al. (2022), but our focus is on treatment effects over the maturity spectrum rather than on the average treatment effect across treated units.

There is a growing literature on how DiD methodology can lead to incorrect inference (see, e.g., Callaway, 2023; Roth et al., 2023). In work that has some parallels to ours, Callaway and Li (2023) show, in the context of epidemiology, that standard DiD analysis produces false treatment effects when treated and control regions are at different stages of a pandemic (e.g., COVID-19). While their's is essentially a longitudinal problem and ours is cross-sectional, we share a common prescription, namely, that it is necessary to condition the treatment effect on the confounding element (stage of pandemic, residual maturity). We show how this can be done in fixed-income settings using tools from the curve-fitting literature. Imprecise and hard to interpret treatment effect estimates have been shown for staggered DiD analysis with heterogeneous treatment effects in the time dimension (Baker et al., 2022; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Athey and Imbens, 2022). We examine heterogeneous treatment effects in the cross-section, which we address by modeling curves.

3. Term structure modeling and data simulation

This section describes how we generate the data used to examine the performance of Specification (1) and its alternatives. For simplicity, and to focus squarely on maturity effects, we simulate zero-coupon bonds. The generated data captures the two key features of real fixed-income data discussed above, namely, (i) idiosyncratic, treatment unrelated and systematic, true treatment effects that are heterogeneous over the maturity spectrum, and (ii) non-matched residual maturities for treated and control bonds. The first key ingredient in our data generation process is a model for pre-event spot rates to which we can add post-event idiosyncratic (treatment unrelated) and systematic (true treatment) effects. The second ingredient is a procedure for generating samples of residual maturities for treated and control bonds that allows for different degrees of heterogeneity in these. This permits us to study the implications of an increasing divergence in residual maturities between treated and control bonds. The realized maturities are then used to generate pre- and post-event yields. Data is simulated with either one or both types of post-event term effect. In the simulation, we use parameter values that generate effects with magnitudes that are commensurate with what we see in real data (e.g., Table 1).

3.1. Term structure and different effects

To generate yields, we use Diebold and Li's (2006) factorization of the Nelson and Siegel (1987) curve.[5] The spot rate, or zero-coupon $yield$, with maturity x at time t is

$$yield_t(x; \lambda_t) = \gamma_{0,t} + \gamma_{1,t} \left(\frac{1 - e^{-\lambda_t x}}{\lambda_t x} \right) + \gamma_{2,t} \left(\frac{1 - e^{-\lambda_t x}}{\lambda_t x} - e^{-\lambda_t x} \right), \qquad (2)$$

where $\gamma_{0,t}$ is a long-term or level factor, $\gamma_{1,t}$ is a short-term or slope factor, $\gamma_{2,t}$ is a medium-term or curvature factor, and λ_t is a decay parameter. Over a hypothesized event, we consider two types of effects. First, idiosyncratic (treatment-unrelated) effects move the yields of "treated" and control bonds alike over the event independent of actual treatment. Second, systematic (true treatment) effects affect only the treated bonds. Both types of effects are heterogeneous over the maturity spectrum and are simulated by manipulating the level, slope, and curvature parameters in the Diebold–Li curve, as described in more detail in subsequent sections.[6]

As a brief overview, in Section 4 we focus on idiosyncratic effects and consider two scenarios. In each case, the effects are generated by curves that follow the Diebold–Li specification. In the first scenario, the effect is stronger for short-term rates; it leads to a yield change of −50 bps at a residual maturity of one year, rising to close to zero (+1 bps) at fifteen years. In the second scenario, the effect is stronger for long-term rates; it is close to zero at the short end (+4 bps at one year), falling to −50 bps at fifteen years. We show that the classical DiD specification leads to a high incidence of large and false treatment effects under either scenario.

In Section 6, we focus on true treatment effects. These also have the functional form of the Diebold–Li curve, but are smaller in magnitude than the idiosyncratic effects. They can be either (i) a twist that pushes yields up by 6 bps at the one-year maturity, declining to −6 bps at ten years; or (ii) a short-end effect that pushes yields down by 6 bps at the one-year maturity, fading to zero at around seven years.[7] We show that this leads to uninformative estimated DiD coefficients under the classical DiD specification. Section 7 combines the idiosyncratic and true treatment effects for a total of four scenarios. This just compounds the problems from when the effects are considered separately. To simulate data, we combine these effects and the underlying baseline spot curve with randomly drawn residual maturities for treated bonds and controls as described next.

[5] As explained by Diebold and Li (2006) their specification suffers less from multicollinearity between the parameters as compared to the original Nelson and Siegel (1987) specification.

[6] For simplicity, we keep the decay parameter fixed at $\lambda_t = \lambda = 0.7308$ across scenarios. As explained by Diebold and Li (2006), λ_t determines the point where the loading on the curvature factor, $\gamma_{2,t}$, obtains its maximum. Based on practice, they pick this to be at a maturity of 30 months and set $\lambda_t = \lambda = 0.0609$. This translates to $\lambda = 0.7308$ when maturity is measured in years as here. The value of the decay parameter is not critical for our purposes.

[7] Treatment effects in fixed-income settings are typically relatively small and may vary over the maturity spectrum (see, e.g., Swanson, 2011). The twist treatment scenario is to some extent motivated by the Fed's "Operation Twist" in 1961 and "QE2" initiated in 2010 which involved selling short maturity Treasuries and buying longer ones. Although the intention of these programs was to lower longer-term yields without moving shorter ones (Alon and Swanson, 2011), more generally, if investors shift habitats, this can cause yields to move in opposite directions in different parts of the maturity spectrum (Vayanos and Vila, 2021). In practice, short and long rates often move in opposite directions (Table 1, Panel C). Events that change the relative attractiveness of long and short bonds can give rise to twists in treatment effects.

Table 3
The maturity structure of outstanding securities in practice.
For the same selection of countries as in Table 1, this table provides an overview of the distribution of the outstanding securities per country at the beginning of 2023. The first column shows the total number of securities with market prices in Bloomberg. The columns to the right provide the number of securities by maturity bucket (as indicated) as a percentage of that country's total number of securities.
Data source: Bloomberg.

Country	Number of securities	[0–2]	(2–5]	(5–10]	(10–15]	(15–20]	(20–30]	>30y
Ireland	18	11	17	39	17	6	11	0
Portugal	22	27	18	27	14	5	9	0
Greece	47	45	23	11	15	4	2	0
Netherlands	50	52	10	18	8	6	4	2
France	76	42	16	17	4	7	8	7
Germany	147	31	18	22	12	8	8	1
U.K.	225	23	12	15	10	15	17	8
Spain	240	16	16	23	11	9	12	13
Italy	285	32	15	18	10	6	9	10
China	610	37	21	15	3	5	10	9
U.S.	674	34	21	13	4	10	18	0
Japan	1137	16	14	18	12	12	19	7

3.2. Simulation of residual maturities

In practice, the distributions of same-issuer bonds are often tilted toward particular segments of the maturity spectrum. As an illustration, for the same countries as in Table 1, Table 3 provides the number of government bonds with market prices in Bloomberg and their distributions across maturity buckets at the beginning of 2023. We see that Dutch issues, for example, are tilted toward the short end; more than fifty percent have maturities of less than two years, with the remaining bonds being spread out at a declining rate after the five to ten year segment. The distribution for Spain, in contrast, is tilted toward the mid range of the maturity spectrum. Within rating categories, corporate bonds are also unevenly distributed across maturities (S&P Global, 2023).

In analyzing the performance of the classical DiD specification in Eq. (1), we wish to use reasonably realistic maturity distributions. To this end, we draw residual maturities for treated bonds and controls from triangular probability density functions (pdf's), $p(x;m)$, where $x \in [0,20]$ is time to maturity measured in years and m is the mode. Specifically,

$$p(x;m) = \begin{cases} 0, & \text{if } x < 0 \text{ or } x > 20, \\ \frac{x}{10m}, & \text{if } 0 \leq x \leq m, \\ \frac{20-x}{10(20-m)}, & \text{if } m < x \leq 20. \end{cases} \tag{3}$$

Relative to what we see in Table 3 for government bonds, the upper limit on residual maturity of twenty years is conservative, but is sufficiently large to generate meaningful term effects. The mode, m, parameterizes a feature of real data that we want to capture, namely, the main location of bonds in terms of residual maturity. Changing m allows us to create heterogeneity in the distributions of residual maturities between treated bonds and controls in a straightforward way. For controls, we use $m = 0.25$ years and for treated bonds $m = 0.25, 1, 3$ or 10 years.[8] The distributions are illustrated for different m's in Fig. 1.

Because there are no coupons, bonds are defined by their residual maturities. The data is simulated by independently repeating the following 1000 times:

1. Control bonds: Draw fifty maturities independently from $p(x;m = 0.25)$.
2. Treated bonds: Draw fifty maturities independently from $p(x;m)$, $m = 0.25, 1, 3, 10$.

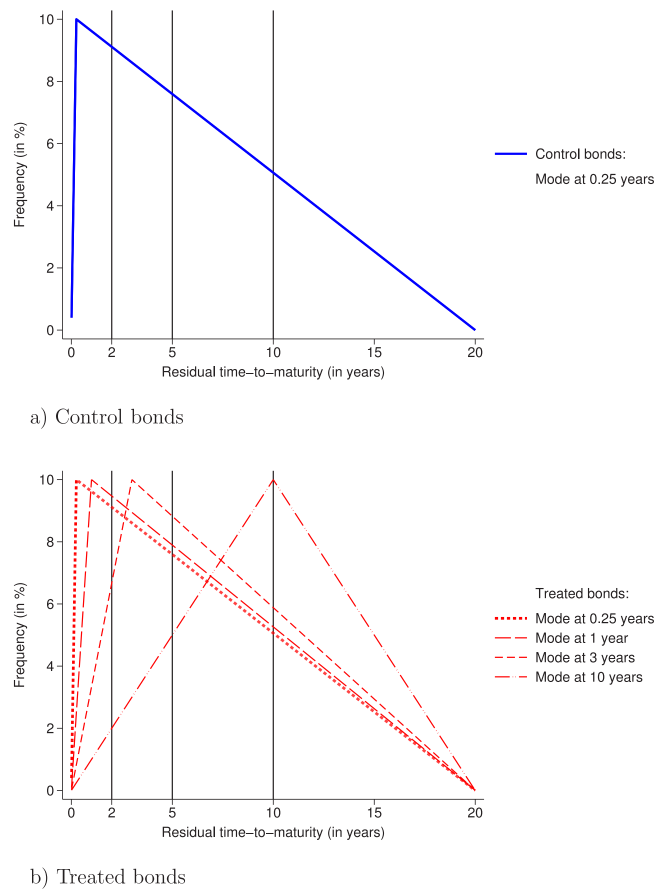


Fig. 1. Triangular probability density functions (pdfs) with different modes *m*.
This figure shows the triangular pdf's used to simulate residual maturity of the one control bond sample with mode $m = 0.25$ years and the four samples of treated bonds with modes $m = 0.25, 1, 3$, and 10 years while residual maturity x ranges from zero to twenty years ($x \in [0,20]$) for either sample. The vertical lines mark the cutoff points in the process of building maturity buckets, namely 2, 5, and 10 years (discussed in later sections of the paper).

This yields 1000 families of simulated bonds. Each family is comprised of fifty controls and four times fifty treated bonds. For each family, or sample draw, we create four sample couplets by combining the fifty control bonds, where $m = 0.25$, with fifty treated bonds, for each of $m = 0.25, 1, 3$, or 10. Thus, we generate 4000 individual samples, each comprised of fifty treated bonds and fifty controls. Because we hold the control bonds constant within each of the 1000 families, this setup

[8] Keeping m for controls fixed reduces the number of cases without affecting the thrust of our results.

Table 4
Overview of simulated residual maturity data.
This table provides an overview of the simulated residual maturity data. We draw 1000 families of samples. Each family is comprised of one control-bond sample and four treated-bond samples. Each sample contains fifty bonds. The residual maturities of the fifty control bonds are drawn from a triangular pdf that ranges from zero to twenty years and has mode $m = 0.25$ years. The residual maturities of the four times fifty treated bonds are drawn from triangular pdfs covering the same range but with modes $m = 0.25, 1, 3, 10$. First, we calculate the average residual maturity for each sample. Panel A shows the distributions of these sample averages separately by treatment group and triangular pdf mode m across the 1000 sample draws. Second, within each family, we build four sample couplets by pairing each sample of treated bonds ($m = 0.25, 1, 3, 10$) with the sample of control bonds ($m = 0.25$). Each sample couplet contains one hundred bonds, namely fifty treated and fifty control bonds. We calculate the average residual maturity of the fifty treated bonds and divide it by the average residual maturity of the fifty control bonds. Panel B shows the distributions of these maturity ratios across the 1000 sample draws, or families, separately for each mode $m = 0.25, 1, 3, 10$ of the treated bonds ($m = 0.25$ for control bonds always). Panel C shows the distributions of the correlation between residual maturity and the treatment assignment across the same 1000 sample draws and modes m of the treated bonds.

Panel A: Distributions of average residual maturity by treatment group and triangular pdf mode

			Distributions of sample averages						
			Number of						
Group	m	unconditional mean	samples	bonds per sample	Mean	SD	Med	Min	Max
Control	0.25	6.75	1,000	50	6.764	0.668	6.765	4.327	8.807
Treated	0.25	6.75	1,000	50	6.734	0.674	6.707	4.322	9.168
	1	7.00	1,000	50	7.034	0.611	7.029	5.361	9.008
	3	7.67	1,000	50	7.648	0.604	7.631	5.819	9.558
	10	10.00	1,000	50	9.979	0.587	10.015	8.043	11.981

Panel B: Distributions of maturity ratios by triangular pdf mode of the treated bonds

		Distributions of ratios of sample averages						
		Number of						
m treated bonds	unconditional means	sample couplets	bonds per sample couplet	Mean	SD	Med	Min	Max
0.25	1.00	1,000	100	1.005	0.138	0.994	0.592	1.524
1	1.04	1,000	100	1.050	0.141	1.039	0.710	1.598
3	1.14	1,000	100	1.142	0.146	1.135	0.691	1.805
10	1.48	1,000	100	1.490	0.177	1.481	1.095	2.231

Panel C: Distributions of correlation between residual maturity and the treatment assignment

		Distributions of correlations						
		Number of						
m treated bonds	unconditional means	sample couplets	bonds per sample couplet	Mean	SD	Med	Min	Max
0.25		1,000	100	−0.004	0.097	−0.004	−0.349	0.318
1		1,000	100	0.029	0.098	0.028	−0.268	0.322
3		1,000	100	0.097	0.098	0.100	−0.301	0.371
10		1,000	100	0.344	0.091	0.349	0.077	0.627

allows us to study the impact of diverging distributions of residual maturities between treated bonds and controls.

The simulated data have the realistic feature that residual maturities are not the same for treated and control bonds. Even if treated bonds are drawn from the same distribution as the controls ($m = 0.25$), their realized maturities differ almost surely. We are interested in examining the performance of the classical and other DiD specifications when underlying distributions diverge. We do this by varying the parameter, m, for treated bonds.

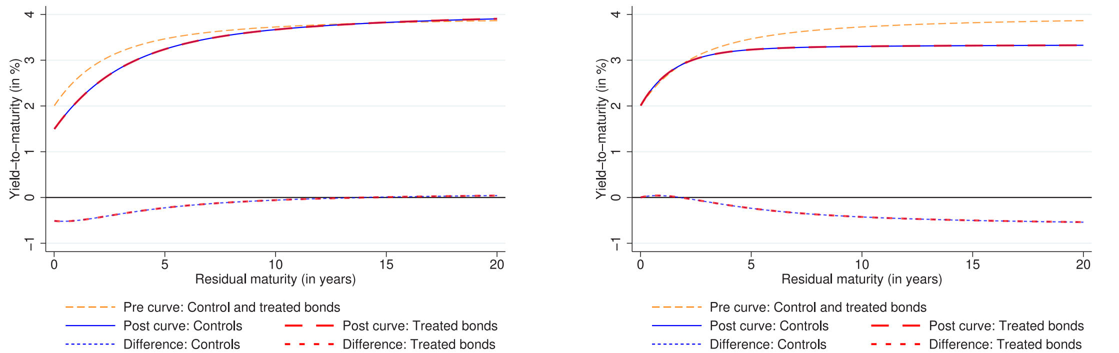
Table 4 presents an overview of the simulated data. Panel A provides summary statistics on average residual maturities across the 1000 control-bond samples and, for each m, the 1000 treated-bond samples. Not surprisingly, in each case, the average and median of the 1000 averages are close to the unconditional mean. However, because each sample comprises only fifty bonds, there is substantial variation across them. For example, while the unconditional mean for the control bonds ($m = 0.25$) is 6.75 years, the distribution of sample averages ranges from 4.327 to 8.807 years.

Panel B provides statistics on the ratios of average maturities across the 4000 sample couplets broken down by the parameter m for treated bonds. For an individual sample couplet, the maturity ratio is the average maturity of treated bonds divided by the average maturity of controls. The unconditional mean of these maturity ratios ranges from 1.00 (when m for treated bonds is 0.25) to 1.48 (when m for treated bonds is 10). For each treated-bond m, mean and median sample maturity ratios are close to the unconditional means. However, there is substantial variation across samples. For example, for $m = 0.25$, the maturity ratio ranges from 0.59 to 1.52 across the 1000 sample couplets. Thus, even when the underlying distributions of residual maturities are identical for treated and control bonds, the differences in residual maturities can be very large in individual samples. This leads to a critical property of fixed-income data from a DiD perspective, namely, nonzero correlation between residual maturity and treatment assignment.

Panel C reports on the distributions of these maturity-treatment correlations. The magnitude of the correlation can be large even in the case that $m = 0.25$ for treated bonds, where the correlation ranges from −0.349 to 0.318, albeit with an average of approximately zero. This illustrates that drawing relatively small samples from wide maturity ranges can induce spurious correlation between residual maturity and the treatment assignment. The situation is worse when the underlying distributions of residual maturities diverge. As m for treated bonds increases, the average maturity-treatment correlation becomes positive across samples and reaches 0.627 in one sample. In short, spurious correlation is compounded by a bias. In the rest of the paper, we examine how heterogeneous term effects and nonzero maturity-treatment correlation give rise to false and garbled treatment effects in the classical DiD specification and what to do about it.

4. False treatment effects

In this section, we run the classical DiD specification in Eq. (1) on simulated data with idiosyncratic, treatment-unrelated effects only and



a) Idiosyncratic yield-curve short-end effect:

i) Parameter values	γ_0	γ_1	γ_2	λ
Pre-curve	4.000	-2.000	0.000	0.7308
Post-curve	4.140	-2.650	-0.800	0.7308

ii) Yields (in %) and differences (in pps) at selected maturities:

	1y	2y	3y	5y	7y	10y	15y
Pre-curve	2.58	2.95	3.19	3.47	3.61	3.73	3.82
Post-curve	2.08	2.51	2.83	3.24	3.47	3.67	3.83
Difference	-0.50	-0.44	-0.36	-0.23	-0.14	-0.06	0.01

b) Idiosyncratic yield-curve long-end effect:

i) Parameter values	γ_0	γ_1	γ_2	λ
Pre-curve	4.000	-2.000	0.000	0.7308
Post-curve	3.350	-1.350	1.000	0.7308

ii) Yields (in %) and differences (in pps) at selected maturities:

	1y	2y	3y	5y	7y	10y	15y
Pre-curve	2.58	2.95	3.19	3.47	3.61	3.73	3.82
Post-curve	2.62	2.93	3.10	3.23	3.28	3.30	3.32
Difference	0.04	-0.01	-0.09	-0.24	-0.34	-0.43	-0.50

Fig. 2. Modeling idiosyncratic effects in the term structure of interest rates.
To model the term structure we employ Diebold and Li's (2006) yield-curve specification. The mini table underneath each plot shows the parameter values to create the true underlying term structures as well as the resulting yield levels and yield differences at selected maturities. Figs. 2a and 2b provide graphical illustrations of the resulting yield and differences curves when there is an idiosyncratic short-end or a long-end effect, respectively, from pre- to post-treatment.

show that this gives rise to a problem of false treatment effects. Initially, we do not include controls for maturity (**Z** is empty). For simplicity, there are only two time periods, labeled pre-treatment ($t = 0$) and post-treatment ($t = 1$). Treatment is to be understood as hypothesized by a researcher. Prior to "treatment", "treated" bonds and controls share the same yield curve. Since there is no treatment effect, they also share the same curve after treatment, but the idiosyncratic effect changes the shape and location of the curve.

For each sample couplet of control and treated bonds (residual maturities), yields are generated by the Diebold and Li (2006) curve in Eq. (2). We consider two scenarios, as shown in Fig. 2. The pre-event curve is the same in both scenarios, but the idiosyncratic effects that move the post-event curve either depresses the short end (Panel A) or the long end (Panel B). The parameter values for the curves are provided in the two panels. As shown in the figure, effects are heterogeneous over the maturity spectrum. The short-end idiosyncratic effect pushes yields down by 50 bps at the one-year maturity, fading to a 1 bps drop at fifteen years. The long-end effect reverses this (approximately).[9]

4.1. Estimation with the classical DiD specification

We estimate Specification (1) without controls using ordinary least squares (OLS) for each of our 8000 simulated datasets (sample couplets) of treated and control bonds (4000 for each type of idiosyncratic effect, see Section 3). Standard errors are formally clustered at the bond level.[10] Fig. 3 plots the 8000 DiD estimates against their t-statistics. Panel A of the table within the figure provides summary statistics. The 8000 DiD estimates range from −24.06 to + 22.99 bps and 3067 of them are statistically different from zero at the 10%-level. In short, the state of the art classical DiD specification produces potentially large treatment effects that are statistically significant at conventional levels *even though there is no true treatment effect.* In fact, in these simulations, there are such false treatment effects in 38.33% of all cases.

Panel B in the table within Fig. 3 provides summary statistics of the DiD estimates separately for each type of idiosyncratic effect, broken down by the mode parameter, m, of the treated bonds ($m = 0.25$ for controls, see Section 3). The incidence of false treatment effects is similar for the two idiosyncratic effects and increasing as the underlying distribution of residual maturities for treated bonds diverges from that of the control bonds (m for treated bonds increases). When the underlying distributions are the same ($m = 0.25$), the incidence of false treatment effects is approximately 10%, and the average DiD estimate is close to zero. However, in individual samples, the DiD estimate can reach economically large values, ranging from −11.59 to 12.01 bps across all samples and the two types of idiosyncratic effects.

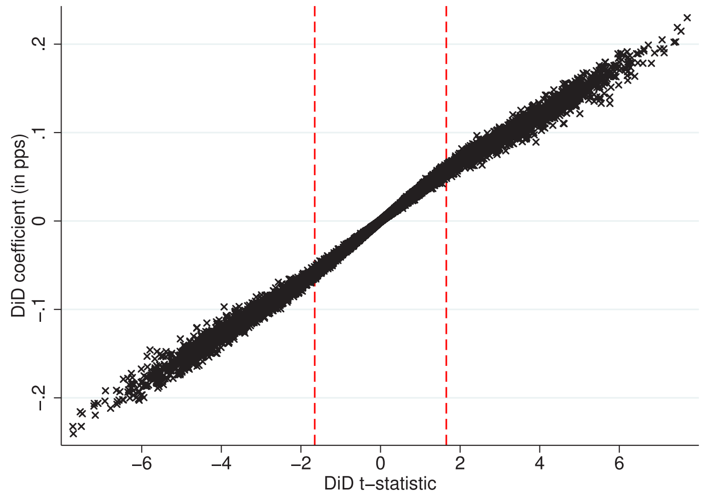
As the underlying distributions of treated- and control-bond residual maturities diverge (m for treated bonds increases), the basic problem of imprecision due to sampling disparity is compounded by a bias. This can be seen by the fact that the average DiD estimate differs from zero, being positive when the idiosyncratic effect predominantly hits the short end and negative when it predominantly hits the long end. As a result, the incidence of false treatment effects increases. For instance, if $m = 10$ for treated bonds and in case of an idiosyncratic short-end (long-end) effect, all 1000 DiD estimates are positive (negative) and more than 99% of them are statistically significant at the 10%-level. Thus, different underlying distributions of treated- and control-bond residual maturities can seriously garble inference from the classical DiD specification when yield curves move around for reasons unrelated to hypothesized treatment.

4.2. The false treatment effect mechanism

False treatment effects result from a combination of idiosyncratic yield-curve effects that are heterogeneous over the maturity spectrum and nonzero correlation between residual maturity and the treatment assignment (see Section 3). Thus, as control and treated bond maturity distributions diverge and idiosyncratic effects increase, for example due

[9] The curves in Fig. 2 slope upwards, but our arguments do not depend on this. False treatment effects can also be shown for downward sloping or flat curves.

[10] Bertrand et al. (2004) show that the persistence of the treatment indicator in DiD settings induces serial correlation in the error term and that clustering at the level of the treated unit helps to diminish this issue (see also Petersen, 2009). Although there are no such correlated errors in our simulated setting, we nevertheless ask the software to cluster at the bond level. Being able to cluster standard errors is an attractive feature of Specification (1).



mode m	Mean	SD	Med	Min	Max	No. of coefficients All	No. of coefficients $\|t\| > 1.653$
Panel A: Both effects and all m (0.25, 1, 3, 10 years) combined							
–	-0.10	7.51	0.05	-24.06	22.99	8,000	3,067
Panel B: Short-end effect by mode m (0.25, 1, 3, 10 years)							
0.25 years	-0.15	3.40	-0.14	-11.59	10.85	1,000	91
1 year	1.25	3.39	1.17	-8.67	11.85	1,000	115
3 years	4.27	3.24	4.30	-8.73	14.24	1,000	354
10 years	12.24	3.04	12.26	2.30	22.99	1,000	992
Panel C: Long-end effect by mode m (0.25, 1, 3, 10 years)							
0.25 years	0.15	3.60	0.13	-11.34	12.01	1,000	88
1 year	-1.20	3.60	-1.10	-12.45	9.31	1,000	111
3 years	-4.31	3.46	-4.35	-14.95	9.72	1,000	325
10 years	-13.05	3.25	-13.08	-24.06	-2.59	1,000	991
Panel D: Correlation between residual maturity and treatment assignment							
0.25 years	-0.004	0.097	-0.004	-0.349	0.318	1,000	Not applicable
1 year	0.029	0.098	0.028	-0.268	0.322	1,000	
3 years	0.097	0.098	0.100	-0.301	0.371	1,000	
10 years	0.344	0.091	0.349	0.077	0.627	1,000	

Fig. 3. False treatment effects graphically.
This figure shows estimated treatment effects when the modeled term structure exhibits heterogeneous idiosyncratic effects along the maturity dimension but there is no true treatment effect present in the data. The specification is $yield_{it} = \alpha_i + \delta_t + \beta_{DiD}\ \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \epsilon_{it}$, where $yield_{it}$ is the yield-to-maturity of bond i at time t, the α_i's (δ_t's) are bond (time) fixed effects, $\mathbb{1}_{Treated,i}$ ($\mathbb{1}_{Post,t}$) is a treatment (event and post-event dates) indicator variable, β_{DiD} the treatment effect, and ϵ_{it} the error term. The specification is estimated with OLS. The (black) crosses show the 8000 estimated DiD coefficients plotted against their t-statistics across the two idiosyncratic effects, the four modes m of treated bonds, and the 1000 families of sample couplets. The vertical dashed (red) lines mark the values of ± 1.653 (two-sided confidence bands using 10%-significance level). The distributions of the corresponding numbers are in Panel A in the table underneath. Panel B (C) shows the DiD coefficient distributions for the idiosyncratic short-end (long-end) effect by mode m. The t-statistics are based on standard errors clustered at the bond level. Panel D shows the distributions of the correlation between residual maturity and the treatment assignment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to long estimation windows, large false treatment effects arise naturally under the classical DiD specification. The sign of the estimated effect is a function of the treatment-maturity correlation and the idiosyncratic effect. When the correlation is positive (negative), an idiosyncratic effect that depresses the short end of the term structure results in a positive (negative) DiD coefficient. Signs are reversed under an idiosyncratic effect that depresses the long-end. As seen in Panel B in the table within Fig. 3, when $m = 10$ for treated bonds, all DiD estimates are positive (negative) under the short-end (long-end) effect. This is because the treatment-maturity correlation is positive in all samples. Internet Appendix A.2 provides further details.

The analysis illustrates that the classical DiD specification in Eq. (1) applied to variables that exhibit time-varying term structures is prone to producing false treatment effects. Even in the absence of a true treatment effect, the specification generates statistically significant, but false, treatment effects that can be economically large and go in either direction. The fundamental problem is that the classical DiD specification is misspecified.

4.3. Estimation separately by individual maturity buckets

To deal with heterogeneous treatment effects, the literature typically uses fixed effects on the discrete units present in the data. For example, in Specification (1), the individual unit is each bond. However, as we have started to show, this is a misspecification when there are heterogeneous term effects which can lead to a potentially high incidence of false positives. A workaround that is sometimes employed is to run the classical DiD specification separately on individual maturity buckets (Bao et al., 2018; Todorov, 2020). In this subsection, we address this approach using the same simulated data as above.

Table 5 shows the results. Bonds are broken down into four maturity buckets, namely, $(0,2]$, $(2,5]$, $(5,10]$, and $(10,20]$ years. Panels A and

Table 5
False treatment effects measured individually by maturity buckets.
This table provides the distributions of estimated treatment effects using OLS on the same data, the same modeled idiosyncratic yield-curve effects, and using the classical DiD specification as in Fig. 3 but separately by four individual buckets with residual maturity in the ranges $(0,2]$, $(2,5]$, $(5,10]$, and $(10,20]$ years. Panel A (B) covers the case when residual maturity of the treated bonds is drawn from a triangular pdf with mode $m = 0.25$ ($m = 10$) years. Each panel shows mean, median, and minimum as well as maximum of the distributions of the estimated treatment effects by the idiosyncratic effects (short- or long-end) and the maturity buckets as well as separately for the cases when $t < -t_{cv}$, $-t_{cv} \le t \le t_{cv}$, and $t_{cv} < t$, where t_{cv} is the critical value of a two-sided t-test at the significance level of 10% (which is 1.645 in case of a z-test) with standard errors clustered at the bond level.

Idiosyncratic effect	Maturity bucket (in years)	Number of measured effects, N	Mean (in bps)	Med (in bps)	$t < -t_{cv}$ N	Min (in bps)	Max (in bps)	$-t_{cv} \le t \le t_{cv}$ N	Min (in bps)	Max (in bps)	$t_{cv} < t$ N	Min (in bps)	Max (in bps)
Panel A: m = 0.25 years													
Short-end	(0–2]	1,000	0.01	0.02	60	−4.24	−1.49	883	−2.77	2.51	57	1.54	4.17
	(2–5]	1,000	−0.01	0.03	53	−7.49	−3.31	900	−4.69	4.73	47	3.55	8.47
	(5–10]	1,000	−0.09	−0.09	63	−6.14	−2.36	889	−3.43	3.46	48	2.39	5.54
	(10–20]	1,000	−0.01	−0.02	46	−3.44	−1.15	900	−2.18	2.11	54	1.31	3.10
Long-end	(0–2]	1,000	−0.01	−0.01	61	−2.85	−0.75	868	−1.97	1.92	71	0.74	3.17
	(2–5]	1,000	0.01	−0.03	47	−8.94	−3.74	900	−5.03	4.97	53	3.50	7.90
	(5–10]	1,000	0.10	0.10	48	−6.24	−2.70	888	−3.87	3.66	64	2.65	6.90
	(10–20]	1,000	0.01	0.02	54	−3.53	−1.49	900	−2.39	2.48	46	1.30	3.92
Panel B: m = 10 years													
Short-end	(0–2]	640	1.13	1.03	94	−5.12	−1.14	308	−2.45	3.90	238	0.58	8.01
	(2–5]	998	1.77	1.77	25	−10.51	−3.32	762	−6.29	6.15	211	3.26	11.89
	(5–10]	1,000	1.37	1.45	11	−4.14	−2.08	787	−2.61	3.11	202	2.03	6.70
	(10–20]	1,000	0.00	−0.01	49	−2.98	−1.18	888	−1.81	1.64	63	1.28	3.15
Long-end	(0–2]	640	−0.48	−0.22	204	−4.82	−0.43	288	−2.24	1.58	148	0.46	2.88
	(2–5]	998	−1.87	−1.84	210	−12.70	−3.35	762	−6.60	6.60	26	3.58	11.01
	(5–10]	1,000	−1.54	−1.63	202	−7.56	−2.28	787	−3.50	2.94	11	2.35	4.67
	(10–20]	1,000	−0.00	0.01	62	−3.58	−1.46	889	−1.87	2.06	49	1.34	3.39

B cover the cases when m for treated-bond maturities equals 0.25 and 10 years, respectively. Each panel provides mean and median estimated DiD coefficients across the 1000 samples by maturity bucket for each of the two types of idiosyncratic effects used above (see Fig. 2). They also show the ranges and number of DiD coefficients that are statistically significantly negative, positive, or not statistically significant (10%-level).

The results show that the maturity-bucket approach does not eliminate false treatment effects. For example, in the $(2,5]$ year bucket in Panel A ($m = 0.25$) under the short-end idiosyncratic effect, there are 53 negative and 47 positive statistically significant DiD coefficients that range from −7.49 to −3.31 and from +3.55 to +8.47 bps, respectively. The remaining 900 coefficients are also different from zero but not statistically significant. The incidence of false positives increases when the underlying distributions diverge. For example, in Panel B ($m = 10$), the number of statistically significant DiD coefficients increases to 236 for the $(2,5]$ year bucket. DiD coefficient point estimates are also more extreme than in Panel A. The mechanism is the same as before, namely, nonzero correlations between treatment assignment and residual maturity. As seen in the table and as discussed above, the sign of the estimation error depends on the nature of the idiosyncratic effect.

For research that uses the maturity bucket approach to study potential treatment effects across the maturity spectrum, these results point to a concern. If the sample results in false treatment effects for some maturity buckets, but not others, the maturity-bucket approach leads to false conclusions about variation in treatment effects over the maturity spectrum. Different researchers working with different samples may find false treatment effects in different parts of the term structure, leading to ambiguous conclusions overall.

Another problem with the maturity-bucket approach is that it can be difficult to implement due to a paucity of bonds and large differences in the underlying, unconditional distributions of residual maturities between treated and control bonds. This issue arises in Panel B ($m = 10$), where, for example, only 640 of the 1000 samples have sufficient treated and control bonds in the $(0,2]$ year bucket to run Specification (1).

In conclusion, the maturity-bucket approach pushes the misspecification problem in the classical DiD specification to the maturity-bucket level. In turn, this gives rise to a new potential problem, namely falsely measured term effects, with false variation across the maturity spectrum. Using a fine maturity-bucket mesh to deal with this is typically impractical due to a paucity of bonds.

4.4. The failure of maturity control

Researchers sometimes use maturity controls in Specification (1). It is interesting to ask whether this helps performance. To address this, we first consider an alternative specification where the bond fixed effect is replaced by an explicit parametric control for the term structure. To give this approach as good a chance as possible to succeed, we use the same functional form for the yield curve as used to generate the data. The question is whether such explicit term-structure control enables a DiD specification to elicit the true underlying effects or, at least, reduce the incidence of false treatment effects. Specifically, we estimate the following using nonlinear least squares (NLS):

$$yield_{it} = \mathbf{B}'\mathbf{L}_{it} + \alpha\ \mathbb{1}_{Treated,i} + \delta\ \mathbb{1}_{Post,t} + \beta_{DiD}\ \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \varepsilon_{it}, \quad (4)$$

where α and δ are coefficients on the indicator variables for treated bonds and the post-event period, respectively, and $\mathbf{B}'\mathbf{L}_{it}$ is a yield curve as in Eq. (2). Specifically, $\mathbf{L}_{it}$ is a three-dimensional vector, $(1, l_1(x_{it};\lambda), l_2(x_{it};\lambda))$, where

$$l_1(x;\lambda) = \frac{1-e^{-\lambda x}}{\lambda x} \quad \text{and} \quad l_2(x;\lambda) = \frac{1-e^{-\lambda x}}{\lambda x} - e^{-\lambda x}, \quad (5)$$

$\mathbf{B}$ is the corresponding vector of coefficients with individual elements β_k, $k = 0,\dots,2$, and the decay parameter, λ, is assumed to be time invariant. As a starting value, we take $\lambda_{Seed} = 1$. Standard errors are clustered at the bond level.

The results turn out to be practically identical to those under the classical DiD specification. Table 6 provides summary statistics on the differences in DiD coefficients and their p-values between the two specifications (classical minus yield-curve control) for the same sample couplets and idiosyncratic effects. The 8000 DiD-coefficient differences all equal 0.000 bps and the corresponding p-value differences range from −0.005 to −0.000. In short, the yield-curve control approach is just as misspecified as the classical DiD approach.

Table 6
Performance of specification with straight yield-curve control.
This table shows summary statistics on the differences in DiD coefficients and their p-values between the classical DiD specification (see Fig. 3) and the specification with straight yield-curve control, namely $yield_{it} = \mathbf{B}'\mathbf{L}_{it} + \alpha\ \mathbb{1}_{Treated,i} + \delta\ \mathbb{1}_{Post,t} + \beta_{DiD}\ \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \varepsilon_{it}$. α and δ are coefficients on the indicator variables for treated bonds and the post-event period, respectively. $\mathbf{B}'\mathbf{L}_{it}$ is the yield curve. Specifically, $\mathbf{L}_{it}$ is a three-dimensional vector, $(1, l_1(x_{it};\lambda), l_2(x_{it};\lambda))$, where $l_{1,t}(x;\lambda) = \left(\frac{1-e^{-\lambda x}}{\lambda x}\right)$ and $l_{2,t}(x;\lambda) = \left(\frac{1-e^{-\lambda x}}{\lambda x} - e^{-\lambda x}\right)$. $\mathbf{B}$ is the corresponding vector of coefficients with individual elements β_k, $k = 0, \ldots, 2$, and the decay parameter, λ, is assumed to be time invariant. The classical DiD specification is estimated with OLS. The specification with straight yield-curve control is estimated with NLS with starting value $\lambda_{Seed} = 1$. In either case, standard errors are clustered at the bond level. The differences are calculated as statistic of the classical DiD specification minus the same statistic of the specification with straight yield-curve control. The underlying data are the 8000 sample couplets comprised of 1000 sample draws, four modes m of the treated bonds, and the two idiosyncratic yield-curve effects at short- and long-end, respectively.

	Mean	SD	Med	Min	Max	N
Differences in DiD coefficients (in bps)	0.000	0.000	0.000	−0.000	0.000	8,000
Differences in DiD p-values	−0.002	0.002	−0.003	−0.005	−0.000	8,000

Table 7
Different maturity controls in the control vector Z.
This table shows estimated treatment effects using a simulated dataset with a heterogeneous idiosyncratic yield-curve short-end effect and m equals one. There is no true treatment effect. The time difference between the pre and post periods is set as equivalent to one month. For different maturity controls in the vector $\mathbf{Z}$, as indicated, the table shows DiD effects estimated with Specification (1) and OLS and, underneath in parentheses, the p-value based on standard errors clustered at the bond level. a, b, and c denote significance (two-sided) at the levels of 1%, 5%, and 10%, respectively. DiD effects statistically significant at the 10%-level or stronger are marked in bold.

	Specification (1): Maturity control in **Z**			
	None	Maturity x	Diebold–Li yield curve	Logarithm of maturity x
$Treated \times Post$	−3.34	−3.34	**0.49**[c]	**−5.61**[c]
	(0.40)	(0.40)	(0.05)	(0.08)
x		*omitted*		
l_1			−4516.86[a]	
			(0.00)	
l_2			−2721.68[a]	
			(0.00)	
$\ln(x)$				157.26[a]
				(0.01)
N	200	200	200	200
R-squared adjusted	0.5908	0.5908	0.9983	0.7184

Since we use the same yield-curve specification in the estimation that is used to generate the data, it is clear that the problem is not with the yield-curve specification itself, but with how it is incorporated into the regression equation. Since $\mathbf{B}'\mathbf{L}_{it}$ just removes the average term structure in the pooled data (treated, controls, pre- and post-treatment), Specification (4) restricts yield-curve movements between the groups to parallel shifts. While this feature is explicit in Eq. (4), the classical DiD specification imposes the same restrictions implicitly through the fixed bond effects. Thus, when idiosyncratic yield-curve effects are not homogeneous over maturity, false treatment effects arise equally under Specifications (1) and (4) through the mechanism of spurious and potentially systematic correlation between residual maturity and the treatment assignment.[11] The classical DiD specification and simple yield-curve control specifications are misspecified unless the parallel-shift assumption holds in the data. Because yield curves in practice change shape as time progresses, however, this is rarely, if ever, satisfied.

Adding residual maturity to Specification (1) does not do anything. Because residual maturity is linear in time, it is a linear combination of the bond- and time-fixed effects and, therefore, redundant. For this reason, the papers in Table 2 that use bond fixed effects and have maturity control in **Z** employ nonlinear transformations of residual maturity. However, as illustrated in the example in Table 7, this can make matters worse. The example runs Specification (1) without controls and with residual maturity, its log, or the Diebold–Li curve with the correct value for the decay parameter ($\lambda = 0.7308$) as controls on a simulated dataset with idiosyncratic effects only. The time difference between the pre and post periods is set as equivalent to one month. The results show that using a nonlinear transformation of residual maturity can increase the statistical significance of the DiD estimator even though, in fact, there is no true treatment effect. Remarkably, "controlling for residual maturity" can contribute to p-hacking. What is the solution? Logically, to deal with heterogeneous effects over the maturity spectrum, we need a DiD approach that specifically allows for such effects.

5. A solution: Flexible yield-curve DiD specification

A problem with the classical DiD specification and extensions of it that include yield curve control variables is that they are prone to producing false treatment effects because they do not adequately address idiosyncratic, treatment-unrelated effects that vary over the maturity spectrum. In this section, we provide a solution to this problem that works well in samples with zero coupon bonds. We also discuss a more general approach in Section 9.

The solution considered here is the fully flexible DiD specification introduced by Nyborg and Woschitz (2021) which estimates the treatment effect as a curve. The specification also estimates a baseline curve and incremental curves for treated bonds and the post-event time period. In this paper, we use the Diebold–Li curve, but other functional forms can also be used. In our case, the specification is given by

$$yield_{it} = \mathbf{B}_1'\mathbf{L}_{it} + \mathbf{B}_2'\mathbf{L}_{it}\ \mathbb{1}_{Treated,i} + \mathbf{B}_3'\mathbf{L}_{it}\ \mathbb{1}_{Post,t} + \mathbf{B}_4'\mathbf{L}_{it}\ \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \varepsilon_{it}, \quad (6)$$

where notation is as above except that each of the four indicators (constant, $\mathbb{1}_{Treated,i}$, $\mathbb{1}_{Post,t}$, and $\mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t}$) has its own Diebold–Li

[11] With longer pre- and post-event periods, the classical DiD specification is worse because the bond fixed effect does not take into account that bonds' maturities change over the event window.

curve, $\mathbf{B}_j'\mathbf{L}_{it}$, with $j = 1, \dots, 4$ and three individual coefficients each, $\beta_{k,j}$, $k = 0, \dots, 2$. The decay parameter, λ, is assumed to be time-invariant and the same for treated and control bonds, which can easily be relaxed. Because it has no bond fixed effects, Specification (6) is substantially more parsimonious than the classical DiD specification.

The first curve, $j = 1$, represents the pre-treatment spot curve of control bonds and is given by

$$s(x;\hat{\lambda}) = \hat{\beta}_{0,1} + \hat{\beta}_{1,1}\ l_1(x;\hat{\lambda}) + \hat{\beta}_{2,1}\ l_2(x;\hat{\lambda}), \tag{7}$$

where $\{\hat{\beta}_{k,1}\}_{k=0}^{2}$ are the estimated regression coefficients, x is residual maturity, and l_1 and l_2 are as in Eq. (5) with λ replaced by $\hat{\lambda}$. The second and third curves, $j = 2, 3$, are defined similarly and represent incremental, or Delta, curves for treated bonds and the post-event period, respectively. The fourth curve, $j = 4$, is the object of interest, that is, the DiD Delta curve:

$$\Delta_4(x;\hat{\lambda}) = \hat{\beta}_{0,4} + \hat{\beta}_{1,4}\ l_1(x;\hat{\lambda}) + \hat{\beta}_{2,4}\ l_2(x;\hat{\lambda}), \tag{8}$$

where $\{\hat{\beta}_{k,4}\}_{k=0}^{2}$ are the estimated regression coefficients. The DiD Delta curve, $\Delta_4(x;\hat{\lambda})$, is a function of residual maturity, x, and thus returns the treatment effect at specific maturities. Treatment-unrelated idiosyncratic yield-curve effects over the event are captured by the post-event Delta curve, $\Delta_3(x;\hat{\lambda})$. Thus, unlike the classical DiD specification, such effects do not contaminate the DiD estimator. By including a separate curve for treated bonds, the specification allows a researcher to capture the incremental effect over the treatment event across the maturity spectrum.

We estimate the fully flexible yield-curve DiD specification in Eq. (6) with NLS using the same 8000 samples of treated bonds and controls as above. In this data, there is no true treatment effect, only idiosyncratic effects that either affect predominantly the short-end of the maturity spectrum or the long end. The decay parameter, λ, is estimated in-sample together with the other parameters using the starting value $\lambda_{Seed} = 1$.[12] Standard errors are clustered at the bond level. For each regression, the estimation gives twelve coefficients and one estimate for lambda. We use these coefficients to calculate treatment effects at selected maturities and compute standard errors using the delta method.[13]

The results are in Table 8. The left side of the table shows the true idiosyncratic effect at the test maturities. For each maturity and each treated-bond m, the right side provides the minimum and maximum estimated treatment effects across all samples. Panel A uses all 8000 datasets while Panel B filters out 2532 samples without treated and control bonds with maturities less than one year. We do this because curve fitting can be problematic without short-term bonds to anchor the curve.

Consider first the case that sample couplets are affected only by spurious correlation between residual maturity and treatment assignment (m for treated bonds equals 0.25). Across the seven selected maturities and the 2000 (1954) datasets in Panel A (B), which is a total of 14,000 (13,678) different cases, Specification (6) estimates treatment effects that are maximally ± 0.01 bps away from the true effect of zero. With such precision, the approximately ten percent of coefficients that are statistically significant are economically insignificant. This is a large improvement compared to the large false treatment effects produced by the classical DiD specification that range from −11.59 to 12.01 bps (Fig. 3).

As the m for treated bonds increases, residual maturities of treated bonds shift out. As discussed, this generates systematic correlation between residual maturity and treatment assignment. It also increases the incidence of samples without short-term bonds because the unconditional distribution of treated-bond maturities becomes tilted to the right relative to the distribution of control-bond maturities. Therefore, we first discuss Panel B, where samples without short-term bonds are excluded. The key result is that the incidence of false treatment effects is eliminated (the maximum treatment effect in absolute value is 0.01 bps). In short, as long as the data is sufficiently complete that curves can be fit with reasonable precision, the fully flexible DiD specification resolves the problem of false treatment effects we saw with the classical DiD specification, the maturity-bucket approach, and the straight yield-curve control approach.

Moreover, the results in Panel A show that even if the bond data is incomplete at the short-end, the fully flexible model can perform very well. For example, the case of $m = 10$ involves a heavy tilt of treated bonds toward the right relative to controls, which could induce curve mismeasurement. Yet, the falsely measured effects between −0.14 and +0.08 bps at the one-year maturity when $m = 10$ are tiny compared to the corresponding effects from the classical DiD specification that lie between −24.06 and +22.99 bps (see Fig. 3).

6. Garbled estimates of true treatment effects

In this section, we examine the performance of the classical DiD specification in Eq. (1) when there are true, maturity-dependent treatment effects. To isolate the estimation impact of heterogeneous treatment effects, we focus on the case without idiosyncratic effects. Thus, while the yield curve of treated bonds changes location and shape upon treatment, the control-bond curve does not move. We show that the classical DiD specification leads to estimates that can be uninformative and misleading.

We use the same sample couplets as above and yields continue to be generated by the Diebold and Li (2006) curve in Eq. (2). As illustrated graphically in Fig. 4, we consider two scenarios for the treatment effects. In the first scenario, the treatment twists the rates of short-term (long-term) bonds up (down). For example, the one-year rate moves up by 6 bps, and the ten-year rate moves down by the same amount. In the second scenario, the treatment predominantly affects short-term bonds, pushing rates down. The magnitude is 6 bps at the one-year maturity, fading to zero at 6.96 years. Parameter values for the pre-treatment curve (which also applies to controls and is the same as in Section 4) and the true treatment effects are provided in the within-figure table.

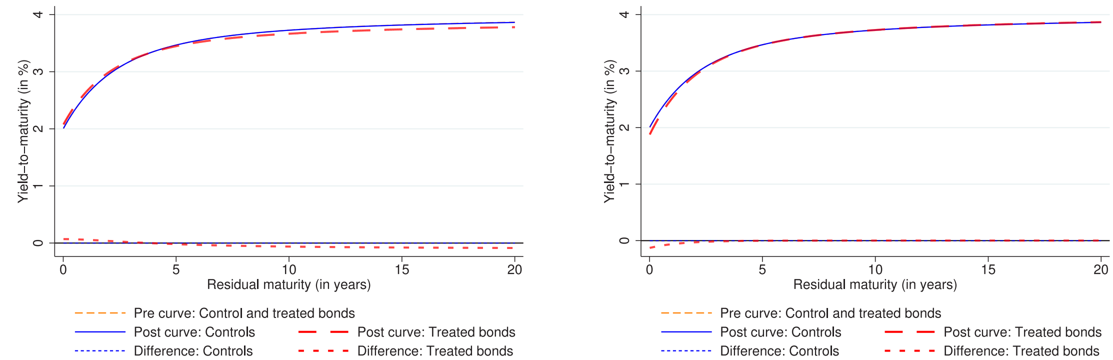
6.1. Estimation using the classical DiD specification

We run the classical DiD specification in Eq. (1) without maturity control using OLS. Standard errors are clustered at the bond level as before. The results are summarized in Figs. 5a and 5b for the twist and short-end scenarios, respectively. From left to right, the figure shows (i) the true maturity-dependent treatment effect, (ii) the distributions of the DiD estimates across samples when m for treated bonds equals 0.25 (in purple) and 10 years (in green; recall that $m = 0.25$ for controls), and (iii) the DiD estimates plotted against their t-statistics for the same m's (and using the same color scheme).

In the twist scenario, the figure shows that for the vast majority of samples, the classical DiD specification produces a statistically significantly negative DiD estimate. For $m = 0.25$, 879 out of 1000 DiD estimates are significantly negative (10% level). The other 121 estimates are insignificant. As a result, a researcher is likely to conclude that the treatment effect is negative *even though the true treatment effect is actually positive out to 3.82 years*, as seen in the figure. If the underlying distribution of treated-bond maturities shifts to the right, the results are even stronger. As seen in the figure, the DiD estimate is now significantly negative for all 1000 samples. This occurs because the DiD coefficient captures an average of the positive treatment effect on short-maturity bonds and the negative effect on longer bonds. As

[12] It is straightforward to extend this to having different λ's for treated and control bonds, pre- and post-treatment. One can also use NLS to estimate λ's first and then run OLS in a second step.

[13] See, for example, Casella and Berger (2001).



a) Yield-curve treatment twist:

i) Parameter values	γ_0	γ_1	γ_2	λ
Pre-curve treated	4.000	-2.000	0.000	0.7308
Post-curve treated	3.890	-1.820	0.180	0.7308

ii) Yields (in %) and differences (in pps) at selected maturities:

	1y	2y	3y	5y	7y	10y	15y
Pre-curve treated	2.58	2.95	3.19	3.47	3.61	3.73	3.82
Post-curve treated	2.64	2.99	3.21	3.45	3.57	3.67	3.74
Difference treated	0.06	0.04	0.02	-0.02	-0.04	-0.06	-0.08

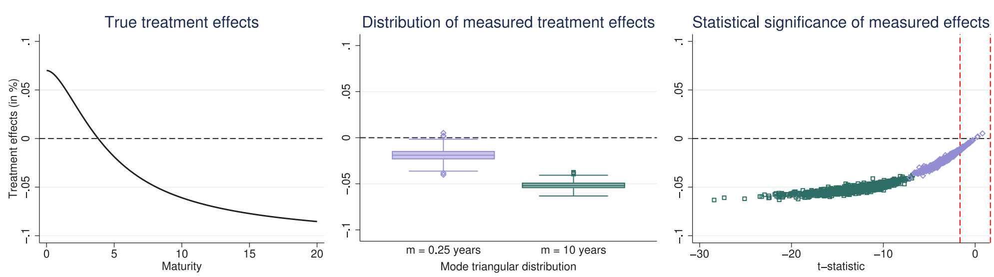
b) Yield-curve treatment short-end effect:

i) Parameter values	γ_0	γ_1	γ_2	λ
Pre-curve treated	4.000	-2.000	0.000	0.7308
Post-curve treated	4.001	-2.131	0.130	0.7308

ii) Yields (in %) and differences (in pps) at selected maturities:

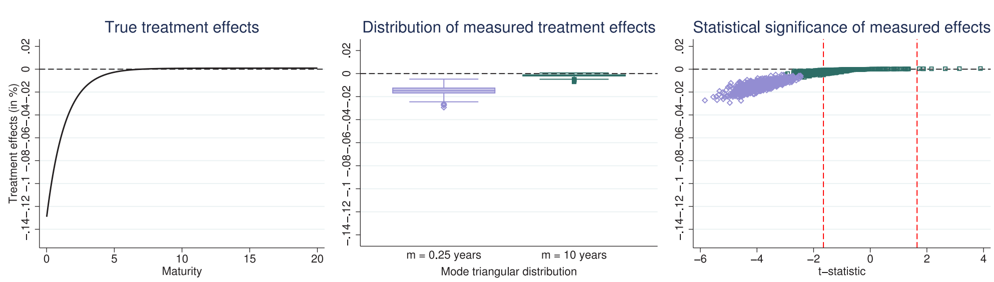
	1y	2y	3y	5y	7y	10y	15y
Pre-curve treated	2.58	2.95	3.19	3.47	3.61	3.73	3.82
Post-curve treated	2.52	2.92	3.18	3.46	3.61	3.73	3.82
Difference treated	-0.06	-0.03	-0.01	0.00	0.00	0.00	0.00

Fig. 4. Modeling term-structure treatment effects.
To model the term structure we employ Diebold and Li's (2006) specification. The mini table underneath each plot shows the parameter values to create the true underlying term structures as well as the resulting yield levels and yield differences at selected maturities. Figs. 4a and 4b provide graphical illustrations of the resulting yield and differences curves when there is a yield-curve treatment twist and a yield-curve treatment short-end effect, respectively, from pre- to post-treatment.



a) Treatment yield-curve twist

$m = 0.25$ years: $\widehat{\beta}_{DiD} \in [-4.02, +0.52]$ bps, $|t| > 1.653$: 879 $m = 10$ years: $\widehat{\beta}_{DiD} \in [-6.32, -3.75]$ bps, $|t| > 1.653$: 1000



b) Treatment short-end effect

$m = 0.25$ years: $\widehat{\beta}_{DiD} \in [-2.94, -0.48]$ bps, $|t| > 1.653$: 1000 $m = 10$ years: $\widehat{\beta}_{DiD} \in [-0.72, +0.06]$ bps, $|t| > 1.653$: 301

Fig. 5. Garbled treatment effects graphically.
Figs. 5a and 5b show true and measured treatment effects on the 1000 families of sample couplets for yield-curve treatment twist and treatment short-end effect, respectively, using OLS to estimate the same specification as in Fig. 3. From left to right, the graphs plot the true treatment effect over maturity, the distributions (box plots) if maturity of the treated bonds is drawn from triangular pdfs with $m = 0.25$ years (purple diamonds) or $m = 10$ years (green squares), and the estimated treatment effects against the t-statistics. The vertical dashed (red) lines in the plots to the far right mark the values of ± 1.653 (two-sided confidence bands using 10%-significance level). The t-statistics are based on standard errors clustered at the bond-level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the residual-maturity distribution shifts out, more weight is put on long bonds because their frequency in the sample increases. Unaware of the misspecification problem in Eq. (1) and the treatment twist, a researcher relying on the classical DiD specification would falsely believe that bond effects are adequately controlled for and erroneously conclude that treatment causes yields to fall regardless of residual maturity.

In the short-end scenario, treatment can go undetected if the distribution of residual maturities for treated bonds is tilted sufficiently away from the region where there is an effect. If $m = 10$ for treated bonds,

Table 8
The fully flexible yield-curve DiD specification to eliminate false treatment effects.
This table shows treatment effects estimated with the fully flexible yield-curve DiD specification $yield_{it} = \mathbf{B}_1' \mathbf{L}_{it} + \mathbf{B}_2' \mathbf{L}_{it} \mathbb{1}_{Treated,i} + \mathbf{B}_3' \mathbf{L}_{it} \mathbb{1}_{Post,t} + \mathbf{B}_4' \mathbf{L}_{it} \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \varepsilon_{it}$ with notation as in Fig. 3, $\mathbf{L}_{it}$ a three-dimensional vector of regressors with elements 1, $l_1(x_{it};\lambda)$, and $l_2(x_{it};\lambda)$, the latter two terms defined as in (5), and $\mathbf{B}_j$ the corresponding three-dimensional vectors of coefficients with individual elements $\beta_{k,j}$, $k = 0,\dots,2$. The latter measure level, slope, and curvature of the baseline curve for control bonds pre treatment ($j = 1$) and the incremental differences of (i) treated bonds pre treatment ($j = 2$), (ii) control bonds post treatment ($j = 3$), and (iii) treated bonds post treatment ($j = 4$). $\mathbf{B}_4$ captures level, slope, and curvature of the DiD Delta curve, $\Delta_4(x)$, which provides the treatment effects at maturity x. The specification is estimated with NLS, $\lambda_{Seed} = 1$, and λ is assumed to be time-invariant and the same for treated and control bonds. There are treatment-unrelated idiosyncratic yield-curve effects either at the short- or the long-end but the true, unconditional treatment effect is zero. At selected maturities, the table shows these true underlying effects and, to the right, the minimum and maximum of the estimated treatment effects across the two types of idiosyncratic yield-curve effects (at short- and long-end) and the 1000 families separately by m using the DiD Delta curve, $\Delta_4(x)$. Standard errors are clustered at the bond level and calculated using the delta method.

Residual maturity (in years)	Idiosyncratic yield-curve effects (in bps)		True treatment effect (in bps)	Distribution of estimated treatment effects (in bps)							
				$m = 0.25$		$m = 1$		$m = 3$		$m = 10$	
	Short-end	Long-end		Min	Max	Min	Max	Min	Max	Min	Max
Panel A: All sample couplets											
1	−50.35	3.91	0	−0.01	0.01	−0.01	0.01	−0.02	0.02	−0.14	0.08
2	−43.65	−1.47	0	−0.00	0.01	−0.01	0.01	−0.01	0.01	−0.06	0.03
3	−35.82	−9.31	0	−0.01	0.00	−0.01	0.00	−0.00	0.00	−0.03	0.01
5	−22.58	−23.60	0	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.01	0.01
7	−13.69	−33.54	0	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00
10	−5.77	−42.50	0	−0.00	0.00	−0.00	0.00	−0.00	0.01	−0.00	0.00
15	0.77	−49.95	0	−0.00	0.00	−0.00	0.01	−0.00	0.01	−0.01	0.00
Number of sample couplets				2,000		2,000		2,000		2,000	
*Panel B: Good sample couplets**											
1	−50.35	3.91	0	−0.01	0.01	−0.01	0.01	−0.01	0.01	−0.01	0.01
2	−43.65	−1.47	0	−0.00	0.01	−0.01	0.01	−0.01	0.00	−0.01	0.01
3	−35.82	−9.31	0	−0.01	0.00	−0.01	0.00	−0.00	0.00	−0.01	0.01
5	−22.58	−23.60	0	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00
7	−13.69	−33.54	0	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00
10	−5.77	−42.50	0	−0.00	0.00	−0.00	0.00	−0.00	0.01	−0.00	0.00
15	0.77	−49.95	0	−0.00	0.00	−0.00	0.01	−0.00	0.01	−0.00	0.00
Number of sample couplets				1,954		1,818		1,210		486	

* There is at least one treated and one control bond in one-year maturity bucket.

the DiD estimates are statistically insignificant in 69.9% of cases despite the strong negative true treatment effect at shorter maturities.

To summarize, the classical DiD specification in Eq. (1) can lead to incorrect conclusions even in the absence of treatment-unrelated idiosyncratic yield-curve effects. The problem is that the classical DiD specification produces an average treatment effect across treated sample bonds. But this is not necessarily a very meaningful or informative economic quantity. When treatment effects are heterogeneous over the term structure – and there is no a priori reason to believe that they are not – the goal of estimation should be to capture and measure this, which Specification (1) is simply not designed to do.

6.2. Yield-curve control

Dealing with the problem we have just outlined requires allowing for heterogeneous effects over the maturity spectrum. As before, the straight yield-curve control in Specification (4) does not solve the problem. For each of the 8000 datasets – 4000 sample couplets of residual maturities times the two true treatment-effect scenarios – we compute the difference between the DiD coefficients from Specifications (1) and (4). All 8000 differences equal 0.000 bps and the differences in p-values range from −0.005 to 0.000. Specifications (1) and (4) are equally poor because they both falsely assume homogeneous effects over the maturity spectrum.

The fully flexible specification in Eq. (6), however, works much better. The results are in Table 9. For brevity, samples where we do not have treated and control bonds with residual maturities of less than one year are excluded. The table provides the true underlying treatment effects at selected maturities for both the twist and the short-end scenarios and, to the right, the range of differences between true and estimated effects. The true effects are, for the most part, measured in single digit basis points. As seen, for all 5486 samples and seven selected maturities, the estimated effects are accurate to within a hundredth of a basis point. Thus, using the fully flexible yield-curve DiD specification in Eq. (6) results in meaningful and accurate treatment-effect estimates over the maturity spectrum.

The DiD estimator in the fully flexible specification, the DiD Delta curve, gives the treatment effect as a function of maturity. This is a meaningful estimator when the true treatment effect also depends on residual maturity. The simulations show that the DiD in curves approach is capable of measuring even relatively small treatment effects with precision over the maturity spectrum. By allowing for maturity-dependent effects, it resolves the problem created by nonzero correlation between residual maturity and treatment assignment that causes the classical and parallel shifts DiD specifications to fail.

7. Both false and garbled treatment effects

In this section, we combine treatment-unrelated idiosyncratic effects (Section 4) with systematic, true treatment effects (Section 6). The bond (residual maturity) samples are the same as above, and we continue to use the Diebold and Li (2006) curve in Eq. (2) to generate bond yields.

Parameter values and yield changes for treated and control bonds at selected maturities are in Table 10. We consider two idiosyncratic and two treatment effects, for a total of four combinations. The two idiosyncratic effects are the same as in Section 4 and work predominantly on either short or long maturities, respectively. The true treatment effects are the same as in Section 6, and either twist the yield curve up at the short end and down at the long end, or move it down predominantly at the short end, respectively. The idiosyncratic effects impact treated and control bonds equally while the treatment effects impact treated bonds only. In our simulation, the idiosyncratic effects are larger than the treatment effects, having a magnitude of around 50 bps at their peak versus 6 bps for the treatment effects.

Table 9

The fully flexible yield-curve DiD specification to eliminate garbled treatment effects.

This table shows treatment effects estimated with the fully flexible yield-curve DiD specification (as in Table 8). The underlying samples are the "good sample couplets" defined as those with at least one treated and one control bond in the one-year maturity bucket (see Panel B of Table 8). There are no treatment-unrelated idiosyncratic yield-curve effects in the data but the treatment effect varies over maturity. At selected maturities, the table shows these true underlying effects and, to the right, the minimum and maximum of the difference between the estimated and the true treatment effects across the two types of true treatment effects and the good sample couplets by m using the DiD Delta curve, $\Delta_4(x)$. Standard errors are clustered at the bond level and calculated using the delta method.

Residual maturity (in years)	Idiosyncratic effect (in bps)	True treatment effect (in bps)		Differences between estimated and true treatment effects (in bps)							
				$m = 0.25$		$m = 1$		$m = 3$		$m = 10$	
		Twist	Short-end	Min	Max	Min	Max	Min	Max	Min	Max
1	0	5.87	−6.23	−0.00	0.00	−0.01	0.01	−0.01	0.01	−0.01	0.01
2	0	3.75	−2.97	−0.00	0.00	−0.00	0.00	−0.00	0.01	−0.01	0.01
3	0	1.58	−1.39	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.01	0.01
5	0	−1.87	−0.26	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00
7	0	−4.11	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00
10	0	−6.09	0.08	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00
15	0	−7.72	0.09	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00
Number of sample couplets				1,954		1,818		1,210		486	

Table 10

Modeling idiosyncratic term-structure and true treatment effects.

To model the term structure we employ Diebold and Li's (2006) yield-curve specification. This table shows the parameter values to create the true underlying term structures as well as the resulting yield levels and yield differences at selected maturities. Panels A and C cover the cases of a yield-curve treatment twist and a yield-curve treatment effect only at the short-end in case of an idiosyncratic short-end effect and Panels B and D, respectively, the same in case of an idiosyncratic long-end effect from pre- to post-treatment.

	Parameter values				Yields (in %) and differences (in pps)						
	γ_0	γ_1	γ_2	λ	1y	2y	3y	5y	7y	10y	15y
Panel A: General short-end effect, treatment yield-curve twist											
Pre-curve	4.000	−2.000	0.000	0.7308	2.58	2.95	3.19	3.47	3.61	3.73	3.82
Post-curve controls	4.140	−2.650	−0.800	0.7308	2.08	2.51	2.83	3.24	3.47	3.67	3.83
Difference					−0.50	−0.44	−0.36	−0.23	−0.14	−0.06	0.01
Post-curve treated	4.030	−2.470	−0.620	0.7308	2.14	2.55	2.85	3.22	3.43	3.61	3.75
Difference					0.06	0.04	0.02	−0.02	−0.04	−0.06	−0.08
Panel B: General long-end effect, treatment yield-curve twist											
Pre-curve	4.000	−2.000	0.000	0.7308	2.58	2.95	3.19	3.47	3.61	3.73	3.82
Post-curve controls	3.350	−1.350	1.000	0.7308	2.62	2.93	3.10	3.23	3.28	3.30	3.32
Difference					0.04	−0.01	−0.09	−0.24	−0.34	−0.43	−0.50
Post-curve treated	3.240	−1.170	1.180	0.7308	2.68	2.97	3.11	3.21	3.23	3.24	3.24
Difference					0.06	0.04	0.02	−0.02	−0.04	−0.06	−0.08
Panel C: General short-end effect, treatment short-end effect											
Pre-curve	4.000	−2.000	0.000	0.7308	2.58	2.95	3.19	3.47	3.61	3.73	3.82
Post-curve controls	4.140	−2.650	−0.800	0.7308	2.08	2.51	2.83	3.24	3.47	3.67	3.83
Difference					−0.50	−0.44	−0.36	−0.23	−0.14	−0.06	0.01
Post-curve treated	4.141	−2.781	−0.670	0.7308	2.02	2.48	2.82	3.24	3.47	3.67	3.83
Difference					−0.06	−0.03	−0.01	0.00	0.00	0.00	0.00
Panel D: General long-end effect, treatment short-end effect											
Pre-curve	4.000	−2.000	0.000	0.7308	2.58	2.95	3.19	3.47	3.61	3.73	3.82
Post-curve controls	3.350	−1.350	1.000	0.7308	2.62	2.93	3.10	3.23	3.28	3.30	3.32
Difference					0.04	−0.01	−0.09	−0.24	−0.34	−0.43	−0.50
Post-curve treated	3.351	−1.481	1.130	0.7308	2.56	2.90	3.08	3.23	3.28	3.30	3.32
Difference					−0.06	−0.03	−0.01	0.00	0.00	0.00	0.00

7.1. The classical DiD specification

When both idiosyncratic and true treatment effects are present in the data, the classical DiD specification in Eq. (1) produces the sum of the corresponding individual false and garbled treatment effects from Sections 4 and 6. This is evident by comparing the DiD estimates with both effects present to the sum of the DiD estimates with the individual effects only. Across the four combinations of idiosyncratic and treatment effects, the four m's for treated-bond residual maturities, and the 1000 sample draws – a total of 16,000 cases – the differences range from −0.003 to 0.003 bps, which is negligible.

The problem with the classical DiD specification in Eq. (1) is that it assumes constant idiosyncratic and true treatment effects over the maturity spectrum. Thus, when these effects are actually heterogeneous, the specification produces a potpourri of false treatment effects and garbled true effects. The magnitude of the error is a function of the underlying yield-curve effects and the treatment-maturity correlation arising from differences in the maturity distributions of treated and control bonds. More extreme differences in effects across maturities and more extreme correlations lead to larger errors. As before, the straight yield-curve control in Specification (4) is just as problematic as the classical DiD specification. Across the four effect combinations, the four treated-bond m's, and the 1000 sample draws, the DiD estimates are identical in all 16,000 cases.

7.2. Fully flexible yield-curve DiD specification

Finally, we report on the results from running the DiD in curves in Eq. (6) on the dataset with idiosyncratic and true treatment effects. The estimated effects are based on the DiD Delta curve, $\Delta_4(x)$. Table 11 shows the results. The table provides the true underlying effects of the four combinations of idiosyncratic effects (predominantly short-end or long-end) and treatment effects (twist or predominantly short-end) at seven selected maturities. To the right, for each of the seven maturities, it shows the minimum and maximum of the range of differences between estimated and true treatment effects across good sample couplets, broken down by m for the treated bonds. As above, a good sample couplet is one in which there is both a treated and a control bond with less than one year of residual maturity. This gives a total of 10,936 runs. The largest estimation error is 0.01 bps in absolute value terms. In other words, DiD in curves accurately captures the true treatment effects.

Table 11
The fully flexible yield-curve DiD specification to eliminate both false and garbled treatment effects.
This table shows treatment effects estimated with the fully flexible yield-curve DiD specification (as in Table 8). The underlying samples are the "good sample couplets" defined as those with at least one treated and one control bond in the one-year maturity bucket (see Panel B of Table 8). There are both treatment-unrelated idiosyncratic yield-curve as well as yield-curve treatment effects in the data and both vary over maturity. At selected maturities, the table shows these true underlying effects and, to the right, the minimum and maximum of the difference between the estimated and the true treatment effects across the four combinations of idiosyncratic effects (at short- or long-end) and treatment effects (twist or short-end) and the good sample couplets separately by m using the DiD Delta curve, $\Delta_4(x)$. Standard errors are clustered at the bond level and calculated using the delta method.

Residual maturity (in years)	Idiosyncratic yield-curve effects (in bps)		True treatment effect (in bps)		Differences between estimated and true treatment effects (in bps)							
					$m = 0.25$		$m = 1$		$m = 3$		$m = 10$	
	Short-end	Long-end	Twist	Short-end	Min	Max	Min	Max	Min	Max	Min	Max
1	−50.35	3.91	5.87	−6.23	−0.01	0.01	−0.01	0.01	−0.01	0.01	−0.01	0.01
2	−43.65	−1.47	3.75	−2.97	−0.00	0.00	−0.01	0.01	−0.01	0.01	−0.01	0.01
3	−35.82	−9.31	1.58	−1.39	−0.01	0.00	−0.01	0.00	−0.01	0.00	−0.01	0.01
5	−22.58	−23.60	−1.87	−0.26	−0.00	0.00	−0.01	0.00	−0.00	0.00	−0.00	0.00
7	−13.69	−33.54	−4.11	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00	−0.00	0.00
10	−5.77	−42.50	−6.09	0.08	−0.00	0.00	−0.00	0.00	−0.00	0.01	−0.00	0.00
15	0.77	−49.95	−7.72	0.09	−0.00	0.00	−0.01	0.01	−0.01	0.01	−0.00	0.00
Number of sample couplets					3,908		3,636		2,420		972	

Table 12
The fully flexible DiD model with alternative curve specifications.
This table shows distributions of the difference between estimated and true treatment effects at selected maturities for the fully flexible yield-curve DiD specification in Eq. (6) using Fisher (1966) and Svensson (1994) curves (see Footnote 14 for details) and OLS and NLS, respectively. The table uses the 4000 simulated maturity sample couplets to produce 24,000 yield sample couplets: 8000 comprise one of the two idiosyncratic yield-curve effects and 16,000 a combination of an idiosyncratic with one of the two treatment effects. In each maturity sample couplet, the shortest (longest) bonds are removed until the shortest (longest) treated and the shortest (longest) control bond are no more than 0.25 (1) years apart from each other (no samples are lost because of this filter). Panel A shows results for good sample couplets. A sample is labeled "good" for maturity x if there are at least three treated and three control bonds in the maturity bucket to the left as well as in that to the right using the maturity-bucket grid formed by the seven selected maturities, i.e., $(0,1]$, $(1,2]$, …, $(10,15]$, and $(15,20]$. If this is not satisfied, the sample is labeled "bad" for that maturity. Panel B (C) shows results for bad sample couplets at maturity x when $m = 0.25$ or $m = 1$ ($m = 3$ or $m = 10$) years. "e" stands for error.

Residual maturity (in years)	Maturity sample couplets	Yield sample couplets	Differences between estimated and true treatment effects (in bps)													
			Fisher (1966)							Svensson (1994)						
			Mean	Min	P1	Med	P99	Max	e>1bp	Mean	Min	P1	Med	P99	Max	e>1bp
Panel A: Good samples																
1	641	3,846	0.11	−1.11	−0.47	0.07	0.88	1.99	23	0.00	−0.01	−0.00	0.00	0.00	0.01	–
2	1,287	7,722	−0.03	−1.55	−0.52	−0.02	0.42	1.00	6	0.00	−0.01	−0.00	0.00	0.00	0.01	–
3	2,105	12,630	−0.04	−1.31	−0.47	−0.02	0.29	0.69	4	0.00	−0.01	−0.00	0.00	0.00	0.01	–
5	3,552	21,312	0.01	−0.53	−0.25	0.02	0.28	0.99	–	0.00	−0.01	−0.00	0.00	0.00	0.01	–
7	3,805	22,830	0.02	−0.77	−0.27	0.01	0.34	1.38	3	−0.00	−0.01	−0.00	−0.00	0.00	0.01	–
10	3,980	23,880	−0.01	−0.99	−0.28	−0.01	0.26	0.49	–	0.00	−0.01	−0.00	0.00	0.00	0.01	–
15	1,447	8,682	0.02	−0.83	−0.44	0.02	0.46	0.86	–	0.00	−0.01	−0.00	0.00	0.00	0.01	–
Total	16,817	100,902							36							0
Panel B: Bad samples when $m = 0.25$ or $m = 1$ years																
1	1,372	8,232	0.08	−2.05	−0.73	0.05	1.09	3.30	135	−0.00	−0.06	−0.01	−0.00	0.01	0.15	–
2	956	5,736	−0.04	−2.14	−0.72	−0.03	0.61	1.74	32	−0.00	−0.02	−0.01	−0.00	0.01	0.04	–
3	618	3,708	−0.06	−1.34	−0.65	−0.05	0.42	1.10	8	−0.00	−0.01	−0.00	−0.00	0.00	0.01	–
5	107	642	0.02	−0.53	−0.42	0.02	0.41	0.64	–	−0.00	−0.01	−0.00	0.00	0.00	0.00	–
7	87	522	0.03	−0.66	−0.50	0.03	0.48	0.70	–	0.00	−0.00	−0.00	0.00	0.00	0.00	–
10	13	78	−0.06	−1.04	−1.04	−0.04	0.36	0.36	1	0.00	−0.00	−0.00	0.00	0.00	0.00	–
15	1,429	8,574	0.02	−3.13	−0.79	0.01	0.94	5.62	118	−0.00	−0.02	−0.01	0.00	0.01	0.02	–
Total	4,582	27,492							294							0
Panel C: Bad samples when $m = 3$ or $m = 10$ years																
1	1,987	11,922	0.27	−3.44	−2.10	0.09	4.72	8.47	2,598	0.00	−4.99	−0.17	−0.00	0.16	5.27	25
2	1,757	10,542	0.05	−3.68	−1.01	0.03	1.35	2.87	388	0.00	−0.89	−0.06	−0.00	0.06	1.51	3
3	1,277	7,662	−0.02	−3.04	−0.98	−0.00	0.77	2.28	91	0.00	−0.19	−0.02	0.00	0.02	0.44	–
5	341	2,046	−0.00	−0.99	−0.53	0.01	0.46	0.81	–	0.00	−0.02	−0.01	0.00	0.01	0.03	–
7	108	648	0.02	−0.66	−0.36	0.01	0.44	0.59	–	0.00	−0.01	−0.00	0.00	0.00	0.00	–
10	7	42	−0.00	−0.40	−0.40	−0.02	0.37	0.37	–	0.00	−0.00	−0.00	0.00	0.00	0.00	–
15	1,124	6,744	0.00	−1.70	−0.67	0.01	0.62	1.60	32	0.00	−0.02	−0.01	0.00	0.01	0.01	–
Total	6,601	39,606							3,109							28

These results illustrate that the fully flexible specification is capable of separating small, maturity-dependent treatment effects from large, maturity-dependent idiosyncratic effects. Through the DiD Delta curve, $\Delta_4(x)$, the specification estimates the treatment effect as a function of maturity, and through the post-treatment Delta curve, $\Delta_3(x)$, it accurately controls for treatment-unrelated idiosyncratic effects. Thus, false treatment effects are eliminated and treatment effects meaningfully expressed as a function of maturity. As a result, correlation between treatment assignment and residual maturity is not an issue.

7.3. Trend plots

Visual inspection of trend lines is a commonly used diagnostic tool to assess the exclusion restriction in DiD analyses. A graphical detection of non-parallel pre-event trends is interpreted as DiD analyses being problematic because measured treatment effects can result from treatment-unrelated variation in the data (Roberts and Whited, 2013). When dealing with variables that have term structures, it is important to pay attention to differences in residual maturities between treated and control units. Suppose, for example, that treated units have shorter maturities than controls and that there are idiosyncratic effects in the pre-event period that pushes the long end up. This would result in non-parallel trends of average outcome variables. Observing this, a researcher may decide to abandon the research. If there are true treatment effects, these will consequently go undetected.

It is also possible that the distributions of treated and control units over residual maturities combine with heterogeneous idiosyncratic effects in such a way that pre-event trends of averages look parallel despite differences in residual maturities, while, subsequently, post-event idiosyncratic movements give rise to false treatment effects under the classical DiD specification as discussed. A similar problem can arise under a steep curve, where differences in residual maturities between treated and control units give rise to differences in pre-event outcome variables that can be so large that it becomes difficult to detect smaller trends. Falsely assured by what looks like parallel trends, a researcher running the classical DiD specification, with or without maturity control, may subsequently falsely conclude that there is a statistically significant treatment effect. The key point is that trend plots need to provide adequate control for differences in residual maturities to be meaningful.

The term-effect problem with respect to trend plots can potentially be mitigated by using maturity buckets – averaging first within buckets and then across – but the usefulness of this approach depends on the closeness of matching and outcome variable variation within buckets. A more complicated approach could involve estimating curves at a suitable frequency and look for systematic deviations from parallel trends in different parts of the maturity spectrum. However, it is even more important to use a well designed DiD methodology. As implied by our results, the fully flexible DiD specification is well specified even if there are non-parallel pre-event trends of treated and control unit averages that are driven by residual maturity differences between the two groups.

8. Alternative curves and an example

In this section, we address the sensitivity of a DiD in curves to the functional form. Using our simulated data, we consider alternatives to the Nelson–Siegel/Diebold–Li curve from the literature. In addition, we provide an example with real data that compares performance using the classical DiD specification versus DiD in curves.

8.1. Alternative curves

Prior to Nelson and Siegel (1987), curves were often fitted as polynomials, especially cubics, sometimes with additional non-exponential terms. A prominent example is Fisher (1966), who proposes a cubic with an added $\ln(x)$ term. Cubics and their variations have the drawbacks that they blow up at long maturities and can impute too much local curvature even though they can also fit real data well (Nelson and Siegel, 1987). Below, we present the results from running the fully flexible DiD specification on our simulated data using two alternative curves: Fisher (1966) and Svensson's (1994) extension of Nelson and Siegel (1987).[14] We use the same 4000 sample couplets of maturities as before.

In practice, as well as in our simulated data, the yield curve typically exhibits substantial curvature at short maturities. This can be especially problematic when we work with a less than perfect curve specification. For this reason, we sequentially remove the shortest bond from each sample couplet until the difference in maturity between the shortest treated and shortest control bond is no more than 0.25 years. We also sequentially remove the longest bond from each sample couplet until the difference in maturity between the longest treated and longest control bond is no more than one year. No sample couplet is lost as a result of this proximity filtering procedure.

For each sample couplet of maturities, we generate two samples of treated and control bond yields using the two idiosyncratic effects discussed above. We then generate four additional samples of yields by adding the two treatment effects (as above) to each idiosyncratic effect. Thus, in total, there are 24,000 individual samples (of Diebold–Li yield couplets), 16,000 of which have true treatment effects.

For each of the two alternative curve specifications, we estimate treatment effects using the fully flexible approach at the following maturities: 1, 2, 3, 5, 7, 10, and 15 years. The grid formed by these test-maturities gives rise to a set of maturity buckets, $\{(a_i, b_i]\}_{i=1}^{8}$, with the leftmost and rightmost buckets being $(0, 1]$ and $(15, 20]$, respectively. For each test-maturity, we label a sample as "good" if there are at least three treated bonds and at least three controls in the maturity buckets to the immediate left and right. If this is not satisfied, the sample is labeled as "bad" for that test-maturity. Under Fisher's (1966) curve specification, the model is estimated with OLS. Under Svensson (1994), we use NLS.

Table 12 reports on the differences between the estimated and true treatment effects at the seven test-maturities. We first discuss Panel A, which is for the good samples. Because of the simulation structure with $m = 0.25, 1, 3, 10$ for treated bonds and $m = 0.25$ for controls, there are fewer good samples for the short and 15 year maturities than for the intermediate test-maturities. As seen, the Fisher (1966) model performs well when samples are good; estimation errors exceed one basis point in only 36 out of 100,902 cases, or 0.04%. Most of these are at the one-year maturity, where the incidence of errors larger than a basis point is still small, namely, 0.60%. The Svensson (1994) model, has a perfect fit, with the largest estimation error being 0.01 bps in absolute value. This is not surprising since it nests Nelson–Siegel/Diebold–Li.

Panel B is for bad samples when the unconditional distributions of control and treated bond maturities are the same or close ($m = 0.25$ or 1 for treated bonds). For these bad samples, estimation performance remains good. For the Fisher (1966) model, there are now 294 out of 27,492 (1.07%) estimates that exceed one basis point. These are predominantly at the extremities of the maturity spectrum. Svensson's model once again performs almost perfectly, with the largest estimation error being 0.15 bps (in absolute value). This is at the one-year maturity. At other maturities, the largest error is 0.04 bps in absolute value.

Panel C is for bad samples when treated bonds are drawn from distributions that are highly skewed relative to that of control bonds ($m = 3$ or 10). The scarcity of treated bonds at the short end of the maturity spectrum gives rise to poor estimation performance at short maturities. However, performance remains good even for the Fisher (1966) model at maturities of three years and up, with only 0.72% of cases having an estimation error larger than one basis point. In this range, Svensson's (1994) model has an almost perfect fit.

These results show that having the wrong curve specification is not necessarily a problem as long as treated and control bond sample distributions are not heavily skewed away from each other. Larger denseness of bonds around the test-maturity helps performance.

8.2. Example with real data

The example uses a dataset of two groups of zero-coupon Spanish government bonds, each with its own set of haircuts in Eurosystem repos as set by the European Central Bank (ECB). In these operations, the Eurosystem provides reserves to banks in the euro area against collateral, including Spanish government bonds. For a given residual maturity, haircuts for Group 1 bonds are low and haircuts for Group 2 bonds are correspondingly high. This disparity is a result of bonds being rated by different rating agencies (or not at all), with different risk

[14] This means that $\mathbf{L}_{it}$ in Specification (6) becomes either (i) Fisher (1966): $(1, x, x^2, x^3, \ln(x))$; or (ii) Svensson (1994): $(1, l_{1,t}(x;\lambda_1), l_{2,t}(x;\lambda_1), l_{3,t}(x;\lambda_2))$, where $l_{1,t}(x;\lambda_1) = \frac{1-e^{-x/\lambda_1}}{x/\lambda_1}$, $l_{2,t}(x;\lambda_1) = \frac{1-e^{-x/\lambda_1}}{x/\lambda_1} - e^{-x/\lambda_1}$, and $l_{3,t}(x;\lambda_2) = \frac{1-e^{-x/\lambda_2}}{x/\lambda_2} - e^{-x/\lambda_2}$. Subscripts on x are not shown.

Table 13
Example – Eurosystem haircut update.
On October 1, 2013, haircuts in Eurosystem repos for reserves were updated differentially for eligible bonds rated AAA to A− (labeled "controls") versus those rated BBB+ to BBB− (labeled "treated") according to the Eurosystem's rating classification, with haircuts increasing relatively more in the latter group across the maturity spectrum. We run several DiD specifications on Spanish government bonds using a forty-day window around this event. Because all treated bonds are zeros, we use zeros only as controls. Bonds with stale market prices in Bloomberg as well as the announcement date, September 27, and the day between this and October 1, September 30, are dropped. Panel A shows the distribution of residual maturities. The truncated sample drops bonds with residual maturities of more than 7.5 years. The other panels show DiD estimates using the classical DiD specification (Panel B) or at selected maturities under DiD in curves (Panel C). Specifications with Diebold–Li curves are estimated with NLS; otherwise, OLS. Standard errors are clustered at the bond level (and calculated using the delta method when NLS is used). Superscripts indicate statistical significance at the 1% (a), 5% (b), and 10% levels (c). Significant coefficients at 10% or better are in bold. *t*-statistics are in parentheses.

	Full sample				Truncated sample			
Panel A: Number of bonds by maturity buckets								
Maturity buckets	Treated		Control		Treated		Control	
(in years)	Number	Percent	Number	Percent	Number	Percent	Number	Percent
0–1	2	15.4	5	6.8	2	18.2	5	14.3
1–3	5	38.5	10	13.7	5	45.5	10	28.6
3–5	2	15.4	9	12.3	2	18.2	9	25.7
5–7	1	7.7	10	13.7	1	9.1	10	28.6
7–10	2	15.4	6	8.2	1	9.1	1	2.9
10–15			11	15.1				
15–20			9	12.3				
20–25			10	13.7				
> 25	1	7.7	3	4.1				
Total	13	100.0	73	100.0	11	100.0	35	100.0

	Full sample		Truncated sample	
Panel B: Classical DiD specification, Eq. (1)				
β_{DiD}	**−0.081**^{a}	**−0.057**^{c}	−0.004	−0.007
	(−2.73)	(−1.84)	(−0.21)	(−0.41)
ln(x)		**0.927**^{c}		−0.266
		(1.96)		(−1.38)
Adj. R-squared	0.6912	0.7152	0.9320	0.9339
Panel C: Fully flexible yield-curve DiD specifications, Eq. (6)				
Maturity (in years)	Diebold–Li (2006)	Fisher (1966)	Diebold–Li (2006)	Fisher (1966)
1	**0.038**^{c}	0.005	**0.041**^{c}	**0.039**^{b}
	(1.95)	(0.32)	(1.72)	(2.27)
2	0.018	−0.010	0.011	**0.022**^{c}
	(1.33)	(−0.91)	(0.92)	(1.82)
3	0.006	0.002	−0.007	−0.000
	(0.46)	(0.13)	(−0.45)	(−0.02)
5	−0.004	0.017	−0.024	−0.014
	(−0.30)	(1.46)	(−1.51)	(−0.94)
7	−0.009	−0.002	−0.029	−0.022
	(−0.49)	(−0.20)	(−1.39)	(−0.80)
Adj. R-squared	0.9903	0.9921	0.9804	0.9807
$\hat{\lambda}$	0.6148	–	0.4726	–

assessments of Spain, and the fact that haircuts in Eurosystem repos are a function of these credit ratings. The event is a Eurosystem haircut update on October 1, 2013, that caused haircuts in these two groups to diverge, with Group 2 bonds experiencing a relative increase in haircuts over the whole maturity spectrum.[15] We label Group 2 bonds as treated and Group 1 bonds as controls. The two groups of bonds are from the public list of eligible collateral as published by the ECB.[16] Market prices are obtained from Bloomberg for a forty-day window around the event. A two-month window is commensurate with best practice in the literature (see Table 2). Bonds that experience rating changes over the event window or have stale or missing market prices are dropped. Because the Spanish government bonds with low ratings and high haircuts (Group 2) with market prices in the Bloomberg system are exclusively zeros, we also use zeros as controls. We are not aware of differences between the bonds in the two groups except for their residual maturities and Eurosystem haircuts.

The economic question is whether haircuts in central-bank repos affect bond yields. They may do so if the reserves provided in these repos have value to banks that can participate in the operations. To examine this, we employ DiD analyses.[17] Results are in Table 13.

The first point to observe is that the distributions of residual maturities of the two groups are dissimilar (Panel A). For example, there is only one treated bond with maturity in excess of ten years, whereas 45.2% of controls are in this range. Thus, we run the analyses on the full dataset as well as a truncated version where maturity distributions are more similar. In the truncated sample, bonds with residual maturity of more than 7.5 years are dropped, leaving the range of maturities as 0.33 to 7.08 years for both groups. Panel B (to the left) shows that the classical DiD specification on the full sample, with or without ln(x) as maturity control, results in an economically large negative treatment estimate (−8.1 bps and −5.7 bps, respectively) that is statistically significant at conventional levels. One could, therefore, be tempted to conclude that higher haircuts improve yields, which makes little economic sense. However, Panel B (to the right) also shows that on the truncated dataset with more congruent distributions of treated and control bonds, the classical DiD specification results in a statistically insignificant treatment estimate in both cases (with and without maturity control). The negative estimate based on the full

[15] Full details on how haircuts are set by the ECB for government bonds, the data, the event, and references are in Internet Appendix A.3.

[16] See https://www.ecb.europa.eu/mopo/coll/assets/html/list-MID.en.html.

[17] The haircut update was announced on September 27, 2013 and implemented two business days later, on October 1, 2013. We exclude the two business days before implementation, thus estimating a joint announcement and implementation effect.

dataset with highly dissimilar maturity distributions is an example of a false treatment effect.

It would be premature, however, to conclude from this that there is no treatment effect. Panel C shows the results from DiD analyses in curves using Diebold and Li (2006) and Fisher (1966) specifications on the full and truncated datasets. The results show a significant treatment effect at the one-year maturity of around four basis points, but no effect at other test maturities. This could be because participants in Eurosystem repos to whom additional reserves from the central bank have marginal value hold shorter-dated paper. Under Diebold–Li curves, the results are approximately the same on the full and the truncated datasets (3.8 bps and 4.1 bps, respectively at the one-year maturity), which serves as an example of the robustness of this particular specification in real data. This may be because bonds are often priced off this or similar curves in practice. While the Fisher specification shows no significant treatment effect under the full dataset, when maturity distributions are more congruent, it gives results that are similar to those under Diebold–Li, in particular, an estimated treatment effect on treated bonds of 3.9 bps at the one-year maturity.

9. Semi-synthetic matching

In this section, we discuss an alternative two-stage matching approach for dealing with the term effect problem. In the first stage, DiDs are calculated individually for each treated bond using a maturity-matched synthetic control. In the second stage, treatment effects are examined over the maturity spectrum. We continue to work with the same samples as above. In this zero coupon setup, if we fit a curve in the second stage with the same functional form as used in the fully flexible DiD specification, point estimates of treatment effects turn out to be the same. More generally, the semi-synthetic matching approach is also well suited to samples with coupon bonds.

9.1. First stage

The basic idea is to compare actual yields of treated bonds to model-implied yields of synthetic controls that are matched on residual maturities and coupons. We calculate the "semi-synthetic DiD" for each treated bond as follows:

1. For each period, pre-event and post-event, estimate the control-bond yield curve.
2. For each treated bond, i, and period, t, compute the yield difference, $\Delta y_{i,t}$, as the actual yield of the treated bond minus the model-implied yield of the matched synthetic control bond.
3. For each treated bond, i, calculate the difference between the yield differences post and pre treatment. This is the semi-synthetic DiD for each bond over the treatment event, specifically, $yield_i^{DiD} = \Delta y_{i,Post} - \Delta y_{i,Pre}$.

The semi-synthetic DiDs remove idiosyncratic term effects that impact treated and control bonds the same way. This resolves the problem of false treatment effects. However, garbled measurement of true treatment effects can still arise, depending on how the semi-synthetic DiDs are used. We consider two types of approaches. In the first and simplest one, the semi-synthetic DiDs are averaged, possibly within selected maturity buckets. In the second and more complex approach, a curve is fitted through the semi-synthetic DiDs.

We use the same 16,000 datasets as in Section 7, generated by 4000 sample couplets of residual maturities and two possibilities each for the idiosyncratic and true treatment effects. In stage one, we estimate semi-synthetic DiDs within these datasets. The yield curve for control bonds is estimated with the Diebold and Li (2006) specification in Eq. (2). For each of the 16,000 datasets, there are fifty treated bonds and, thus, fifty semi-synthetic DiDs.

9.2. Second-stage: Average semi-synthetic DiDs

For each of the 16,000 samples, we regress the fifty semi-synthetic DiDs on a constant,

$$yield_i^{DiD} = \beta_{DiD} \times C + \varepsilon_i, \tag{9}$$

to produce the average semi-synthetic DiD. Since the synthetic DiDs are designed to eliminate false treatment effects, we compare the resulting DiD estimates from Eq. (9) to those from the classical DiD specification in Eq. (1) when only true treatment effects are present. Across the 16,000 cases, the differences in DiD coefficients range from −0.003 bps to 0.004 bps.[18] Thus, averaging over semi-synthetic DiDs eliminates false treatment effects but not garbled true treatment effects, and the discussion on this issue from Section 6 applies. Averaging semi-synthetic DiDs within maturity buckets can reduce the garbling problem, but the paucity of data in practice limits this approach.

9.3. Second stage: Fitting a curve

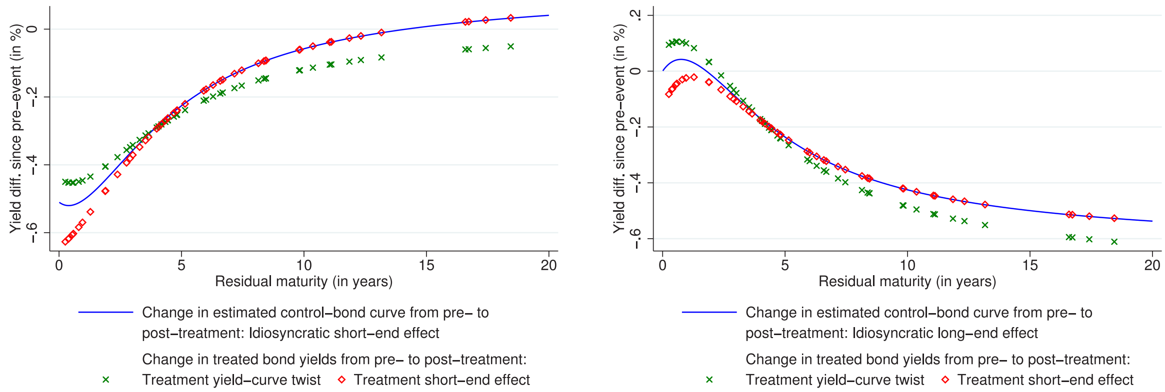
There are many ways to fit a second-stage curve, for instance, one can run linear or nonlinear regressions or use specifications from the yield-curve literature as above. In practice, it may be advisable to consider different curve specifications for robustness. Below, we fit a Diebold–Li curve and compare the results to those from the fully flexible approach in Eq. (6). In addition, we consider linear regression. Although misspecified, linear regression is standard and, therefore, interesting to look at.

We fit Diebold–Li curves for the semi-synthetic DiDs from the first stage against residual maturity in the same 10,936 good datasets as in Table 11 (NLS with $\lambda_{Seed} = 1$). Across the same seven selected maturities, we obtain identical results as under the fully flexible yield-curve DiD approach. The largest difference in absolute value terms is 0.01 bps.[19] Thus, fitting a well specified curve in a second stage resolves the remaining problem, namely, garbled measurement of true treatment effects.

Fig. 6 provides graphical intuition for this result. The figure is based on a randomly selected sample couplet with $m = 0.25$ for the treated bonds. Both types of effects are present. The solid (blue) line shows the change in the estimated Diebold–Li control-bond curve over the treatment for an idiosyncratic short-end effect on the left and a long-end effect on the right. In each plot, the (green) crosses and (red) diamonds show the change in the treated bond yields over the event induced by the treatment twist and short-end effect, respectively, together with the respective idiosyncratic effect. Therefore, the differences between the (green) crosses and the (blue) solid line represent the semi-synthetic DiDs under the treatment twist, while the differences between the (red) diamonds and the (blue) solid line are the semi-synthetic DiDs under the predominantly short-end treatment effect. Applying second-stage curve fitting after first-stage semi-synthetic matching and using the Diebold and Li (2006) curve in each step essentially does in two steps what the fully flexible yield-curve DiD specification in Eq. (6) does in one.

Using linear regression in the second stage works less well, but might provide the right qualitative conclusion. Consider, for example, the twist effect in the sample of fifty treated and fifty control bonds shown in Panel A of Fig. 6. Regressing the fifty semi-synthetic DiDs

[18] Details are in Table A.4 in the Internet Appendix.
[19] Details are in Table A.5 in the Internet Appendix.



a) Idiosyncratic short-end effect

b) Idiosyncratic long-end effect

Fig. 6. Illustration of semi-synthetic matching.
This illustration is based on a random sample couplet when $m = 0.25$ years for both control and treated bonds. Figs. 6a and 6b provide graphical illustrations for semi-synthetic matching when there is an idiosyncratic yield-curve effect only at the short-end or only at the long-end, respectively. In each plot, given the idiosyncratic yield-curve effects there is either an additional yield-curve treatment twist or a yield-curve treatment effect predominantly at the short-end. The solid (blue) line in each plot shows the change in the estimated control-bond curve from pre- to post treatment using Diebold–Li (2006). The (green) crosses and (red) diamonds in each plot show the changes in the treated bond yields from pre- to post-treatment for a yield-curve treatment twist and predominantly short-end effect, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on residual maturity, x, yields the following fitted line (t-statistics in brackets):

$$\widehat{yield}_i^{DiD} = \underset{(9.53)}{4.274} - \underset{(-16.00)}{0.915}\, x_i \qquad \text{and} \qquad R^2_{adj} = 0.8389. \tag{10}$$

The positive intercept reflects the positive short-end effect, and the negative slope coefficient reflects the negative long-end effect. However, because the semi-synthetic DiDs are not linear in residual maturity, using a linear specification in the second-stage estimation gives less accurate results. For example, according to the linear regression the effect is zero at 4.67 years while the true effect is zero at 3.82 years. A linear second-stage specification is problematic if the treatment effect has more than one zero or is otherwise highly nonlinear.

9.4. Advantages and disadvantages of different approaches

Both the DiD in curves in Eq. (6) and semi-synthetic matching with second-stage curve fitting resolve the problems of false treatment effects and garbled measurement of true treatment effects. The point estimates of the treatment effect are identical under the two approaches when the same curve specification is used. They are to a large extent just two different ways to estimate the same DiD Delta curve. Notwithstanding this, the two approaches have different advantages and disadvantages. With spot rates on the left-hand side of the regression equation, the fully flexible yield-curve DiD approach estimates the DiD Delta curve with one single regression and allows clustering standard errors at the bond level. This ability to cluster is lost under semi-synthetic matching. As discussed in the literature, calculating standard errors when synthetic controls are used may require bootstrapping methods (Xu, 2017; Abadie, 2021).

However, semi-synthetic matching also has some advantages, such as being more easily adapted to samples with coupon bonds. Furthermore, semi-synthetic matching puts less demands on maturity coverage for both groups, treated bonds or controls, since curves are only estimated for one of these in the first stage. Although we used synthetic controls above, one could reverse the process and calculate semi-synthetic DiDs with actual control-bond yields and synthetic yields for treated bonds. In semi-synthetic matching, error is reduced if one uses the synthetic yields of the group where curves can be estimated with the most precision. The semi-synthetic approach is suited to staggered DiD.

In the literature, one can also find examples of fully-synthetic matching, where both treated and control spot curves are synthetic (Eser and Schwaab, 2016; Lentner, 2023). This requires good samples for both sets of bonds. An advantage of the semi-synthetic approach relative to a fully synthetic procedure is that potential imprecision from curve fitting is limited to one group of bonds.

Some key advantages (+) and disadvantages (−) of the fully flexible yield-curve DiD specification and the semi-synthetic matching approach:

- Fully flexible yield-curve DiD specification:
  - \+ Simple implementation: One single regression (using standard software)
  - \+ Allows clustering standard errors at the bond level
  - − Designed for zero coupon bonds
  - − Requires good residual maturity coverage for both groups
- Semi-synthetic matching:
  - − Implementation is more laborious
  - − Does not allow clustering standard errors at the bond level
  - \+ Works with coupon bonds
  - \+ Do not need to estimate curves for both groups (treated bonds and controls).

10. False and garbled effects in non-DiD settings

The issues discussed in this paper are not unique to DiD analysis, but are relevant also in simpler settings when the left-hand side variable exhibits a term structure. Consider, for example, a cross-section of fixed-income securities where a subgroup is assigned a characteristic that is hypothesized to have impact on yields. We label this subgroup *assigned* and consider several specifications that are often used in the literature to capture the incremental yield of assigned bonds. The issue is, similarly to before, that this incremental difference, or assignment effect, can be heterogeneous over the maturity spectrum. As a result, the standard specifications are misspecified in significant ways.

Table 14
False and garbled assignment effects in pure cross-sectional data.
This table shows the results of estimating assignment effects with the specification indicated in each panel. Panels A to D provide the distributions of assignment effects. Panel E shows the true underlying effects at selected maturities and, to the right, distributions of the difference between estimated and true assignment effects by maturity. In terms of data, Panels A to D are based on the 1000 sample draws that contain sample couplets for four modes m of the assigned bonds, namely $m = 0.25, 1, 3, 10$ (non-assigned bonds have $m = 0.25$ always). This is a total of 4000 sample couplets (see Table 4). Each sample couplet represents one cross-section of assigned and non-assigned bonds. Panel E is based on the subset of 2734 "good sample couplets" defined as those with at least one assigned and one non-assigned bond in the one-year maturity bucket (see Panel B of Table 8). The underlying shape of the yield curve in all panels corresponds to the pre-event curve from the DiD analysis (see Fig. 4). In Panels A to C there are no true assignment effects. In Panels D and E there are true assignment effects. They correspond to the treatment yield-curve twist of the DiD analysis (see Fig. 4a). The largest and smallest numbers in absolute value terms in the distribution within each panel are marked in bold.

Mode m	Residual maturity (in years)	Number of sample couplets	True assignment effects (in bps)	Distribution of estimated assignment effects (in bps)				
				Mean	SD	Med	Min	Max
Panel A: False effects with Eqn (1′): $yield_i = \beta_0 + \beta_a \mathbb{1}_{Assigned,i} + \varepsilon_i$								
0.25	–	1,000	0	−0.47	8.57	−0.45	**−30.56**	30.56
1	–	1,000	0	4.42	8.22	4.20	−18.79	30.87
3	–	1,000	0	12.63	7.64	12.66	−14.62	36.43
10	–	1,000	0	27.65	7.16	27.51	2.51	**57.62**
Panel B: False effects with Eqn (1″): $yield_i = \beta_0 + \beta_a \mathbb{1}_{Assigned,i} + \beta_1 x_i + \varepsilon_i$								
0.25	–	1,000	0	−0.22	4.71	−0.31	**−13.01**	15.02
1	–	1,000	0	2.31	4.51	2.36	−11.83	18.17
3	–	1,000	0	6.27	3.92	6.27	−4.30	16.67
10	–	1,000	0	6.69	3.60	6.76	−3.54	**19.44**
Panel C: False effects with Eqn (4′): $yield_i = \mathbf{B}' \mathbf{L}_i + \beta_a \mathbb{1}_{Assigned,i} + \varepsilon_i$								
0.25	–	1,000	0	−0.00	0.00	−0.00	−0.00	0.00
1	–	1,000	0	−0.00	0.00	−0.00	−0.00	0.00
3	–	1,000	0	0.00	0.00	0.00	−0.00	0.00
10	–	1,000	0	−0.00	0.00	0.00	**−0.00**	**0.00**
Panel D: Garbled effects with Eqn (4′): $yield_i = \mathbf{B}' \mathbf{L}_i + \beta_a \mathbb{1}_{Assigned,i} + \varepsilon_i$								
0.25	–	1,000		−2.02	0.54	−2.02	−3.88	**−0.26**
1	–	1,000	Maturity-	−2.24	0.48	−2.25	−3.53	−0.41
3	–	1,000	dependent	−2.87	0.48	−2.88	−4.37	−1.12
10	–	1,000		−4.35	0.50	−4.36	**−5.71**	−2.68
Panel E: Correct effects with Eqn (6′): $yield_i = \mathbf{B}_1' \mathbf{L}_i + \mathbf{B}_2' \mathbf{L}_i \mathbb{1}_{Assigned,i} + \varepsilon_i$								
–	1	2,734	5.87	0.00	0.00	0.00	−0.01	**0.01**
–	2	2,734	3.75	−0.00	0.00	−0.00	**−0.01**	0.01
–	3	2,734	1.58	−0.00	0.00	−0.00	−0.01	0.01
–	5	2,734	−1.87	−0.00	0.00	0.00	−0.00	0.00
–	7	2,734	−4.11	−0.00	0.00	−0.00	−0.00	0.00
–	10	2,734	−6.09	−0.00	0.00	−0.00	−0.00	0.00
–	15	2,734	−7.72	−0.00	0.00	−0.00	−0.00	0.00

Specification (1′) is an adaptation of the classical DiD specification in Eq. (1) to a plain cross-sectional setup. β_a measures the assignment effect in the cross-section of assigned and non-assigned bonds. Specification (1″) controls for maturity, x, linearly. Specification (4′) controls for the term structure, analogously to the straight yield-curve control in Eq. (4). Finally, Specification (6′) estimates separate Diebold–Li curves for each group of bonds, analogously to the fully flexible DiD specification in Eq. (6).

$$yield_i = \beta_0 + \beta_a \mathbb{1}_{Assigned,i} + \varepsilon_i \tag{1′}$$

$$yield_i = \beta_0 + \beta_a \mathbb{1}_{Assigned,i} + \beta_1 x_i + \varepsilon_i \tag{1″}$$

$$yield_i = \mathbf{B}'\mathbf{L}_i + \beta_a \mathbb{1}_{Assigned,i} + \varepsilon_i \tag{4′}$$

$$yield_i = \mathbf{B}_1'\mathbf{L}_i + \mathbf{B}_2'\mathbf{L}_i \mathbb{1}_{Assigned,i} + \varepsilon_i \tag{6′}$$

To study this setting, we use the same 4000 sample couplets of residual maturities as before (1000 sample draws and four modes $m = 0.25, 1, 3, 10$ of treated bonds), with "treated" ("control") relabeled "assigned" ("non-assigned"). Yields are generated by Diebold–Li curves, as before. We use the pre-event curve and the twist effect in Fig. 4 as the baseline curve for non-assigned bonds and the assignment effect, respectively. We first consider the case that there is no true assignment effect. Instead, unknown to a researcher, there is a common curve for assigned and non-assigned bonds. Panels A, B, and C of Table 14, show the results for the first three specifications. Not unexpectedly, the two first specifications result in a high incidence of large false assignment effects, with the first being the worst.

However, the third specification with straight yield-curve control eliminates the false assignment effects. Interestingly, this contrasts with the parallel case in the DiD setup, where straight yield-curve control resulted in large false treatment effects. The reason straight yield-curve control performs worse in the DiD setup in this case is the tendency of this specification to ascribe differential movements over an event window as being due to treatment, when this may simply be due to common effects that are heterogeneous over the maturity spectrum.

When there is a true assignment effect that is heterogeneous over the maturity spectrum, however, Specification (4′) does not work well because it assumes, falsely, that the assignment effect is independent of residual maturity. Effectively, it measures an average assignment effect over assigned bonds, which is not necessarily an informative quantity. This is seen in Panel D, where the twist effect is added to assigned bonds. The estimated assignment effects on all 4000 sample couplets are negative, being between −5.71 and −0.26 bps. In contrast, the true effect at the one-year residual maturity, for instance, is a positive 5.87 bps. Panel E shows that the fully flexible assignment specification in Eq. (6′) resolves the problem; the difference between the estimated assignment effect and the true assignment effect is zero at all maturities. Thus, when assignment effects vary over the maturity spectrum – and there is no a priori reason to believe that they do not – it is important to incorporate this into the specification, just as in DiD analyses.

As in the DiD setup, one can also proceed in two stages by first estimating the yield differences between assigned bonds and maturity-matched synthetic non-assigned bonds (or the reverse). Averaging in the second stage, as in Ang et al. (2010), would give the same results as in Specification (4′). Curve fitting in a second stage is identical to Specification (6′), assuming consistent functional forms. The cross-sectional setup described here can be extended to a panel setting where the assignment Delta curve can be estimated on a date by date basis and averaged, for example, across dates as in the Fama-MacBeth procedure (Nyborg and Woschitz, 2021).

11. Concluding remarks

It is common practice in finance to use difference-in-differences (DiD) analysis to test the impact of hypothesized treatment on variables that exhibit term structures. However, for such variables, this paper shows that the classical DiD specification in Eq. (1) systematically produces false and garbled treatment effects. This is the case even under random treatment assignment. These problems arise because of two ubiquitous features of real data, namely: (i) term effects, which vary over time and may or may not be related to hypothesized treatment, and (ii) different sample distributions of maturities for treated and control units. The second feature gives rise to nonzero correlation between the treatment assignment and residual maturity which, given the first feature, leads to a problem of false and garbled treatment effects. Although we have cast our analysis in the context of bond yields and the term structure of interest rates, the issues discussed in this paper are relevant for any variable that exhibits a term structure, for example, option-implied volatilities, futures prices, risk premia, and so on. If a variable exhibits a term structure, there is little reason to believe that treatment effects should be homogeneous over the maturity spectrum, that is, have no term structure.

The classical DiD equation is misspecified because it erroneously assumes that treatment unrelated effects for each unit are fixed, when they actually depend on residual maturity, and because it does not allow for heterogeneous treatment effects over the maturity spectrum. Researchers sometimes try to control for the term structure by augmenting or modifying the classical DiD specification with control variables that capture residual maturities. However, this does not do much in terms of reducing the fundamental misspecification unless these control variables are also interacted with the specification's indicator variables. As we show, allowing for such interaction is essential with respect to addressing the maturity mismatch feature of most data and treatment effects that are heterogeneous over the maturity spectrum. Failure to do this can lead to large false treatment effects, whose signs are sample dependent, overlaid on garbled estimates of true treatment effects, whose signs are also sample dependent.

We discuss two ways to resolve the problem. The first approach is designed for zero coupon bonds or analogous variables whose values can be written as functions of residual maturity. It captures the treatment effect as a curve over the maturity spectrum (the DiD Delta curve) by including a baseline pre-event curve in the DiD specification as well as incremental curves for (i) post event, (ii) treated bonds, and (iii) the interaction of these, which is the DiD Delta curve. This is very much like a standard DiD specification, but over curves. The drawback of this approach is that it is not well suited to studying yields of coupon bonds or other situations where the link between residual maturity and the value of the dependent variable is similarly complex.

Thus, we also discuss a two-step approach that we refer to as semi-synthetic matching. In a first step, DiDs of individual treated units are calculated relative to synthetic matches. In a second step, these individual units can be examined over the maturity spectrum. The two methods are identical for zero coupon bonds when curves are estimated with the same functional form. However, the one-step approach has the advantage that it allows clustering standard errors.

Naturally, as for any econometric procedure, the performance of these approaches relate to how noisy the underlying data is and also to how precisely curves can be estimated. Although yield curves are often estimated using the Nelson and Siegel (1987) model, as we do in this paper, as these authors discuss, cubics often have better fit within the sample range of maturities while being equally parsimonious. In practice, the most appropriate curve specification is likely to vary across applications. Finally, we show that the issues discussed in this paper are not unique to DiD analysis, but also to more simpler assignment specifications when the dependent variable has a term structure. Analogously to the DiD setup, in the simple assignment inference problem it is also essential to allow for heterogeneous effects over the maturity spectrum for results to be meaningful.

CRediT authorship contribution statement

Kjell G. Nyborg: Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Jiri Woschitz:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abadie, A., 2021. Using synthetic controls: Feasibility, data requirements, and methodological aspects. J. Econ. Lit. 59, 391–425.

Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. J. Amer. Statist. Assoc. 105, 493–505.

Abadie, A., Gardeazabal, J., 2003. The economic costs of conflict: A case study of the basque country. Am. Econ. Rev. 93, 113–132.

Adelino, M., Ferreira, M.A., 2016. Bank ratings and lending supply: Evidence from sovereign downgrades. Rev. Financ. Stud. 29, 1709–1746.

Allen, F., Gu, X., Li, C.W., Qian, J., Qian, Y., 2023. Implicit guarantees and the rise of shadow banking: The case of trust products. J. Financ. Econ. 149, 115–141.

Almeida, H., Cunha, I., Ferreira, M.A., Restrepo, F., 2017. The real effects of credit ratings: The sovereign ceiling channel. J. Financ. 72, 249–290.

Alon, T., Swanson, E., 2011. Operation twist and the effect of large-scale asset purchases. FRBSF Econ. Lett..

Amiram, D., Beaver, W.H., Landsman, W.R., Zhao, J., 2017. The effects of credit default swap trading on information asymmetry in syndicated loans. J. Financ. Econ. 126, 364–382.

Ang, A., Bhansali, V., Xing, Y., 2010. Taxes on tax-exempt bonds. J. Financ. 65, 565–601.

Athey, S., Imbens, G.W., 2022. Design-based analysis in difference-in-differences settings with staggered adoption. J. Econometrics 226, 62–79.

Ayotte, K., Gaon, S., 2011. Asset-backed securities: Costs and benefits of "bankruptcy remoteness". Rev. Financ. Stud. 24, 1299–1335.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229–1279.

Baker, A.C., Larcker, D.F., Wang, C.C.Y., 2022. How much should we trust staggered difference-in-differences estimates? J. Financ. Econ. 144, 370–395.

Bao, J., Hou, K., 2017. De facto seniority, credit risk, and corporate bond prices. Rev. Financ. Stud. 30, 4038–4080.

Bao, J., O'Hara, M., Zhou, X.A., 2018. The volcker rule and corporate bond market making in times of stress. J. Financ. Econ. 130, 95–113.

Ben-Michael, E., Feller, A., Rothstein, J., 2022. Synthetic controls with staggered adoption. J. R. Stat. Soc. Ser. B Stat. Methodol. 84, 351–381.

Benetton, M., Fantino, D., 2021. Targeted monetary policy and bank lending behavior. J. Financ. Econ. 142, 404–429.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? Q. J. Econ. 119, 249–275.

Bitler, M.P., Gelbach, J.B., Hoynes, H.W., 2006. What mean impacts miss: Distributional effects of welfare reform experiments. Am. Econ. Rev. 96, 988–1012.

Callaway, B., 2023. Difference-in-differences for policy evaluation. In: Zimmerman, K.F. (Ed.), Springer International Publishing, Cham, pp. 1–61.

Callaway, B., Li, T., 2019. Quantile treatment effects in difference in differences models with panel data. Quant. Econ. 10, 1579–1618.

Callaway, B., Li, T., 2023. Policy evaluation during a pandemic. J. Econometrics 236, 105454.

Callaway, B., Li, T., Oka, T., 2018. Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods. J. Econometrics 206, 395–413.

Callaway, B., Sant'Anna, P.H.C., 2021. Difference-in-differences with multiple time periods. J. Econometrics 225, 200–230.

Casella, G., Berger, R.L., 2001. Statistical Inference, second ed. Cengage Learning.

Chan, L.H., Chen, K.C.W., Chen, T.-Y., 2013. The effects of firm-initiated clawback provisions on bank loan contracting. J. Financ. Econ. 110, 659–679.

Chava, S., Livdan, D., Purnanandam, A., 2009. Do shareholder rights affect the cost of bank loans? Rev. Financ. Stud. 22, 2973–3004.

Choi, J., Hoseinzade, S., Shin, S.S., Tehranian, H., 2020. Corporate bond mutual funds and asset fire sales. J. Financ. Econ. 138, 432–457.

Cornaggia, J., Cornaggia, K.J., Israelsen, R.D., 2018. Credit ratings and the cost of municipal financing. Rev. Financ. Stud. 31, 2038–2079.

Dannhauser, C.D., 2017. The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs). J. Financ. Econ. 125, 537–560.

de Chaisemartin, C., D'Haultfoeuille, X., 2020. Two-way fixed effects estimators with heterogeneous treatment effects. Am. Econ. Rev. 110, 2964–2996.

Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. J. Econometrics 130, 337–364.

Elton, E.J., Gruber, M.J., Agrawal, D., Mann, C., 2001. Explaining the rate spread on corporate bonds. J. Financ. 56, 247–277.

Eser, F., Schwaab, B., 2016. Evaluating the impact of unconventional monetary policy measures: Empirical evidence from the ecb's securities markets programme. J. Financ. Econ. 119, 147–167.

Fama, E.F., Bliss, R.R., 1987. The information in long-maturity forward rates. Am. Econ. Rev. 77, 680–692.

Favara, G., Gao, J., Giannetti, M., 2021. Uncertainty, access to debt, and firm precautionary behavior. J. Financ. Econ. 141, 436–453.

Filipovic, D., Pelger, M., Ye, Y., 2023a. Shrinking the term structure.

Filipovic, D., Pelger, M., Ye, Y., 2023b. Stripping the discount curve - a robust machine learning approach.

Fisher, D., 1966. Expectations, the term structure of interest rates, and recent british experience. Economica 33, 319–329.

Fleckenstein, M., Longstaff, F.A., 2020. The US treasury floating rate note puzzle: Is there a premium for mark-to-market stability? J. Financ. Econ. 137, 637–658.

Fleckenstein, M., Longstaff, F.A., Lustig, H., 2014. The TIPS-treasury bond puzzle. J. Financ. 69, 2151–2197.

Gao, P., Lee, C., Murphy, D., 2020. Financing dies in darkness? The impact of newspaper closures on public finance. J. Financ. Econ. 135, 445–467.

Gobillon, L., Magnac, T., 2016. Regional policy evaluation: Interactive fixed effects and synthetic controls. Rev. Econ. Stat. 98, 535–551.

Goodman-Bacon, A., 2021. Difference-in-differences with variation in treatment timing. J. Econometrics 225, 254–277.

Gürkaynak, R.S., Sack, B., Wright, J.H., 2007. The U.S. treasury yield curve: 1961 to the present. J. Monet. Econ. 54, 2291–2304.

Hasan, I., Hoi, C.K.S., Wu, Q., Zhang, H., 2014. Beauty is in the eye of the beholder: The effect of corporate tax avoidance on the cost of bank loans. J. Financ. Econ. 113, 109–130.

Heckman, J.J., Smith, J., Clements, N., 1997. Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. Rev. Econ. Stud. 64, 487–535.

Helwege, J., Turner, C.M., 1999. The slope of the credit yield curve for speculative-grade issuers. J. Financ. 54, 1869–1884.

Huang, J.-Z., Huang, M., 2012. How much of the corporate-treasury yield spread is due to credit risk? Rev. Asset Pricing Stud. 2, 153–202.

Huang, J., Östberg, P., 2023. Difference-in-differences with economic factors and the case of housing returns.

Lentner, P., 2023. Price pressure during central bank asset purchases: Evidence from covered bonds.

Liu, Y., Wu, J.C., 2021. Reconstructing the yield curve. J. Financ. Econ. 142, 1395–1425.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. J. Bus. 60, 473–489.

Nyborg, K.G., Woschitz, J., 2021. The price of money: How collateral policy affects the yield curve.

Painter, M., 2020. An inconvenient cost: The effects of climate change on municipal bonds. J. Financ. Econ. 135, 468–482.

Pelizzon, L., Riedel, M., Simon, Z., Subrahmanyam, M.G., 2024. Collateral eligibility of corporate debt in the eurosystem. J. Financ. Econ. 153, 103777.

Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. Rev. Financ. Stud. 22, 435–480.

Qiu, J., Yu, F., 2009. The market for corporate control and the cost of debt. J. Financ. Econ. 93, 505–524.

Roberts, M.R., Whited, T.M., 2013. Endogeneity in empirical corporate finance 1. In: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), 2, Elsevier B.V., pp. 493–572.

Rodano, G., Serrano-Velarde, N., Tarantino, E., 2016. Bankruptcy law and bank financing. J. Financ. Econ. 120, 363–382.

Roth, J., Sant'Anna, P.H.C., Bilinski, A., Poe, J., 2023. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. J. Econometrics 235, 2218–2244.

Santos, J.A.C., Winton, A., 2008. Bank loans, bonds, and information monopolies across the business cycle. J. Financ. 63, 1315–1359.

Sarig, O., Warga, A., 1989. Some empirical estimates of the risk structure of interest rates. J. Financ. 44, 1351–1360.

S&P Global, 2023. Investor fact book.

Sun, L., Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. J. Econometrics 225, 175–199.

Svensson, L.E.O., 1994. Estimating and interpreting forward interest rates: Sweden 1992–1994. IMF Working Papers 94.

Swanson, E.T., 2011. Let's twist again: A high-frequency event-study analysis of operation twist and its implications for QE2. Brookings Pap. Econ. Act. 151–188.

Titman, S., Tsyplakov, S., 2010. Originator performance, CMBS structures, and the risk of commercial mortgages. Rev. Financ. Stud. 23, 3558–3594.

Todorov, K., 2020. Quantify the quantitative easing: Impact on bonds and corporate debt issuance. J. Financ. Econ. 135, 340–358.

Vayanos, D., Vila, J.-L., 2021. A preferred-habitat model of the term structure of interest rates. Econometrica 89, 77–112.

Xu, Y., 2017. Generalized synthetic control method: Causal inference with interactive fixed effects models. Political Anal. 25, 57–76.