Robust difference-in-differences analysis when there is a term structure[1]

Kjell G. Nyborg

University of Zurich,
Swiss Finance Institute,
and CEPR

Jiri Woschitz

BI Norwegian Business School

August 2023

[1]We gratefully acknowledge financial support from the Swiss National Science Foundation (project #100018_172679 "Trading and Financing during Market Stress"). We would also like to thank participants of the brownbag seminar at BI Norwegian Business School, the Financial Econometrics Conference to mark Stephen Taylor's Retirement in Lancaster, and the 2023 Financial Markets and Corporate Governance Conference at Deakin University, as well as Massimiliano Bondatti (discussant) for useful comments and suggestions. Nyborg: Department of Banking and Finance, University of Zurich, Plattenstrasse 14, CH-8032 Zurich, Switzerland. email: kjell.nyborg@bf.uzh.ch. Woschitz: Department of Finance, BI Norwegian Business School, B4y, NO-0442 Oslo, Norway. email: jiri.woschitz@bi.no.

Abstract

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

It is common practice in finance to use difference-in-differences analysis to examine fixed-income pricing data. This paper uses simulations to show that this method applied to pricing variables that exhibit a term structure, such as yields or credit spreads, systematically produces false and mismeasured treatment effects. This holds true even if the treatment is randomly assigned. False and mismeasured treatment effects result from heterogeneous effects in different parts of the term structure in combination with unequal distributions of residual maturities in the treated and control bond samples. Neither bond fixed effects nor explicit yield-curve control in the specification resolve the issues. By combining difference-in-differences analysis with yield-curve modeling this paper provides new methodology to overcome these issues.

JEL classification: C20, G12, E43, E47

Keywords: Fixed-income pricing, yield curve, term structure of interest rates, yields, credit spreads, difference-in-differences analysis, false and mismeasured treatment effects

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 the interest it pays, i.e. its yield (or credit spread), and DiD analysis is applied by running a classical DiD regression of the form

$$yield_{it} = \alpha_i + \delta_t + \beta_{DiD}\ \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \varepsilon_{it}, \tag{1}$$

where $yield_{it}$ is security i's yield-to-maturity at time t and the right-hand side of the equation is represented by the typical DiD structure: α_i and δ_t correspond to security- and time-fixed effects, respectively, $\mathbb{1}_{Treated,i}$ and $\mathbb{1}_{Post,t}$ to treatment and post-event date indicator variables, β_{DiD} is the treatment effect, and ε_{it} is an error term. DiD methodology is designed to deal with endogeneity, namely 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, fixed-income securities data have two ever-present features that severely inhibit the ability of the classical DiD specification in Equation (1) and variations thereof to accurately measure true treatment effects. As we will show, this specification is even so problematic that it is prone to leading researchers to conclude that there are treatment effects when there are, in fact, none. When there are true treatment effects, these are invariably mismeasured, leading to garbled inference. In this paper, we explain these problems and propose methods for overcoming them.

The first regular, but problematic, feature in fixed-income data is that pricing variables, such as yields or credit spreads, are time-varying issuer-specific functions of residual maturity. Fixed-income securities are priced against the term structure of interest rates, typically of the issuer and using its individual maturities, and pricing terms vary over time and heterogeneously over the maturity spectrum. For example, in a sample of twelve selected countries (details below), the standard deviation of the weekly change in the ten-year minus three-month term spread is 57.5 basis points, and it is rare for the term spread not to move from one date to another. Second, residual maturity is a continuous variable. In practice, it is hardly ever possible to match two same-issuer securities on maturity because individual issuers have relatively few outstanding securities over a wide range of maturities. For example, in January 2023, seven of said twelve countries have less than 100 securities issued, with residual maturities covering a range from zero to more than thirty years.

In this paper, we use simulations to show that, because of these two ubiquitous features of fixed-income pricing data, the classical DiD specification in Equation (1) applied to pricing variables that exhibit time-varying term structures systematically produces false and mismeasured treatment effects. This holds even if the treatment is assigned randomly. One may erroneously conclude that there is a statistically significant treatment effect when, in fact, a true treatment effect is nonexistent; or, conversely, one finds no treatment effect when in reality there is one. It is even possible to measure a significant *negative* treatment effect while the true impact is *positive*, or vice versa. The reason for false and mismeasured treatment effects is Specification (1)'s inability to properly control for term effects, which we elaborate on below. This is particularly unfortunate since the DiD approach is designed precisely for the purpose of dealing with endogeneity to quantify causal effects.

To show false and mismeasured treatment effects, we run Specification (1) under two types of true yield-curve effects, namely, treatment-unrelated idiosyncratic effects and true, systematic treatment 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. The analysis shows that, even in absence of a true underlying treatment effect, the classical DiD specification in Equation (1) measures a false treatment effect – "false" because, in fact, there is no treatment effect. Specification (1) gives wrong inference because it is misspecified. It assumes there is a fixed bond effect, when the bond effect, in fact, is variable as a function of its time to maturity and coupon structure and depends on treatment-unrelated idiosyncratic shifts in the underlying spot curve.

In contrast, systematic treatment effects only impact yields of treated bonds. We also allow these to be heterogeneous over the term structure. When there actually is a true treatment effect, Specification (1) remains problematic. First, there can still be treatment-unrelated idiosyncratic effects. Second, samples of treated and control bonds are highly unlikely to be matched on maturity and coupon structure. If there is a treatment effect only, Specification (1) generates a mismeasured treatment effect – "mismeasured" because it produces an average treatment effect that ignores maturity while the true treatment effect is maturity-dependent. As we show, this can confound inference and lead to incorrect conclusions. If there are both systematic and treatment-unrelated idiosyncratic effects, Specification (1) produces a combination of false and mismeasured treatment effects.

The limitation of the classical DiD specification arises from its inability to control for idiosyncratic and systematic effects that vary over the maturity spectrum. It is misspecified because it ignores a key feature of fixed-income data, namely, time varying term spreads. Instead, it essentially assumes that idiosyncratic and systematic effects are homogeneous across maturity. False and mismeasured treatment effects, as we show, result from heterogeneous true effects in different parts of the term structure combined with unequal distributions of residual maturity in the treated- and control-bond samples. This induces nonzero correlation between duration and the treatment indicator variable, confounding inference when treated and control bonds have different residual maturities and coupon structures.

Nonzero correlation between duration and the treatment indicator variable arises naturally in fixed-income settings because treated and control bonds are highly unlikely to be perfectly matched on residual maturity and coupon structure. In the context of zero-coupon bonds, even if residual maturities of treated and control bonds are drawn from the same distribution, a paucity of bonds in practice means that regular, idiosyncratic movements in the yield curve induces spurious correlation. In this case, the estimated treatment effect using Specification (1) is measured with imprecision, although not unconditionally biased. Because standard errors calculated the usual way do not control for the spurious correlation, however, t-statistics are overestimated in absolute value and inference is, accordingly, garbled. If the residual maturities of the two bond groups are drawn from different distributions, spurious correlation between residual maturity and the treatment indicator variable is compounded by systematic correlation. Combined with maturity-dependent idiosyncratic and treatment effects, this leads to a combination of bias and imprecision in the estimated treatment effect. The error can go either way. When there are only idiosyncratic effects, the null hypothesis of no treatment effect is systematically over-rejected.

Thus, it is important to control for term effects when using DiD analysis in fixed-income settings. However, 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 illustration, we analyze a specification that substitutes the bond fixed effects with a model of the yield curve that is consistent with that used to simulate the true underlying curves. It turns out that the DiD estimator in this case is essentially identical to what is produced by Specification (1). The problem with these two 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, the true underlying effects are heterogeneous over the term structure.

To deal with heterogeneous treatment effects the literature typically either estimates heterogeneous treatment effects over the distribution of the dependent variable or uses fixed effects on the discrete right-hand side units present in the data.[1] The former does not work well with yield as the dependent variable because the shape and location of the yield curve fluctuate over time. The latter is sometimes applied in fixed-income settings by using maturity buckets (e.g., Bao, O'Hara, and Zhou, 2018; Todorov, 2020). However, running Specification (1) on individual maturity buckets simply pushes the problems discussed above 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. Residual maturity in a fixed-income setting is a continuous habitat variable and, as such, requires a different approach.

The perfect solution would be to match each treated bond with a control bond having the same residual maturity and coupon structure. However, this is rarely feasible in practice. We provide two alternative approaches that are more practical. First, we replace the bond and time fixed effects in Specification (1) with separate parameterized yield curves for control and treated bonds both pre- and post-treatment. The treatment effect is now estimated as a "Delta curve," namely, as the incremental difference between the yield curves of treated bonds and controls over the event. With spot rates as the dependent variable, treatment effects at selected maturities can be estimated by running one single regression using standard software. We show that this "fully flexible yield-curve DiD specification" resolves the problems of both false and mismeasured treatment effects. If there are true, maturity-dependent treatment effects, these are identified and separated from maturity-dependent treatment-unrelated idiosyncratic yield-curve effects. Furthermore, since the specification uses the full panel structure of the data, it permits clustering standard errors

[1] To name a few, for the former see Heckman, Smith, and Clements (1997), Bitler, Gelbach, and Hoynes (2006), Callaway and Li (2019), the latter de Chaisemartin and D'Haultfœuille (2020), and both in one Callaway, Li, and Oka (2018).

at the bond level as recommended by Bertrand, Duflo, and Mullainathan (2004).

Our second approach is what we call "semi-matching" (or, synthetic matching). This essentially matches treated bonds with synthetic control bonds having the same residual maturity. We compare and contrast these two methods and show that they are identical when the synthetic matching approach involves a second stage where curve fitting is applied to the individual bond-level difference-in-differences. Thus, we show that there are two ways to generate the DiD Delta curve over the term structure.

The paper relates to several literatures. First and foremost, it contributes to a large finance literature as papers aim to use DiD analysis to measure treatment effects with dependent variables that exhibit time-varying term structures. We show false and mismeasured effects in the most trivial setting with security level data, yield as dependent variable, and security fixed effects on the right-hand side of the regression equation. In this form, this relates already to many analyses in finance; it affects many different types of fixed-income securities such as bonds, bills, notes, loans, asset-backed securities, mortgage loans, etc. and both in primary and secondary markets. Importantly, however, false and mismeasured treatment effects survive (1) if the unit of analysis is not the security level but an aggregation of it such as the firm, the country, or the bank-firm relationship, (2) with other dependent pricing variables such as expected returns, loan rates or spreads, yield spreads, logarithms of these variables, etc. and (3) if, instead of bond fixed effects, the right-hand side of the regression equation explicitly controls for maturity or functions of it.[2] To give a feeling of how widespread these issues are, Table 1 provides a list of relatively recent top finance publications that use the DiD method in ways potentially affected.

Insert Table 1 here.

Second, a large literature in finance attempts to estimate the unobserved yield curve parameters (e.g., Nelson and Siegel, 1987; Svensson, 1994; Liu and Wu, 2021). Because these parameters can be interpreted as unobserved factors (Diebold and Li, 2006) and DiD analysis is a special case of fixed-effects setting, our paper relates to a literature on con-

[2]Regarding (2), any pricing variable that exhibits a time-varying term structure can lead to false and mismeasured treatment effects. Loan or yield spreads are typically calculated with maturity-matched interpolated LIBOR or treasury rates. Bao and Hou (2017), for example, show that an issuer's relatively longer-dated bonds have larger yield spreads and more co-movement with the issuer's equity and, hence, provide evidence for a term structure in yield spreads rather than in yields. See also John, Lynch, and Puri (2003), Chava, Livdan, and Purnanandam (2009), Ayotte and Gaon (2011).

founding factor structures in fixed-effects settings (Bai, 2009). Gobillon and Magnac (2016) and Xu (2017) show that the DiD effect is biased if one ignores correlation between factors and the treatment. Our contribution is twofold. First, we demonstrate that heterogeneous idiosyncratic yield-curve effects over maturity ("factors") *naturally* confound the estimation of Specification (1) in the context of fixed-income pricing data. Second, the proposed approaches overcome false and mismeasured treatment effects because they accurately deal with heterogeneous term effects ("factors"). Both approaches make use of the rich yield-curve fitting literature in finance to identify the true effects but they do so in different ways. The fully flexible yield-curve DiD method, our first approach, models flexible yield curves as part of the DiD estimator and does not only control for "factors" but, simultaneously, provides maturity-dependent treatment effects. As we show, the latter is relevant because Specification (1) produces mismeasured treatment effects even after controlling for "factors." Semi-matching, our second approach, instead, removes the "factor structure" before applying DiD, as in the synthetic control literature (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010; Abadie, 2021).

Third, the paper relates to a large literature that more generally shows how DiD methodology can lead to incorrect measurements of effects, as synthesized, for instance, by Callaway (2023) and Roth, Sant'Anna, Bilinski, and Poe (2023). Specification (1) assumes a fixed bond effect, when the bond effect, in fact, is variable as a function of maturity, coupon structure, and idiosyncratic yield-curve effects, which is a misspecification and leads to incorrect inference. In that vein, Callaway and Tong (2023), for example, show that the DiD method is less suitable to conduct policy evaluation during a pandemic because the unit-level fixed effect, in fact, is not fixed but highly nonlinear. In our case, we can make use of the rich curve fitting literature in finance to identify the true effects. Furthermore, Imbens and Wooldridge (2009) discuss the overlap assumption, which, in our case, implies that the support of the residual maturity distribution is the same for control and treated sample bonds. We show that Specification (1) leads to false and mismeasured treatment effects almost surely due to small bond samples and wide maturity ranges even if the overlap assumption holds. Moreover, mismeasured treatment effects have been shown in the context of staggered DiD analysis with heterogeneous treatment effects in the time dimension.[3] We examine heterogeneous treatment effects in the cross-sectional dimension, which

[3]See, for instance, Sun and Abraham (2021), Callaway and Sant'Anna (2021), Goodman-Bacon (2021),

we can address by modeling the spot yield curve.

The rest of the paper is structured as follows. Section 2 presents the term structure modeling and the data simulation. Section 3 shows how Specification (1) can lead to false treatment effects. Section 4 provides methodology to overcome false treatment effects. Sections 5 and 6 show the cases of mismeasured as well as false and mismeasured treatment effects combined, respectively. Section 7 provides further methodology to resolve false and mismeasured treatment effects. Section 8 concludes.

2. Term structure modeling and data simulation

Our approach is to posit idiosyncratic (treatment unrelated) as well as systematic treatment effects in the underlying spot yield curve (pricing kernel) and, then, examine the performance of Specification (1) and its alternatives on random samples of bonds. For simplicity and to isolate the effect of time to maturity, we use zero-coupon bonds. Thus, the cash flow structure of the bonds are completely captured by their respective times to maturity. We then draw random samples of residual maturities separately for treated and control bonds. Since the feature of real data that we want to capture is heterogeneity in the distribution of residual maturities between treated and control bonds, we also allow for treated and control bonds to be drawn from different distributions. In this section, to generate the data, we focus on the two features that, combined, lead to false and mismeasured treatment effects, which are heterogeneous idiosyncratic or actual treatment yield-curve effects over maturity and unequal residual maturity distributions in the control and treated bond samples.

2.1 Modeling term structure effects

In practice, fixed-income securities are priced against the term structure of interest rates, typically of the issuer, using its individual maturities. Thereby, pricing terms vary over time and heterogeneously so in different parts of the term structure. Therefore, pricing variables such as yields or credit spreads, which is what Specification (1) has on its left-hand side, exhibit issuer-specific term structures whose shapes change over time. This feature, which is inherent to fixed-income pricing data, is shown in Table 2. Using yield curve data from

Athey and Imbens (2022), Baker, Larcker, and Yang (2021).

Bloomberg from January 2000 to December 2022 for a selected group of countries, the table shows distributions of daily and monthly changes in the ten-year minus three-month term spread in Panels A and B, respectively.

Insert Table 2 here.

In Table 2 the twelve countries are ordered according to the range in the change of the term spread. For example, changes in daily term spreads vary from ±30 basis points (bps) in Japan to between −210 to +190 bps in China.[4] In Spain, which is one of the two countries in the middle of the list, daily (monthly) term spread changes vary between −121 and +63 (−143 and +137) bps, which shows that yield curves are issuer-specific and vary heterogeneously along maturity over time. The standard deviation of the change in the term spread in the pooled sample of countries is 87.9 bps with monthly data, 57.5 bps with weekly data (not reported in Table 2), and 27.4 bps with daily data.

Motivated by the magnitudes of yield curve movements observed in practice (Table 2), we consider two types of yield-curve effects. First, idiosyncratic yield-curve effects are unrelated to the treatment. These are idiosyncratic movements in the term structure that result from economic forces other than the treatment. They move "treated" and control bonds from pre- to post-treatment irrespective of the treatment.[5] Second, following the same logic but in the absence of treatment-unrelated idiosyncratic yield-curve effects, we model heterogeneous yield-curve treatment effects. From pre- to post-treatment, they affect only the treated bonds but heterogeneously over maturity.

To model heterogeneous idiosyncratic and treatment yield-curve effects we employ Diebold and Li's (2006) factorization of the Nelson-Siegel (1987) term structure parameterization.[6] The spot, or zero-coupon, rate, or *yield*, with maturity x at time t is

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

where $\gamma_{0,t}$ is a long-term or level factor, $\gamma_{1,t}$ a short-term or slope factor, $\gamma_{2,t}$ a medium-term or curvature factor, and λ_t the decay parameter. To model effects in the term structure, we

[4]In Greece, a country that was hit exceptionally hard during the European sovereign debt crisis, daily term spreads move exceptionally between −2,012 and 1,822 bps.

[5]For example, Foley-Fisher, Ramcharan, and Yu (2016) show how the Fed's maturity extension program depresses yields of long-term but not short-term bonds.

[6]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.

manipulate the level, slope, and curvature parameters while following Diebold and Li (2006) and holding the decay parameter fixed, namely $\lambda_t = \lambda = 0.7308$.[7]

Exhibit 1: Overview effects and sections discussing them

		Treatment effect	
		No	Yes
Idiosyncratic effect	No	–	Section 5
	Yes	Section 3	Section 6

Inference using Specification (1) will be erroneous. We show this by, first, allowing for idiosyncratic yield-curve effects only, then for systematic treatment effects only, and finally we combine the two types of effects. Exhibit 1 provides an overview of the modeled effects. First, we model two idiosyncratic yield-curve effects. The idiosyncratic short-end effect corresponds to a yield reduction of −50 bps at a residual maturity of one year and an effect close to zero (+1 bps) at fifteen years (details will be provided later). Contrarily, the long-end effect amounts to a yield decline of −50 bps at fifteen years and an effect close to zero (+4 bps) at one year. As a preview, the analysis in Section 3 shows that the classical DiD specification produces potentially large, statistically significant treatment effects even if a true treatment effect is entirely absent from the data.

Second, we model two systematic treatment effects. Either the treatment leads to a yield-curve twist, which pushes up (down) the yields of treated bonds at the one-year (ten-year) maturity by +6 (−6) bps, or the treatment affects treated bonds only at the short-end with −6 (0) bps at the one-year (ten-year) maturity. By choosing relatively small treatment effects, we mimic reality where treatment effects are typically smaller than the idiosyncratic yield-curve effects. As shown in Section 5 and much along Kahn and Whited (2018), the classical DiD specification produces an average treatment effect, which is a quantity that ignores an important dimension in this case, namely the spot curve, or the pricing kernel. Because the true treatment effect is dependent on maturity, this quantity is "mismeasured" as it may lead to incorrect conclusions. For example, if the treatment twists the curve for treated bonds (short-end up, long-end down) and, incidentally, treated bonds pile up at the long-end, one finds a negative treatment effect while the true effect is positive at the short-end. The mismeasurement depends on the distribution of the treated bonds over maturity as the treatment effect itself depends on maturity.

[7]The authors explain that λ_t determines the point where the loading on the curvature factor, $\gamma_{2,t}$, obtains its maximum and pick this, based on practice, to be at a maturity of 30 months. If maturity is measured in months $\lambda_t = \lambda = 0.0609$, which translates to $\lambda = 0.7308$ if maturity is measured in years.

Third, in Section 6 we combine idiosyncratic and treatment effects (four combinations). Heterogeneous yield-curve effects along maturity, either idiosyncratic or treatment effects or both, are the first critical feature that leads to false and mismeasured treatment effects.

2.2 Simulation of residual maturity

In practice control and treated bonds are often distributed differentially over maturity. This property arises naturally in fixed-income data as residual maturity is a continuous variable; individual issuers inherently issue only a limited number of securities but the range of residual maturity is large. As an illustration, Table 3 provides the number of securities together with the percentage of debt outstanding by maturity bucket for the same twelve countries as used previously (Panel A: January 2023; Panel B: January 2011).

Insert Table 3 here.

The table illustrates two key aspects. First, across panels, the number of securities lies between 16 in Ireland and 559 in Japan and is relatively small compared to the wide maturity range, which lies between zero and more than thirty years. In 2023, seven of the twelve countries have less than 100 securities outstanding. Second, the countries have relatively more short-dated debt but the exact maturity structure is issuer- and time-specific. For example, while the US' maturity structure is tilted toward the short-end, the UK's is tilted toward the (5-10]-year bucket. The Netherlands has short maturities in 2011 but rather uniformly distributed ones in 2023. Non-governmental issuers typically issue even fewer securities. The small number of securities per issuer in conjunction with wide maturity ranges make it difficult to maturity-match two same-issuer securities. To work with non-matched samples and neglect maturity is inherently prone almost surely to unequal maturity distributions in the treated and control bond samples.

To analyze the performance of Specification (1), our goal is to mimic maturity structures realistically. We use simulations to emulate bond maturity. Motivated by the maturity structures in Table 3, in particular, we draw residual maturity from a triangular probability density function (pdf), which is a continuous pdf and, hence, suitable in simulating residual maturity. Moreover, the triangular pdf enables us to generate residual maturity distributions similar to the maturity structures in Table 3. To be specific, we draw residual time-to-maturity, x, for control and treated bonds from triangular pdfs, $p(x; m)$, that range from zero to twenty years, $x \in [0, 20]$, and have mode m:[8]

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

The mode m parameterizes the feature of real data that we want to capture, namely the location of the "peak" of the triangular pdf that allows us to create heterogeneity in the distributions of residual maturities between control and treated bonds. For controls we use $m = 0.25$ years. For treated bonds we use $m = 0.25, 1, 3$ and 10 years. In particular, we repeat the following procedure 1,000 times:

1. Control bonds: Draw 50 maturities from $p(x; m = 0.25)$,

2. Treated bonds: Draw 50 maturities from each of $p(x; m)$, $m = 0.25, 1, 3, 10$.

Thus, we have 1,000 families f, each comprised of 50 control bond maturities and four times 50 treated bond maturities. For each family, f, we then create four sample couplets, each comprised of the 50 control-bond maturities, where $m = 0.25$, and the four times 50 treated bond maturities based on $m = 0.25, 1, 3,$ and 10 years. The idea is to examine inference mistakes as the underlying distributions of control and treated bond maturities get more different by increasing m for the treated bonds as illustrated in Figure 1. However, even if treated bonds are based on $m = 0.25$ years, just as the control bonds, the realized control and treated bond maturities will differ almost surely as, inherently, there is a paucity of bonds given the wide possible maturity range. This by itself creates inference mistakes (with non-trivial probability), which are exacerbated when the underlying maturity distributions of the treated bonds move away from that of the control bonds.

Insert Figure 1 here.

For illustration purposes, in Table 4 we compute the average of residual maturity of the fifty bonds in each sample and calculate for each sample couplet its ratio (average residual maturity of treated bonds divided by that of controls). Panels A and B show population means as well as distributions, respectively, of the sample averages and the average-maturity ratios across the 1,000 families by treatment group and mode m.

[8]Notice that for countries the upper limit of twenty years is a rather conservative choice (see Table 3).

Insert Table 4 here.

If residual maturities of control and treated bonds are drawn from the same triangular pdf ($m = 0.25$), by definition the unconditional population ratio in Panel B is one as the unconditional population means in Panel A are the same (6.75 years). With increasing mode for the treated bonds the unconditional population ratio in Panel B rises above one. While means and medians of the residual-maturity averages conditional on the samples in Panel A and their ratios in Panel B are very similar to the unconditional population means, importantly, the dispersion is fairly extreme. For example, if residual maturity is drawn from triangular pdfs with m equals to 0.25 years, the average-maturity ratios conditional on the samples in Panel B vary by between 0.59 to 1.52 significantly around the population mean of one. This is an artifact of drawing relatively small samples from wide ranges of maturities, as seen in Table 3.[9]

The unequal maturity distributions of the control and treated samples bonds, in turn, lead to a critical property. Panel C shows the distribution of the correlation between residual maturity and the treatment indicator variable over the 1,000 families by m. If residual maturity of control and treated bonds is drawn from the same $p(x, m = 0.25)$, the correlation distribution exhibits large dispersion that ranges from -0.349 to 0.318 (while means and medians are close to zero). Drawing relatively small samples from wide maturity ranges introduces spurious correlation between residual maturity and the treatment. Instead, if the treated bonds' residual maturities are tilted toward the long-end, i.e. if m increases, the distributions are still dispersed but now means and medians of the correlations lie systematically above zero. Hence, drawing residual maturities of control and treated bonds systematically from different distributions introduces spurious combined with systematic correlation. This correlation between residual maturity and the treatment indicator variable is the second critical feature that leads to false and mismeasured treatment effects.

The following sections make use of the simulated maturity data and the modeled term structure effects to analyze the performance of Specification (1) and its alternatives.

[9] Given the maturity structures in Table 3, the triangular pdf seems a reasonable benchmark from which to draw residual maturity. However, if anything, this choice is conservative as the dispersion in the average-maturity ratio would increase if residual maturity is drawn, for example, from a uniform distribution.

3. False treatment effects

In this section we run Specification (1) on the simulated data when the term structure exhibits heterogeneous idiosyncratic effects across maturity, which are not related to the treatment whatsoever; true treatment effects are entirely absent from the data. We analyze the simplest possible case, namely when there are only two time periods, labeled "pre-" and "post-treatment." To model the scenarios, we use Diebold and Li's (2006) yield curve specification in (2). In either scenario, each period has its own yield curve.[10]

Insert Figure 2 here.

Figure 2 provides the underlying yield curve parameter values used to build the curves together with the resulting yield levels and differences at selected maturities and graphically illustrates the two scenarios.[11] From pre to post treatment, the idiosyncratic effect either pushes down the yield curve particularly at the short-end with -50 ($+1$) bps at the one-year (fifteen-year) maturity or particularly at the long-end with $+4$ (-50) bps, respectively.

3.1 Main result: False treatment effects

We estimate treatment effects by running Specification (1) using ordinary least squares methodology (OLS). Since Bertrand, Duflo, and Mullainathan (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, standard errors are clustered at the individual bond level. In our case, this does not matter as we have "perfect" draws and only two time periods. But in practice it might matter and, thus, the possibility to be able to cluster is an attractive feature of Specification (1).

Figure 3 provides first results. Each graph plots the 1,000 estimated DiD effects against its t-statistics. The first (second) row of graphs covers the idiosyncratic short-end (long-end) effect and graphs on the left (right) the case when $m = 0.25$ ($m = 10$) years.

Insert Figure 3 here.

[10] Logic: Prior to treatment, control and treated bonds share the same yield curve (pre-treatment curve). Since there is no treatment effect, control and treated bonds also share the same yield curve after treatment but the idiosyncratic effect has shifted it to a different location (post-event curve).

[11] With the chosen parameter values the curves are upward sloping but the argument is independent of the shape of the yield curve. False treatment effects can also be shown for downward sloping or flat curves.

The figure illustrates that the classical DiD specification generates false treatment effects almost surely if there are heterogeneous idiosyncratic effects along maturity. Figures 3a and 3c show that even if residual maturity of the control and treated bonds is drawn from the *same* underlying triangular pdf with m equals to 0.25 years, the specification produces false treatment effects. The effects can go in either direction and the larger in absolute value, the more likely they are statistically different from zero. The latter results from underestimated standard errors as these are calculated based on the conditional sample distribution and not on the unconditional population distribution. Across the 1,000 families and in case of an idiosyncratic yield-curve short-end (long-end) effect, the false treatment effects range from -11.59 to 10.85 (-11.34 to 12.01) bps and 91 (88) of them are statistically significant at the 10%-level. Notice, however, that mean and median of the distributions for either type of idiosyncratic effect are close to zero. Hence, what we observe is not a bias but imprecision in the measured treatment effects and its t-statistics, as we elaborate on shortly.

What happens, however, if residual maturities of the treated bonds are drawn from a triangular pdf with m equals to 10 years? In case of an idiosyncratic yield-curve short-end (long-end) effect all the 1,000 estimated treatment effects are positive (negative), range from 2.30 to 22.99 (-24.06 to -2.59) bps, and 992 (991) of them are statistically different from zero at the 10%-level as illustrated in Figure 3b (3d). However, now mean and median of the distributions across the 1,000 coefficients are not close to zero. In case of an idiosyncratic short-end (long-end) effect mean and median are positive (negative) and statistically different from zero. Thus, the imprecision in measured treatment effects described above is now compounded with a bias.

As will be shown next, these issues result from a combination of heterogeneous idiosyncratic effects in different parts of the term structure and unequal residual maturities in the samples of control and treated bonds.

3.2 The main driver of false treatment effects

A key driver of false treatment effects is the average-maturity ratio of treated sample bonds relative to controls and the implied correlation, either spurious or spurious combined with systematic correlation, between residual maturity and the treatment. To illustrate this we proceed as follows. For each m we sort the 1,000 families of sample couplets on the average-maturity ratio and index them in ascending order from 1 to 1,000. Table 5 then shows the implied correlation between residual maturity and the treatment indicator variable as well as the estimated treatment effects for a selection of nine sample couplets. For each m, the first and the last sample couplets in the ordered distribution, with order index $j = 1$ and $1,000$, are the sample couplets with the minimum and maximum average-maturity ratio, respectively. In between, with order indices $j = 10, 50, 250, 501, 751, 951, 991$, are the samples couplets with average-maturity ratio at, approximately, percentile $j/10$.

Insert Table 5 here.

Panel A provides the average-maturity ratio for the nine selected sample couplets by m. From the upper left corner (order index 1 and $m = 0.25$) to the lower right corner (order index 1,000 and $m = 10$) this ratio tends to go up.[12] Panel B shows the implied correlation between residual maturity and the treatment indicator variable for the nine selected sample couplets by m. If m equals to 0.25 years the correlation is negative for about the first half of the 1,000 sample couplets (with $j = 1, ..., 501$) and positive for the second half ($j = 502, ..., 1,000$). This is what we refer to as spurious correlation. If m increases the correlation tends to become positive but still exhibits the large dispersion, which is what we refer to as spurious combined with systematic correlation.

Panels C and D show the DiD coefficients and, underneath in parentheses, the associated p-values for each of the nine selected sample couplets by m. In case of an idiosyncratic yield-curve short-end (long-end) effect in Panel C (D), the estimated DiD effects tend to increase (decrease) with the average-maturity ratio in Panel A and the implied maturity-treatment correlation in Panel B both within and across m. The table shows that the false treatment effects result from unequal distributions of residual maturity in the treatment group samples and the implied maturity-treatment correlations and heterogenous idiosyncratic effects over the term structure. For example, for the same sample couplet, e.g. if $j = 1$ and $m = 0.25$, the false treatment effect is tilted away from the true, unconditional zero-effect in opposite direction for an idiosyncratic effect at the short-end (-11.59 bps) compared to at the long-end ($+12.01$ bps). This also illustrates that, conditional on what part of the term structure moves, the bias will be conditional on the selected bonds.

Furthermore, the table highlights that larger coefficients in absolute value, which are tilted away more from the true, unconditional zero-effect, are more often statistically signif-

[12] Panel A of Table 5 is essentially just a more granular depiction of Panel B in Table 4.

icant. In Table 5, coefficients that are statistically significant at the 10% level are marked in bold and the letters a, b, and c indicate significance at the 1%, 5%, and 10% levels, respectively. The table shows that the larger is the average-maturity ratio in Panel A and the implied maturity-treatment correlation in Panel B, both within and across m, the more likely the false treatment effect is statistically different from zero. The reason is that, based on standard methods, the standard errors are underestimated as they are calculated using the conditional sample distribution and not the unconditional population distribution.

The analysis so far illustrates that the classical DiD specification applied to fixed-income securities data with yield as dependent variable produces false treatment effects. Even in the absence of a true treatment effect the specification generates potentially economically large and statistically significant but false treatment effects. They can go in either direction, which is dependent on the combination of heterogeneous treatment-unrelated idiosyncratic effects in different parts of the term structure and unequal distributions of residual maturity in the samples of control and treated bonds. The latter introduces spurious or even a combination of spurious and systematic correlation between residual maturity and the treatment indicator variable which leads, respectively, to imprecise or even a combination of imprecise and biased false treatment effects. Furthermore, when using standard methods, the larger is the false treatment effect in absolute value, the more likely it also is (erroneously) statistically different from zero. The reason is that the standard errors are underestimated as they are not calculated unconditionally but conditionally on the sample.

3.3 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. Applied to fixed-income securities with yield as dependent variable, researchers sometimes measure the DiD effects separately by individual maturity buckets (see, for example, Bao, O'Hara, and Zhou, 2018; Todorov, 2020).

Table 6 shows the results if we run Specification (1) on four buckets with residual maturities in the ranges $[0, 2]$, $(2, 5]$, $(5, 10]$, and $(10, 20]$ years. Panels A and B cover the cases of m equals to 0.25 and 10 years, respectively. Each panel provides the distributions of estimated treatment effects by the idiosyncratic effect and maturity bucket as well as separately for statistically significant and non-significant coefficients (two-sided 10%-level).

Insert Table 6 here.

The results show that running Specification (1) separately for individual maturity buckets does not eliminate false treatment effects. In our case, the false treatment effects are the largest in the $(2, 5]$-year maturity bucket. For example, if there is an idiosyncratic yield-curve short-end effect and m equals to 0.25 years, as in Panel A, across the 1,000 estimated treatment effects $53 + 47 = 100$ range from -7.49 to -3.31 and from $+3.55$ to $+8.47$ bps and are statistically significant at the 10%-level. The remaining 900 coefficients are also different from zero but not statistically significant. Mean and median of the distributions across the 1,000 estimated treatment effects are, however, close to zero. This shows that the false treatment effects are not the result of a bias but of imprecise measurement. The reason is that Specification (1) ignores maturity and the unequal residual maturity distributions of control and treated sample bonds introduces spurious correlation between residual maturity and the treatment indicator variable also on the maturity-bucket level.

However, if m increases to 10 years, as in Panel B, in 2 cases it is not possible to estimate effects in the $(2, 5]$-year bucket because of no treated observations. Out of the remaining 998 coefficients $25 + 211 = 236$ are statistically significant and they range more extremely from -10.51 to -3.32 and from $+3.26$ to $+11.89$ bps. This time neither mean nor median of the distributions across the 998 coefficients are zero. The reason is that, this time, the false treatment effects are the result of spurious combined with systematic correlation between residual maturity and the treatment and are, therefore, not only imprecisely measured but also biased. How large the systematic correlation between residual maturity and the treatment is depends on the relative probability masses of the triangular pdfs of control and treated bonds in the $(2, 5]$-year maturity bucket. In this case, as seen in Figure 1, the systematic correlation is positive, which, with an idiosyncratic yield-curve short-end (long-end) effect, on average, leads to positive (negative) false treatment effects.

In case of an idiosyncratic yield-curve long-end effect the results are similar but, if anything, even more extreme. Importantly, a specification that is immune against bias and imprecise measurement and, therefore, would eliminate false treatment effects, should elicit effects of zero in all of these cases.

The analysis illustrates that maturity-bucket controls do not resolve the measurement of false treatment effects but shift the issue to the maturity-bucket level. Moreover, taking this approach further and choosing maturity buckets of shorter length leads, in the extreme case, to many infinitesimal short maturity buckets and no or only very few bonds remain in each bucket. Hence, because residual maturity is a continuous habitat variable we would prefer to work with an approach that acknowledges maturity-dependent effects.

3.4 Naive specification adjustment: Explicit yield curve control

Instead of using Specification (1) and controlling for bond specific characteristics via bond fixed effects, researchers sometimes use DiD specifications that explicitly control for bond maturity or functions of it (see Table 1). In this section we study the, in our case, most extreme version of this, namely explicit parametric control for the term structure itself. Moreover, we use the exact same yield-curve specification in the estimation as used to model the true underlying yield-curve effects. The question is whether such explicit term-structure control enables a DiD specification to elicit the true underlying effects or, at least, reduce the size of false treatment effects. Specifically, we run

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

where the notation is as above, α (δ) is the parameter that corresponds to $\mathbb{1}_{Treated,i}$ ($\mathbb{1}_{Post,t}$), and the new term $\mathbf{B}'\mathbf{L}_{it}$ explicitly controls parametrically for the yield curve. As alluded to, we continue to employ Diebold and Li's (2006) term structure specification in (2) and, hence, we use the exact same yield-curve specification in the estimation as used to model the true underlying yield-curve effects. $\mathbf{L}_{it}$ is a three-dimensional vector with elements 1, $l_1(x_{it};\lambda)$, and $l_2(x_{it};\lambda)$, with the latter two terms defined as

$$l_{1,t}(x;\lambda) = \left(\frac{1-e^{-\lambda x}}{\lambda x}\right) \quad \text{and} \quad l_{2,t}(x;\lambda) = \left(\frac{1-e^{-\lambda x}}{\lambda x} - e^{-\lambda x}\right), \tag{5}$$

$\mathbf{B}$ the corresponding vector of coefficients with individual elements β_k, $k = 0, \ldots, 2$, and the decay parameter λ assumed to be independent of time t.

In terms of estimation, since we know the underlying parameter value of lambda, which is $\lambda = 0.7308$, we could simply plug it into the expressions in (5) and use OLS to run Specification (4). Instead, we use nonlinear least squares methodology (NLS) to estimate λ in-sample together with the other parameters and take, as start value, $\lambda_{Seed} = 1$. Standard errors are clustered at the bond level, which, in our case, however, makes no difference.

It turns out that the specification with explicit term structure control in (4) produces the exact same false treatment effects as the classical DiD specification in (1). As illustration we compute the difference between the coefficients estimated with Specifications (1) and (4) for the two idiosyncratic yield-curve effects and the 4,000 sample couplets (four modes and 1,000 families). Across the 8,000 differences in DiD coefficients between Specifications (1) and (4) minimum and maximum amount to -0.0000 and 0.0000 bps, respectively.[13] Hence, the two specifications produce the exact same false treatment effects.

The fact that we use the same yield-curve specification in the estimation as used to model the true underlying yield curves illustrates that it is not the yield-curve specification that creates the problems but the way it is embedded in the regression specification. The regression specification in (4) explicitly controls for the yield curve parametrically via $\mathbf{B}'\mathbf{L}_{it}$. However, $\mathbf{B}'\mathbf{L}_{it}$ just removes the average term structure in the pooled data (treated, controls, pre-, post-treatment). The specification therefore restricts yield-curve movements between the different groups to parallel yield-curve level-shifts. While this feature is explicit with $\mathbf{B}'\mathbf{L}_{it}$, the classical DiD specification imposes the same parallel level-shift restrictions more implicitly through the bond fixed effects If, however, the idiosyncratic yield-curve effects come as movements other than parallel level-shifts and, simultaneously, the samples are affected by spurious or even spurious combined with systematic maturity-treatment correlation, then either of the Specifications (1) and (4) produce false treatment effects. Put differently, as these specifications restrict idiosyncratic yield-curve effects to parallel level-shifts they are misspecified when the true underlying effects are not limited to parallel level-shifts. The results illustrate that these specifications are not suitable when there are heterogeneous idiosyncratic effects in different parts of the yield curve.

The next section provides methodology to deal with heterogeneous idiosyncratic yield-curve effects along maturity and, therefore, overcomes false treatment effects.

4. A solution: Flexible yield-curve DiD specification

As touched on in the Introduction, a simple solution to the problem would be to perfectly match each treated sample bond with a control bond on residual maturity. This approach

[13] We have also calculated the 8,000 differences in p-values, which range from -0.0049 to -0.0000, showing that Specification (4) is slightly more conservative than (1).

eliminates one of the two components that lead to false treatment effects, namely unequal maturity distributions in the samples of control and treated bonds. However, in the context of fixed-income securities data, perfect matching on maturity is rarely feasible in practice because individual issuers only issue relatively few securities compared to the wide maturity ranges. In Section 7 we approach the challenge from this side and combine a matching procedure with yield-curve modeling.

In this subsection, however, we provide an alternative approach to overcome false treatment effects. The goal is to accurately deal with the other relevant component, namely heterogeneous idiosyncratic effects along the term structure. The limitation of the specifications discussed so far, which ultimately leads to false treatment effects, is its inability to control for idiosyncratic yield-curve effects if those are heterogeneous over maturity. Therefore, false treatment effects are the result of a misspecification, namely assuming that treatment-unrelated yield-curve effects are homogeneous along maturity. The solution provided in this section combines the DiD method with fully flexible yield curves.

The fully flexible yield-curve DiD specification, first used by Nyborg and Woschitz (2023), does not impose any particular relation between pre- and post-curves of treated and control bonds. On the contrary, the specification implicitly estimates yield curves separately for each group (control, treated, pre-, post-treatment) and uses those to quantify the DiD in yield curves. Specifically, the fully flexible specification is

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

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 curve, $\mathbf{B}_j'\mathbf{L}_{it}$, with $j = 1, \ldots, 4$ and three individual coefficients each, $\beta_{k,j}$, $k = 0, \ldots, 2$. For simplicity, the fourth parameter, λ, is assumed to be time-invariant and the same for treated and control bonds.[14]

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

$$s^{dl}(x; \widehat{\lambda}) = \widehat{\beta}_{0,1} + \widehat{\beta}_{1,1} \ l_1(x; \widehat{\lambda}) + \widehat{\beta}_{2,1} \ l_2(x; \widehat{\lambda}), \quad (7)$$

where $\{\widehat{\beta}_{k,1}\}_{k=0}^{2}$ are the estimated regression coefficients, x is residual maturity, and l_1 and l_2 are as in (5) with λ replaced by $\widehat{\lambda}$. The other three curves are either differences curves,

[14] This assumption can easily be relaxed.

namely when $j = 2$ or 3, or the difference-in-differences curve, when $j = 4$, given by

$$\Delta_j^{dl}(x; \widehat{\lambda}) = \widehat{\beta}_{0,j} + \widehat{\beta}_{1,j}\ l_1(x; \widehat{\lambda}) + \widehat{\beta}_{2,j}\ l_2(x; \widehat{\lambda}), \tag{8}$$

where $\{\widehat{\beta}_{k,j}\}_{k=0}^{2}$ are the estimated regression coefficients. The differences and difference-in-differences curves capture the incremental differences for i) treated bonds ($j = 2$), ii) the post-treatment period ($j = 3$), and iii) treated bonds over the post-treatment period ($j = 4$). Adding them to the spot curve for control bonds pre treatment, $s^{dl}(x; \widehat{\lambda})$, returns the spot curve, respectively, for i) treated bonds pre treatment, $s^{dl}(x; \widehat{\lambda}) + \Delta_2^{dl}(x; \widehat{\lambda})$, ii) control bonds post treatment, $s^{dl}(x; \widehat{\lambda}) + \Delta_3^{dl}(x; \widehat{\lambda})$, and iii) treated bonds post treatment, $s^{dl}(x; \widehat{\lambda}) + \sum_{j=2}^{4} \Delta_j^{dl}(x; \widehat{\lambda})$.

The difference-in-differences curve, $\Delta_4^{dl}(x)$, measures the DiD in yield curves of treated relative to control bonds over the treatment. $\Delta_4^{dl}(x)$ is a function of residual maturity x and, therefore, returns the treatment effect at specific maturities. Treatment-unrelated idiosyncratic yield-curve effects from pre- to post-treatment will be captured by the differences curve on the post-treatment indicator variable, $\Delta_3^{dl}(x)$. Hence, the specification is able to separate treatment-unrelated idiosyncratic yield-curve effects from actual treatment effects and allows for both heterogeneous idiosyncratic as well as heterogeneous treatment effects, simultaneously, in different parts of the yield curve.

To analyze its performance, we estimate Specification (6) with NLS. Lambda is estimated in-sample together with the other parameters using the start value $\lambda_{Seed} = 1$ (as above; true $\lambda = 0.7308$).[15] Standard errors are clustered at the bond level. For each regression, the estimation gives twelve coefficients and one estimate for lambda. Because it is difficult to grasp the economics from these coefficients, however, we use them to calculate the treatment effect at selected maturities. The corresponding standard errors are calculated with the delta method.[16]

Table 7 shows the results. At selected maturities of $x = 1, 2, 3, 5, 7, 10$, and 15 years, the table shows the true underlying treatment effects, which unconditionally are zero at all maturities. To the right, the table provides minimum and maximum of the estimated treat-

[15] Alternatively, we could use NLS to first estimate the yield curves separately for treated and control bonds pre- and post-treatment and, then, average across those within-group $\widehat{\lambda}$s, plug the average into (5) and use OLS to estimate Specification (4). For an application see Nyborg and Woschitz (2023).

[16] See, for example, Casella and Berger (2001). The Internet Appendix illustrates the procedure in detail.

ment effects across the two idiosyncratic yield-curve effects and the 1,000 families separately by m using the estimated difference-in-differences curve, $\Delta_4^{dl}(x)$, from Specification (6).

Insert Table 7 here.

In Panel A, when m equals to 0.25 years, Specification (6) estimates treatment effects across all maturities which are maximally ± 0.01 bps away from the true unconditional zero-effect across the 2,000 regressions. These are highly accurate estimates and a large improvement compared to the corresponding false treatment effects produced with Specification (1) that range from -11.59 to 12.01 bps (Figure 3). With m equals to 0.25 years, however, we only consider the case when the sample couplets are affected only by spurious correlation between residual maturity and the treatment indicator variable.

When spurious maturity-treatment correlation is compounded by systematic correlation, i.e. when m increases, the tails of the distributions exhibit relatively more measurement error at the short-end of the yield curve. The largest error is measured when m equals to 10 years at the one-year maturity. However, across the 2,000 regressions the difference ranges between -0.14 and $+0.08$ bps, which is still an improvement compared to the corresponding false treatment effects produced with Specification (1) which range from -24.06 to $+22.99$ bps (Figure 3). Moreover, Panel B shows that the increasing measurement error in m results from a lack of data. As a crude test, we repeat the analysis in Panel A but restrict the data to "good sample couplets," namely to those that contain at least one control and one treated bond in the [0,1]-year maturity bucket. Using those 5,468 "good sample couplets" (from the total of 8,000), the measurement error at the short-end disappears. The estimation error lies, again, maximally ± 0.01 bps away from the true unconditional zero-effect across all maturities and all m. This shows that the estimation error in Panel A results from a lack of data at the short-end in some samples.

In short, the fully flexible yield-curve DiD specification in (6) is able to control for relatively large treatment-unrelated, maturity-dependent idiosyncratic effects if there is sufficient data to estimate the yield curves. The specification eliminates false treatment effects independent of whether the samples are affected only by spurious correlation between residual maturity and the treatment or spurious combined with systematic correlation.

5. Mismeasured treatment effects

In this section, we model systematic actual treatment effects to examine whether Specification (1) identifies them. We focus exclusively on the case when only the yield curve of treated bonds, through the treatment effect, experiences a shift. Hence, there are no treatment-unrelated idiosyncratic effects.[17] The treatment either leads to a yield-curve twist or affects the treated bonds' yield curve only at the short-end. As a preview, the estimated treatment effect will be the average DiD effect conditional on the sample, which ignores maturity. Because the true underlying treatment effect, however, is maturity-dependent, this average will depend on where the treated bonds are located in the distribution of maturity and the true underlying treatment effects there.

Insert Figure 4 here.

To model the underlying term structures we continue to use Diebold and Li's (2006) yield-curve specification in (2). Figure 4 provides the underlying parameter values and shows the resulting yield curves graphically together with the numbers for yield levels and differences. The treatment yield-curve twist leads to a yield-curve increase for treated bonds of +6 bps at a maturity of one year and a reduction of −6 bps at ten years. The short-end treatment effect corresponds to a yield reduction for treated bonds of −6 bps at a maturity of one year and no effect (0 bps) at ten years relative to the pre-curve. In either case control bonds stay on the pre-treatment curve.

5.1 Main result: Mismeasured treatment effects

We use OLS to run Specification (1) with standard errors clustered at the bond level on this data. Figures 5a and 5b provide the results, respectively, for a treatment yield-curve twist and a treatment short-end effect. From left to right, the graphs show 1) the true underlying treatment effects as a function of maturity, 2) the distributions of estimated treatment effects when m is equal to 0.25 or 10 years for the treated bonds, and 3) the DiD effects against its t-statistics for these two modes.

[17]Logic: Control and treated bonds share the same term structure prior to treatment. Since there is no treatment-unrelated idiosyncratic effect, control bonds stay on the pre-treatment curve when the treatment takes place. Only the yield curve of treated bonds moves to a different location.

Insert Figure 5 here.

The classical DiD specification estimates an average treatment effect that (erroneously) ignores maturity. In case of a treatment yield-curve twist (Figure 5a) and if m equals to 0.25 years, based on t-statistics calculated with standard methods and conditional on the sample, across the 1,000 estimates one concludes in 121 cases that there is no treatment effect (at the 10%-significance level), and in 879 cases that the treatment effect is statistically significantly *negative* even if, in fact, the true underlying treatment effect at one year is a *positive* +6 bps. If the treated bonds are shifted toward the long-end, i.e. if m equals to 10 years, this becomes more extreme. In all the 1,000 cases the treatment effect is statistically significantly *negative* despite the *positive* true treatment effect of +6 bps at one year.

In case of a yield-curve treatment short-end effect (Figure 5b) and if m equals to 0.25 years, the average effect ranges from -2.94 to -0.48 bps. In terms of the coefficient's sign, researchers will draw the correct conclusion but the size of the effect is not identified. However, if the treated bonds are tilted sufficiently to the long-end, i.e. if m increases to 10 years, the estimated effects range from -0.72 to $+0.06$ bps. In 699 cases one concludes that there is no treatment effect because the DiD estimate is not statistically significant at the 10%-level if, in fact, the true treatment effect at one year is a negative -6 bps.

The analysis shows that the classical DiD specification may lead to incorrect conclusions. This is the case even in the absence of treatment-unrelated idiosyncratic yield-curve effects if the treatment effects themselves vary over maturity. The reason is that the classical DiD specification produces an average treatment effect that abstracts from maturity. Therefore, this average is tilted toward the effect at those maturities where the treated sample bonds are located. Much in the spirit of the discussion in Kahn and Whited (2018), the classical DiD specification is not able to elicit the true underlying treatment effect because the quantity it measures ignores maturity. The classical DiD specification is not designed to capture heterogeneous treatment effects in different parts of the term structure.

5.2 Naive specification adjustment does not work

How does the naive specification adjustment with the explicit term structure control handle the mismeasurement of treatment effects? The analysis shows that running Specification (4) produces the same mismeasured treatment effects as Specification (1). For illustration purposes, we compute the difference between the coefficients estimated with Specifications (1) and (4) for the two yield-curve treatment effects and the 4,000 sample couplets. Across the 8,000 differences in DiD coefficients between Specifications (1) and (4) even minimum and maximum amount to -0.0000 and 0.0001 bps, respectively.[18] Thus, the specifications produce the exact same mismeasured treatment effects.

As in Subsection 3.4, $\mathbf{B}'\mathbf{L}_{it}$ removes the average yield curve in the pooled sample of control, treated, pre- and post-event observations. Specifications (1) and (4) restrict the treatment effect to a parallel yield-curve level-shift from pre- to post-treatment. They are, therefore, misspecified if the true underlying treatment effect depends on maturity. Either of the specifications may lead to incorrect conclusions if the true treatment effect is maturity-dependent and, simultaneously, the samples are affected by spurious or spurious combined with systematic correlation between residual maturity and the treatment. This illustrates that these specifications are not suitable when there are heterogeneous treatment effects in different parts of the yield curve.

5.3 Fully flexible yield-curve DiD specification works well

In this subsection, we run Specification (6) on the data without treatment-unrelated idiosyncratic yield-curve effects but with true, maturity-dependent treatment effects. Basically, we repeat the analysis from Section 4 to see whether the fully flexible yield-curve DiD specification is able to identify the true underlying effects.

At selected maturities, Table 8 provides the true, maturity-dependent treatment effects. Either the data exhibits a yield-curve treatment twist or a yield-curve treatment short-end effect; there are no treatment-unrelated idiosyncratic effects. To the right, the table shows the minimum and maximum of the difference between the estimated and the true treatment effect across the two types of effects and the 1,000 families separately by m using the estimated difference-in-differences curve, $\Delta_4^{dl}(x)$, from Specification (6).

Insert Table 8 here.

The quantity that the difference-in-difference curve, $\Delta_4^{dl}(x)$, measures is a function of maturity. In Panel A, when m equals to 0.25 years and, therefore, the sample couplets are affected only by spurious correlation between residual maturity and the treatment,

[18]The differences in p-values range from -0.0049 to 0.0000, with (4) being slightly more conservative.

Specification (6) estimates precise treatment effects across all maturities and the 2,000 regressions that are maximally ±0.01 bps away from the true treatment effects. When spurious maturity-treatment correlation is compounded by systematic correlation, i.e. when m increases, we observe the same pattern as in Section 4. For example, when m equals to 10 years, the measurement error at the short-end increases from ±0.01 bps to between −0.21 and 0.18 bps. However, to focus on the "good sample couplets," those with at least one treated and one control bond in the [0,1]-year bucket (Panel B), makes the estimation error disappear and, again, the effects lie ±0.01 bps around the true treatment effects.

The fully flexible yield-curve DiD specification in (6) estimates the treatment effect as a function of maturity, which is meaningful (Kahn and Whited, 2018) since, in practice, the true treatment effects may depend on maturity. As shown, the specification is able to precisely measure even relatively small treatment effects independent of whether the samples are affected only by spurious correlation between residual maturity and the treatment or spurious combined with and systematic correlation.

6. Combine false and mismeasured treatment effects

In this section, we combine treatment-unrelated idiosyncratic with systematic treatment effects. We continue to employ Diebold and Li's (2006) yield-curve specification in (2) to build the underlying term structures. The yield curve parameter values and resulting yield levels and differences are provided in Table 9.

Insert Table 9 here.

By choosing values for the yield curve parameters, we generate an idiosyncratic short-end (long-end) effect, which reduces yields at a residual maturity of one year (fifteen years) by −50 bps and is close to zero at a maturity of fifteen years (one year). On top, we add a treatment yield-curve twist (short-end effect), which pushes yields of treated bonds up (down) by 6 bps at one year and pushes yields down by 6 bps (has a zero effect) at ten years relative to the idiosyncratic effects. Hence, we have a total of four combinations of idiosyncratic and treatment effects. The true underlying effects in this section are the same as the individual true effects from Sections 3 and 5 combined.

6.1 Main results

When idiosyncratic and actual treatment effects are present in the data simultaneously, it turns out that the estimated treatment effect for each sample couplet and combination of true effects equals the sum of the corresponding individual false and mismeasured treatment effects from Sections 3 and 5.[19] To illustrate this, we estimate the treatment effect with Specification (1) in the data with both effects present simultaneously and compare them to the sum of the individually estimated components, namely the false effect from Section 3 and the mismeasured effect from Section 5 using the same sample couplet and true underlying effects. We run a total of 16,000 different regressions, namely for the four combinations of idiosyncratic and treatment effects, the four m, and the 1,000 families. The difference between the coefficient estimated in the data with both effects present simultaneously and the sum of the individual coefficients estimated separately in the data when only one of the effects is present across the 16,000 regressions ranges from -0.003 to 0.003 bps and is, therefore, virtually zero.

Inherently, the treatment effect is assumed to be the same for all the treated bonds and the idiosyncratic effect the same for all the sample bonds. Specification (1), therefore, neither allows for heterogeneous idiosyncratic nor for heterogeneous systematic treatment effects along maturity. If these effects, however, are heterogeneous over maturity, Specification (1) produces a combination of false and mismeasured effects. The magnitude depends on the true underlying yield-curve effects and the spurious or spurious combined with systematic correlation between residual maturity and the treatment as implied by the relative maturity distributions in the treated and control bond samples.

6.2 Naive specification adjustment does not work

The naive specification adjustment with explicit term structure control in (4) does also not help to overcome the combined version of false and mismeasured treatment effects. As before, we compute the difference between the coefficients estimated with Specifications (1) and (4) for the four combinations of idiosyncratic and treatment effects and the 4,000 sample couplets. Across the 16,000 differences in DiD coefficients between Specifications (1) and

[19]This results from the simplicity of having just two time periods and no bond-individual noise (the bonds always lie on the term structure).

(4) even minimum and maximum amount to −0.0000 and 0.0000 bps, respectively.[20] This shows that the two specifications produce the exact same false and mismeasured treatment effects for the same reasons as explained in Subsections (3.4) and (5.2). Specifications (1) and (4) are misspecified if the true underlying yield-curve effects vary along maturity as these specifications restrict the true effects to parallel yield-curve level-shifts.

6.3 Fully flexible yield-curve DiD specification works well

In this subsection, we repeat the analysis from Section 4 and run Specification (6) on the data that exhibit both treatment-unrelated idiosyncratic yield-curve as well as yield-curve treatment effects while both type of effects vary along maturity.

At selected maturities, Table 10 provides true underyling treatment effects of the four combinations of idiosyncratic effects (short- or long-end) and treatment effects (twist or short-end). To the right, the table shows minimum and maximum of the difference between the estimated and the true treatment effects across the four effect combinations and the 1,000 families separately by m using $\Delta_4^{dl}(x)$ from Specification (6).

Insert Table 10 here.

The table shows that Specification (6) is able to separate even relatively small, maturity-dependent treatment effects from large, treatment-unrelated, and maturity-dependent idiosyncratic effects. In Panel A, when m equals 0.25 years and, therefore, the sample couplets are affected only by spurious maturity-treatment correlation, the estimated treatment effects across all maturities and the 4,000 regressions differ by maximally ±0.01 bps from the true treatment effects. When m increases, i.e. when spurious is compounded by systematic maturity-treatment correlation, the measurement error at the short-end increases to between −0.14 and 0.15 bps (when $m = 10$ at one-year maturity). As before, focusing on the "good sample couplets" (Panel B), again eliminates this estimation error.

In short, the fully flexible yield-curve DiD specification in (6) eliminates both false and mismeasured treatment effects independent of whether the samples are affected only by spurious maturity-treatment correlation or spurious combined with systematic correlation. Through the difference-in-differences curve, $\Delta_4^{dl}(x)$, the specification meaningfully estimates the treatment effect as a function of maturity (Kahn and Whited, 2018) and through the

[20] The differences in p-values range from −0.0056 to −0.0000, with (4) being slightly more conservative.

differences curve on the post-treatment indicator variable, $\Delta_3^{dl}(x)$, it accurately controls for treatment-unrelated idiosyncratic yield-curve effects. The specification separates small, maturity-dependent treatment effects from large, treatment-unrelated, maturity-dependent idiosyncratic effects and provides precisely estimated coefficients.

7. Semi-matching

Section 4 eliminates false and mismeasured treatment effects by combining the standard DiD method with flexible yield-curve modeling. Specification (6) accurately deals with one of the two components that create the problems, namely with heterogeneous effects in different parts of the term structure. This section shows that it is possible to eliminate false and mismeasured treatment effects by, instead, dealing with the other relevant component, namely unequal maturity distributions in the treatment sample groups.

As described earlier, a simple solution to the problem would be to perfectly match each treated with a control bond on maturity. However, since perfect matching on maturity is only rarely feasible in practice, researchers sometimes implement imperfect matching procedures (see, e.g., Ang, Bhansali, and Xing, 2010; Choi, Hoseinzade, Shin, and Tehranian, 2020). In this section we implement imperfect matching for a fixed-income pricing variable by combining a matching procedure with yield-curve modeling such as, for example, used by Lentner (2023).

In this section, we apply what we call "semi-matching," which works as follows. As perfect matching is not feasible, each treated bond is matched with a synthetic control bond whose yield is inferred from a contemporaneous yield curve of control bonds. The term "semi-matching" reflects that exact matching is not possible; only semi-matching on a synthetic control bond whose yield is estimated from the surrounding bonds via yield-curve modeling is feasible. We apply semi-matching as follows:

1. Separately for pre- and post-treatment periods, use Diebold and Li's (2006) specification in (2) to estimate the yield curve of control bonds;[21]
2. Apply semi-matching: Separately for each period, subtract the spot yield of the maturity-matched synthetic control bond from the yield of each treated sample bond;

[21] As before, we use NLS to estimate lambda in-sample using a start value $\lambda_{Seed} = 1$.

3. For each maturity-matched pair of treated and synthetic control bond, calculate the difference in the pair's yield-difference from pre- to post treatment, which represents the DiD in yields for each bond pair i, $yield_{it}^{DiD}$.

Semi-matching is illustrated in Figure 6 using a random sample couplet with m equal to 0.25 years. The figure plots the difference between the estimated Diebold-Li control-bond curves from pre- to post-treatment for an idiosyncratic yield-curve short-end (long-end) effect on the left (right). The (red) diamonds and (green) crosses show the difference of the treated bonds' yields from pre- to post-treatment in the cases, respectively, of a treatment yield-curve twist and a yield-curve treatment short-end effect.

Insert Figure 6 here.

In the following, we show in three steps that 1) semi-matching applied on a sample couplet eliminates false but not mismeasured treatment effects and may, thus, still lead to incorrect conclusions; 2) the analysis by maturity bucket overcomes false and reduces the likelihood of mismeasured treatment effects; 3) the analysis bond-by-bond resolves both false and mismeasured treatment effects. Finally, we show that estimating a yield curve through the individual bond-level DiDs after having applied semi-matching results in the difference-in-differences curve, $\Delta_4^{dl}(x)$, from Specification (6).

7.1 Mismeasurement

Semi-matching eliminates false treatment effects that derive from treatment-unrelated idiosyncratic yield-curve effects. As an illustration, we proceed as follows: First, we run a regression of the bond-level DiDs in yields on a constant C,

$$yield_{it}^{DiD} = \beta_{DiD} \times C + \varepsilon_{it}, \tag{9}$$

to estimate the average treatment effect, β_{DiD}, for each sample couplet. We run Specification (9) on the simulated data used in Section 6 with both true underlying idiosyncratic and treatment effects present simultaneously. This is a total of 16,000 different regressions (four combinations of idiosyncratic and treatment effects, four m, and 1,000 families). Importantly, idiosyncratic effects are present in that data.

Second, for each sample couplet, we compute the difference between the semi-matched

DiD coefficient from Specification (9), run on the data with idiosyncratic effects, and the DiD coefficient from Specification (1), but run on the data where only treatment (no idiosyncratic) effects are present (as in Section 5). Importantly, the latter coefficients are affected only by mismeasurement as idiosyncratic effects are absent from the data.

Across the 16,000 cases, namely the four combinations of idiosyncratic and treatment effects, the four m, and the 1,000 families, the differences in coefficients range from -0.003 to 0.004 bps.[22] This shows that semi-matching applied to an entire sample couplet produces the same mismeasurement as in Section 5 but, importantly, in the data with idiosyncratic effects. Hence, independent of spurious correlation between maturity and the treatment or of spurious combined with systematic correlation, semi-matching eliminates false treatment effects that stem from treatment-unrelated idiosyncratic yield-curve effects. However, mismeasured treatment effects may still lead to incorrect conclusions.

7.2 Semi-matching by maturity buckets

This subsection shows that semi-matching applied by maturity bucket eliminates false and reduces the likelihood of mismeasured treatment effects. As an illustration, we run Specification (9) on the data that exhibit both true idiosyncratic and treatment effects simultaneously, as in the previous subsection, but this time separately by maturity buckets as used in Subsection 3.3, namely $[0, 2]$, $(2, 5]$, $(5, 10]$, and $(10, 20]$ years.

Table 11 shows the results. Panels A and B cover the two treatment effects, twist and short-end effect, respectively. Separately by maturity bucket, each panel shows the number of estimated treatment effects, minimum and maximum number of involved bonds, the true treatment effects at maturity-range start and end, and minimum and maximum of the distributions of the estimated treatment effects for each idiosyncratic effect. Per panel and maturity bucket this is a total of 4,000 different regressions (four m and 1,000 families).

Insert Table 11 here.

Across panels, the number of estimated coefficients is below 4,000 for the $[0, 2]$- and the $(2, 5]$-year buckets as for some sample couplets no treated bonds are available in those maturity buckets. The remaining 3,614 and 3,998 coefficients for these two buckets are es-

[22]Notice that this extreme similarity is not generic. It is the result of the simplicity of the setting with only two time periods and no noise of yields around the term structures.

timated with maximally twenty-four treated bonds. Hence, the more granular the maturity buckets, the shorter its maturity ranges, and the fewer bonds per bucket.

Comparing true and estimated treatment effects shows that, for each maturity bucket, the estimated coefficients cover, with one exception, the range of true treatment effects.[23] Semi-matching produces an average treatment effect per maturity bucket, which misguides less often as, on the maturity-bucket level, more likely the sign of the coefficient is correct. However, it is still possible to draw incorrect conclusions regarding the sign of the coefficient. In our case, this is possible in maturity bucket $(2, 5]$ in Panel A and $(5, 10]$ in Panel B.

Overall, this illustrates the trade-off between the accuracy of measured treatment effects and the power of the test. With more maturity buckets of relatively shorter length, the precision of the estimated effects increases but is based on fewer bonds. Semi-matching applied on the maturity-bucket level eliminates false and reduces the likelihood of mismeasured treatment effects but the latter is not entirely ruled out.

7.3 Semi-matching bond-by-bond

The limiting case is to analyze the effects bond-by-bond as in the synthetic control literature (see, e.g., Abadie, 2021). Figure 6 shows the small treatment effects relative to the large idiosyncratic yield-curve effects and that both effects vary over maturity. Applying semi-matching via yield-curve modeling eliminates treatment-unrelated idiosyncratic yield-curve effects and is, as no averaging takes place, simultaneously immune against the mismeasurement of treatment effects. However, as discussed in detail in the synthetic control literature (see, e.g., Xu, 2017), estimating adequate standard errors is more laborious.[24]

One way to express the results of bond-by-bond semi-matching is to estimate a curve through the bond-level DiDs in yields. Separately for the four combinations of idiosyncratic and treatment effects and the same sample couplet as in Figure 6, we use NLS and $\lambda_{Seed} = 1$ to fit a Diebold-Li curve through the bond-level DiDs in yields in Figure 6. Exhibit 2 shows the true as well as the estimated treatment effects at selected maturities.

In short, Exhibit 2 shows the same treatment effects as the fully flexible yield-curve DiD specification in (6). Hence, estimating a curve through the bond-level DiDs in yields

[23] The one exception relates to the $(2, 5]$-year bucket in Panel B, where the maximum of -0.29 bps in case of an idiosyncratic short-end effect is outside the range of true effects of $[-2.97, -0.26]$ bps.

[24] Standard errors are typically derived from bootstrapping methods.

after having applied semi-matching results in the difference-in-differences curve, $\Delta_4^{dl}(x;\widehat{\lambda})$, as used in Section 4 and provided in (8).

Exhibit 2: Estimated treatment effects (in bps) with semi-matching

Residual maturity (in years)	True treatment effects		Idiosyncratic and treatment effects			
			short-end		long-end	
	twist	short-end	twist	short-end	twist	short-end
1	5.87	-6.23	5.88	-6.23	5.87	-6.23
2	3.75	-2.97	3.75	-2.97	3.74	-2.97
3	1.58	-1.39	1.58	-1.39	1.58	-1.39
5	-1.87	-0.26	-1.87	-0.26	-1.87	-0.26
7	-4.11	0.00	-4.11	0.00	-4.11	0.00
10	-6.09	0.08	-6.09	0.08	-6.09	0.08
15	-7.72	0.09	-7.71	0.09	-7.71	0.09

Overall, the advantage of the fully flexible yield-curve DiD specification is its simple and fast implementation. In our case, with zero-coupon yields it accurately estimates the DiD in yield curves with one single regression. Furthermore, it is possible to cluster standard errors at the bond level (Bertrand, Duflo, and Mullainathan, 2004). However, semi-matching is more broadly applicable if, besides maturity, other fixed-income characteristics such coupons, callability, ratings, etc. are of relevance. Contrarily, semi-matching is laborious, for example, when it comes to the estimation of standard errors (see, e.g., Xu, 2017).

8. Concluding remarks

Difference-in-differences (DiD) methodology is frequently used in finance to measure the causal impact of a treatment on yields or other pricing variables that exhibit time-varying term structures. Using simulations, this paper shows that the classical DiD specification with yield as dependent variable systematically produces false and mismeasured treatment effects, which holds even if the treatment is assigned randomly. To illustrate false and mismeasured treatment effects we simulate residual maturity of treated and control bond samples and model two types of yield-curve effects that both vary along maturity. The first type of effect is an idiosyncratic yield-curve effect that is not related to the treatment whatsoever and affects "treated" and control bonds irrespective of the treatment. The analysis shows that heterogeneous idiosyncratic effects over maturity lead to systematically false treatment effects even in the absence of true underlying treatment effects.

The second type of effect is a systematic actual yield-curve treatment effect that affects only the treated bonds. However, heterogeneous treatment effects across maturity may lead, even without true underlying idiosyncratic effects, to mismeasured treatment effects. Both false and mismeasured treatment effects can be economically large and can go in either direction. Based on standard methods, the likelihood of statistical significance increases with the coefficients' absolute size. As shown, neither explicit term structure control in the specification nor regressions separately by individual maturity buckets overcome the issues.

The limitation of these specifications is the inability to capture idiosyncratic and treatment effects if these effects vary over maturity. The magnitudes of false and mismeasured treatment effects depend on 1) heterogeneous idiosyncratic and treatment effects along maturity and 2) unequal residual maturity distributions in the treated and control bond samples. The root lies in the specifications' implicit restrictions of movements in the underlying yield curves to parallel level-shifts between the involved groups (controls, treated bonds, pre- and post-treatment). In fact, false and mismeasured treatment effects survive any DiD specification that implicitly assumes parallel yield-curve level-shifts if the true underlying effects are not limited to parallel level-shifts.

The paper provides new methodology to overcome both false as well as mismeasured treatment effects by combining DiD analysis with yield-curve modeling. First, the fully flexible yield-curve DiD specification takes the yield curve parameterization explicitly into the DiD estimator. The specification measures the DiD in yield curves between the involved groups and eliminates both false and mismeasured treatment effects. Second, semi-matching applies the first approach step-by-step to data.

False and mismeasured treatment effects are shown in the most trivial setting with bond yield on the left-hand side of the regression equation and bond-fixed effects on the right-hand side. Importantly, however, both false and mismeasured treatment effects survive (1) when the unit of analysis is an aggregation of the bond level, (2) with other dependent pricing variables (as long as they exhibit time-varying term structures), and (3) for other right-hand side controls of maturity. Overall, this shows that DiD analysis must be applied with great caution in fixed-income settings, especially with respect to residual maturity.

References

Abadie, Alberto, 2021, Using synthetic controls: Feasibility, data requirements, and methodological aspects, *Journal of Economic Literature* 59(2), 391-425.

Abadie, Alberto, Alexis Diamond, Jens Hainmueller, 2010, Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program, *Journal of the American Statistical Association* 105(490), 493-505.

Abadie, Alberto, Javier Gardeazabal, 2003, The economic costs of conflict: A case study of the basque country, *American Economic Review* 93(1), 113-132.

Adelino, Manuel, Miguel A. Ferreira, 2016, Bank ratings and lending supply: Evidence from sovereign downgrades, *Review of Financial Studies* 29(7), 1709-1746.

Ang, Andrew, Vineer Bhansali, Yuhang Xing, 2010, Taxes on tax-exempt bonds, *Journal of Finance* 65(2), 565-601.

Athey, Susan, Guido W. Imbens, 2022, Design-based analysis in difference-in-differences settings with staggered adoption, *Journal of Econometrics* 226(1), 62-79.

Ayotte, Kenneth, Stav Gaon, 2011, Asset-Backed Securities: Costs and Benefits of "Bankruptcy Remoteness," *Review of Financial Studies* 24(4), 1299-1335.

Bai, Jushan, 2009, Panel data models with interactive fixed effects, *Econometrica* 77(4), 1229-1279.

Baker, Andrew C., David F. Larcker, Charles C. Y. Wang, 2022, How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics* 144(2), 370-395.

Bao, Jack, Maureen O'Hara, Xing (Alex) Zhoua, 2018, The Volcker Rule and corporate bond market making in times of stress, *Journal of Financial Economics* 130(1), 95-113.

Bao, Jack, Kewei Hou, 2017, De facto seniority, credit risk, and corporate bond prices, *Review of Financial Studies* 30(11), 4038-4080.

Benetton, Matteo, Davide Fantino, 2021, Targeted monetary policy and bank lending behavior, *Journal of Financial Economics* 142(1), 404-429.

Bertrand, Marianne, Esther Duflo, Sendhil Mullainathan, 2004, How much should we trust differences-in-differences estimates?, *Quarterly Journal of Economics* 119(1), 249-275.

Bitler, Marianne P., Jonah B. Gelbach, Hilary W. Hoynes, 2006, What mean impacts miss: Distributional effects of welfare reform experiments, *American Economic Review* 96(4), 988-1012.

Callaway, Brantly, 2022, Difference-in-differences for policy evaluation, in Klaus F. Zimmermann (ed.), *Handbook of Labor, Human Resources and Population Economics*, 1-61, Springer, Cham.

Callaway, Brantly, Tong Li, Tatsushi Oka, 2018, Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods, *Journal of Econometrics* 206(2), 395-413.

Callaway, Brantly, Tong Li, 2019, Quantile treatment effects in difference in differences models with panel data, *Quantitative Economics* 10, 1579-1618.

Callaway, Brantly, Tong Li, 2023, Policy evaluation during a pandemic, *Journal of Econometrics* 236(1), 105454.

Callaway, Brantly, Pedro H.C. Sant'Anna, 2021, Difference-in-differences with multiple time periods, *Journal of Econometrics* 225(2), 200-230.

Casella, George, Roger L. Berger, 2001, *Statistical inference*, Brooks/Cole Cengage Learning, Belmont.

Chava, Sudheer, Dmitry Livdan, Amiyatosh Purnanandam, 2009, Do shareholder rights affect the cost of bank loans?, *Review of Financial Studies* 22(8), 2973-3004.

Choi, Jaewon, Saeid Hoseinzade, Sean Seunghun Shin, Hassan Tehranian, 2020, Corporate bond mutual funds and asset fire sales, *Journal of Financial Economics* 138(2), 432-457.

Cornaggia, Jess, Kimberly J. Cornaggia, Ryan D. Israelsen, 2018, Credit ratings and the cost of municipal financing, *Review of Financial Studies* 31(6), 2038-2079.

Dannhauser, Caitlin D., 2017, The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs), *Journal of Financial Economics* 125(3), 537-560.

de Chaisemartin, Clément, Xavier D'Haultfœuille, 2020, Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects, *American Economic Review* 110(9), 2964-2996.

Diebold, Francis X., Canlin Li, 2006, Forecasting the term structure of government bond yields, *Journal of Econometrics* 130(2), 337-364.

Ding, Yi, Wei Xiong, Jinfan Zhang, 2022, Issuance overpricing of China's corporate debt securities, *Journal of Financial Economics* 144(1), 328-346.

Ferreira, Miguel A., Pedro Matos, 2012, Universal banks and corporate control: Evidence from the global syndicated loan market, *Review of Financial Studies* 25(9), 2703-2744.

Foley-Fisher, Nathan, Rodney Ramcharan, Yu Edison, 2016, The impact of unconventional monetary policy on firm financing constraints: Evidence from the maturity extension program, *Journal of Financial Economics* 122(2), 409-429.

Francis, Bill, Iftekhar Hasan, Michael Koetter, Qiang Wu, 2012, Corporate boards and bank loan contracting, *Journal of Financial Research* 35(4), 521-552.

Gao, Pengjie, Chang Lee, Dermot Murphy, 2020, Financing dies in darkness? The impact of newspaper closures on public finance, *Journal of Financial Economics* 135(2), 445-467.

Gobillon, Laurent, Thierry Magnac, 2016, Regional policy evaluation: Interactive fixed effects and synthetic controls, *Review of Economics and Statistics* 98(3), 535-551.

Goodman-Bacon, Andrew, 2021, Difference-in-differences with variation in treatment timing, *Journal of Econometrics* 225(2), 254-277.

Hasan, Iftekhar, Chun Keung (Stan) Hoi, Qiang Wu, Hao Zhang, 2014, Beauty is in the eye of the beholder: The effect of corporate tax avoidance on the cost of bank loans, *Journal of Financial Economics* 113(1), 109-130.

Heckman, James J., Jeffrey Smith, Nancy Clements, 1997, Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts, *Review of Economic Studies* 64(4), 487-535.

Imbens, Guido W., Jeffrey M. Wooldridge, 2009, Recent developments in the econometrics of program evaluation, *Journal of Economic Literature* 47(1), 5-86.

John, Kose, Anthony W. Lynch, Manju Puri, 2003, Credit ratings, collateral, and loan characteristics: Implications for yield, *Journal of Business* 76(3), 371-409.

Kahn, R., Toni M. Whited, 2018, Identification is not causality, and vice versa, *Review of Corporate Finance Studies* 7(1), 1-21.

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

Liu, Yan, Jing Cynthia Wu, 2021, Reconstructing the yield curve, *Journal of Financial Economics* 142(3), 1395-1425.

Nelson, Charles R., Andrew F. Siegel, 1987, Parsimonious modeling of yield curve, *Journal of Business* 60(4), 473-489.

Nyborg, Kjell G., Jiri Woschitz, 2023, The price of money: The reserves convertibility premium over the term structure, Working paper.

Painter, Marcus, 2020, An inconvenient cost: The effects of climate change on municipal bonds, *Journal of Financial Economics* 135(2), 468-482.

Qiu, Jiaping, Fan Yu, 2009, The market for corporate control and the cost of debt, *Journal of Financial Economics* 93(3), 505-524.

Rodano, Giacomo, Nicolas Serrano-Velarde, Emanuele Tarantino, 2016, Bankruptcy law and bank financing, *Journal of Financial Economics* 120(2), 363-382.

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, John Poe (2023), What's trending in difference-in-differences? A synthesis of the recent econometrics literature, *Journal of Econometrics* 235(2), 2218-2244.

Sun, Liyang, Sarah Abraham, 2021, Estimating dynamic treatment effects in event studies with heterogeneous treatment effects, *Journal of Econometrics* 225(2), 175-199.

Svensson, Lars E.O., 1994, Estimating and interpreting forward interest rates: Sweden 1992–1994, Working Paper No. WP/94/I, 14, International Monetary Fund. Washington. D.C.

Todorov, Karamfil, 2020, Quantify the quantitative easing: Impact on bonds and corporate debt issuance, *Journal of Financial Economics* 135(2), 340-358.

Titman, Sheridan, Sergey Tsyplakov, 2010, Originator performance, CMBS structures, and the risk of commercial mortgages, *Review of Financial Studies* 23(9), 3558-3594.

Xu, Yiqing, 2017, Generalized synthetic control method: Causal inference with interactive fixed effects models, *Political Analysis* 25(1), 57-76.

Table 1: Recent top finance publications potentially affected.

This tables shows a collection of recent top finance publications potentially affected by the issues discussed in the present paper. The list was created by a manual search of The Journal of Finance (JF), The Journal of Financial Economics (JFE), and The Review of Financial Studies (RFS) over the period July 1 to September 10, 2021, using relevant combinations of key words. "Treas." is short for Treasury and "mat." is short maturity.

Authors	**Publ. year**	**Jour-nal**	**Analysis level**	**Dependent variable**	**Independent variable(s) to capture maturity**[(1)]
Chava, Livdan, Purnanandam	2009	RFS	Loan, or firm	Log changes in loan spread over LIBOR	–
Qiu, Yu	2009	JFE	Firm	Credit spread over mat.-matched Treas.	Duration and convexity[(2)]
Titman, Tsyplakov	2010	RFS	Mortgage	Credit spread over mat.-matched Treas.	Mortgage resid. time-to-mat.
Ayotte, Gaon	2011	RFS	ABS issuance	ABS spread over mat.-matched swap rates	Average life and its quadratic term
Hasan, Hoi, Wu, Zhang	2014	JFE	Loan facility	Log loan spread over LIBOR	Log resid. time-to-mat.
Rodano, Serrano-Velarde, Tarantino	2016	JFE	Bank-firm	Loan interest rate	Bank-firm fixed effects, loan mat.[(3)]
Adelino, Ferreira	2016	RFS	Loan facility	Loan spread over LIBOR	–
Cornaggia, Cornaggia, Israelsen	2018	RFS	Bond	Yield, credit spread over dur.-matched Treas.[(4)]	Duration
Dannhauser	2017	JFE	Bond	Yield spread over mat.-matched swap rates[(5)]	Bond fixed effects
Bao, O'Hara, Zhou	2018	JFE	Bond	Yield spread over mat.-matched Treas.	Log resid. time-to-mat.
Todorov	2020	JFE	Bond	Yield	Bond fixed effects
Gao, Lee, Murphy	2020	JFE	Bond	Yield spread over coupon-equiv. Treas. yield[(6)]	Resid. time-to-mat. and its inverse
Painter	2020	JFE	Bond	Annualized issuance cost[(7)], yield	Log resid. time-to-mat.[(8)]
Benetton, Fantino	2021	JFE	Bank-firm	Loan rate	Bank-firm fixed effects
Ding, Xiong, and Zhang	2021	JFE	Issuance	Issuance spread over Chinese Treas.	Resid. time-to-mat.

(1) Dashes mean that there are no independent variables to capture maturity

(2) Bond-level characteristics are converted into firm-level measures (value-weighted). Authors also run bond level regressions: Independent variables are not aggregated

(3) Loan maturity is measured with indicator variables for < 1 year, 1-5 years, and > 5 years

(4) Either using each bond's time-to-maturity or with the callable bonds' call dates in lieu of their maturity dates

(5) Monthly volume-weighted yield of a bond over the maturity-matched swap rate

(6) See Footnote 7 in Gao, Lee, and Murphy (2020) for details

(7) For details see Painter (2020), page 470

(8) Sample split at maturity of 25 years throughout the paper

Table 2: Time-variation in the term spread in practice.

This table shows the distribution of changes in the term spread for a selected group of countries over the period from January 3, 2000 to December 14, 2022. The term spread is measured in basis points (bps) and calculated as ten-year minus three-month zero-coupon spot yield. Panel A shows daily changes in the term spread (using end-of-day pricing data) and Panel B monthly changes (using end-of-month data). Data source: Bloomberg.

Country	Mean	SD	Min	P1	P5	Med	P95	P99	Max	N
Panel A: Distribution of daily changes in ten-year minus three-month term spread (in bps)										
Japan	0	3	-30	-9	-4	0	4	10	30	5,985
Germany	0	5	-37	-13	-7	0	7	14	39	5,987
United States	0	6	-61	-16	-9	0	10	18	63	5,983
France	0	5	-78	-12	-7	0	7	13	86	5,217
United Kingdom	0	6	-84	-16	-8	0	9	17	81	5,985
Spain	0	7	-121	-17	-9	0	9	19	63	5,959
Ireland	0	7	-117	-17	-8	0	8	19	76	5,982
Netherlands	0	5	-100	-12	-7	0	7	13	99	5,831
Italy	0	9	-178	-24	-11	0	12	24	110	5,987
Portugal	0	13	-180	-37	-11	0	12	37	142	5,985
China	0	12	-210	-34	-15	0	15	33	190	4,766
Greece	0	90	-2,012	-126	-27	0	22	121	1,822	5,976
Panel B: Distribution of monthly changes in ten-year minus three-month term spread (in bps)										
Japan	0	10	-31	-22	-14	-1	18	28	49	219
United States	0	30	-103	-58	-45	-1	61	83	93	219
France	-1	24	-97	-61	-35	-3	38	109	118	193
Germany	-1	23	-92	-51	-35	-3	35	64	149	219
Netherlands	-1	25	-109	-59	-36	-3	36	53	151	207
United Kingdom	0	27	-143	-58	-38	-2	43	79	135	219
Spain	1	31	-143	-87	-36	-2	47	127	137	216
China	0	34	-116	-102	-54	0	46	116	166	179
Ireland	0	34	-115	-82	-52	-2	48	112	208	219
Italy	0	32	-148	-73	-44	-3	38	116	219	219
Portugal	1	49	-145	-121	-61	-4	53	193	285	219
Greece	0	283	-1,810	-1,060	-219	0	221	876	1,745	219

Table 3: The maturity structure of outstanding debt in practice.

This table provides the number of outstanding securities as well as outstanding debt by maturity buckets for the same selection of countries as in Table 2 at the beginning of 2023 (Panel A) and at the beginning of 2011 (Panel B). For each country, outstanding debt by maturity bucket is provided as percentage of the total outstanding debt by that country. Data source: Bloomberg.

Country	# of sec.	[0-2]	(2-5]	(5-10]	(10-15]	(15-20]	(20-30]	>30y
Panel A: At the beginning of 2023								
Netherlands	31	23	21	26	9	11	9	1
Portugal	32	17	28	35	14	2	5	0
Ireland	59	10	20	38	11	3	13	4
Spain	82	23	25	30	8	5	7	2
Greece	82	21	18	27	18	6	9	1
Germany	84	30	24	26	6	4	10	0
France	97	21	25	31	6	7	7	4
United Kingdom	122	14	17	19	10	10	17	13
Italy	205	28	25	25	9	5	7	1
United States	444	42	24	17	0	6	10	0
China	493	33	29	23	1	2	9	4
Japan	559	31	20	21	8	8	9	2
Panel B: At the beginning of 2011								
Netherlands	41	35	25	22	6	4	4	3
Portugal	45	26	27	30	11	0	5	0
Ireland	16	12	20	59	9	0	0	0
Spain	63	33	24	23	7	4	8	2
Greece	105	22	28	28	10	5	7	1
Germany	274	33	26	26	3	4	8	1
France	92	30	24	25	8	4	6	3
United Kingdom	100	14	19	21	9	8	16	13
Italy	178	30	23	22	9	5	9	1
United States	305	41	26	23	3	3	5	0
China	285	32	24	21	12	5	5	1
Japan	466	36	25	22	6	7	3	0

Table 4: Overview on simulation of residual maturity.

This table provides an overview of the simulated residual maturities. Family f comprises five simulated residual maturity samples, namely one for control bonds with $m = 0.25$ and four for treated bonds with $m = 0.25, 1, 3, 10$. m is the mode of the triangular pdf from which residual maturity is drawn. Each sample is comprised of fifty bonds. For each family f, we build four sample couplets by pairing each sample of treated bonds ($m = 0.25, 1, 3, 10$) with the sample of control bonds ($m = 0.25$). Thus, each sample couplet contains fifty control and fifty treated bonds. In total, we draw 1,000 families. Panel A shows the distributions of average-maturity across the 1,000 families separately for each treatment group and mode. Panel B provides the distributions of average-maturity ratios across the families by mode m. For each sample couplet, the ratio is calculated as average residual maturity of the fifty treated bonds divided by average residual maturity of the fifty control bonds. Panel C shows the distributions of the correlation between residual maturity and the treatment indicator variable by mode m.

Panel A: Average-maturity across families of samples by treatment group and mode

Group	m	Population mean	Sample distributions: No. of families	Mean	SD	Med	Min	Max
Control	0.25	6.75	1,000	6.764	0.668	6.765	4.327	8.807
Treated	0.25	6.75	1,000	6.734	0.674	6.708	4.322	9.168
	1	7.00	1,000	7.034	0.611	7.029	5.361	9.008
	3	7.67	1,000	7.648	0.604	7.631	5.819	9.558
	10	10.00	1,000	9.979	0.587	10.015	8.043	11.981

Panel B: Average-maturity ratios across families of sample couplets by mode

m treated bonds	Population mean	Sample distributions: No. of families	Mean	SD	Med	Min	Max
0.25	1.00	1,000	1.005	0.138	0.994	0.592	1.524
1	1.04	1,000	1.050	0.141	1.039	0.710	1.598
3	1.14	1,000	1.142	0.146	1.135	0.691	1.805
10	1.48	1,000	1.490	0.177	1.481	1.095	2.231

Panel C: Correlation between residual maturity and treatment indicator variable

m treated bonds	Sample distributions: No. of families	Mean	SD	Med	Min	Max
0.25	1,000	-0.004	0.097	-0.004	-0.349	0.318
1	1,000	0.029	0.098	0.028	-0.268	0.322
3	1,000	0.097	0.098	0.100	-0.301	0.371
10	1,000	0.344	0.091	0.349	0.077	0.627

Table 5: False treatment effects.

This table shows the estimation of false treatment effects using the simulated data with heterogeneous idiosyncratic yield-curve effects along maturity but without true treatment effects. The specification is $yield_{it} = \alpha_i + \delta_t + \beta_{DiD}\ \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \varepsilon_{it}$, 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 results are presented as follows: For each mode $m \in \{0.25, 1, 3, 10\}$ of the treated bonds the 1,000 families of sample couplets are ordered according to the average-maturity ratio (treated divided by control bonds). For each m, we select the sample couplets with order index as indicated in the table. Panel A provides the corresponding average-maturity ratio and Panel B the corresponding correlation between residual maturity and the treatment indicator variable. For the corresponding sample couplet in Panels A and B, Panel C (D) shows the estimated DiD effects and, underneath in parentheses, p-values based on standard errors clustered at the bond level for an idiosyncratic yield-curve short-end (long-end) effect. a, b, and c denote significance (two-sided) at the levels of 1%, 5%, and 10%, respectively. Coefficients statistically significant at the 10%-level or stronger are marked in bold.

	Order index of families of sample couplets								
m	1	10	50	250	501	751	951	991	1000
Panel A: Ratio of average residual time-to-maturity of treated over control bonds									
0.25	0.592	0.742	0.799	0.904	0.994	1.086	1.259	1.379	1.524
1	0.710	0.766	0.838	0.948	1.039	1.142	1.304	1.419	1.598
3	0.691	0.838	0.913	1.044	1.135	1.228	1.400	1.527	1.805
10	1.095	1.144	1.232	1.362	1.482	1.594	1.790	2.027	2.231
Panel B: Correlation between residual maturity and treatment indicator variable									
0.25	-0.349	-0.220	-0.162	-0.078	-0.004	0.064	0.157	0.225	0.318
1	-0.268	-0.213	-0.140	-0.039	0.029	0.093	0.196	0.234	0.322
3	-0.301	-0.145	-0.068	0.036	0.094	0.162	0.239	0.308	0.371
10	0.077	0.139	0.176	0.266	0.353	0.413	0.494	0.565	0.567
Panel C: Idiosyncratic yield-curve short-end effect									
0.25	**-11.59**^{a}	**-7.22**^{b}	-3.29	-3.89	0.89	-0.58	4.16	**9.32**^{a}	**10.79**^{a}
	(0.00)	(0.04)	(0.35)	(0.28)	(0.81)	(0.86)	(0.26)	(0.01)	(0.00)
1	**-8.67**^{a}	**-7.79**^{b}	-4.82	0.54	1.32	2.19	**6.01**^{c}	**7.84**^{b}	**11.32**^{a}
	(0.01)	(0.02)	(0.14)	(0.89)	(0.70)	(0.55)	(0.09)	(0.03)	(0.00)
3	**-8.73**^{a}	-2.01	-0.82	2.30	4.24	**5.88**^{c}	**8.63**^{b}	**11.35**^{a}	**14.24**^{a}
	(0.01)	(0.55)	(0.82)	(0.46)	(0.21)	(0.09)	(0.01)	(0.00)	(0.00)
10	**6.30**^{c}	4.26	**8.61**^{b}	**9.88**^{a}	**10.80**^{a}	**14.14**^{a}	**15.34**^{a}	**19.47**^{a}	**22.99**^{a}
	(0.06)	(0.14)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Panel D: Idiosyncratic yield-curve long-end effect									
0.25	**12.01**^{a}	**7.67**^{b}	3.86	4.17	-1.06	0.25	-4.35	**-9.94**^{a}	**-11.34**^{a}
	(0.00)	(0.04)	(0.30)	(0.27)	(0.78)	(0.94)	(0.26)	(0.01)	(0.00)
1	**9.31**^{a}	**8.31**^{b}	5.32	-0.21	-1.51	-2.08	**-6.45**^{c}	**-8.28**^{b}	**-11.84**^{a}
	(0.01)	(0.02)	(0.13)	(0.96)	(0.68)	(0.60)	(0.08)	(0.03)	(0.00)
3	**9.72**^{a}	2.59	1.16	-2.15	-4.40	-6.12	**-8.82**^{b}	**-11.50**^{a}	**-14.95**^{a}
	(0.00)	(0.47)	(0.77)	(0.52)	(0.22)	(0.10)	(0.02)	(0.00)	(0.00)
10	**-6.56**^{c}	-4.54	**-8.60**^{b}	**-10.53**^{a}	**-11.79**^{a}	**-15.32**^{a}	**-16.51**^{a}	**-20.63**^{a}	**-24.06**^{a}
	(0.07)	(0.14)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 6: 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 same classical DiD specification as in Table 5 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 $t < -t_{cv}$, $-t_{cv} \leq t \leq t_{cv}$, 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 meas. effects, N	Mean (in bps)	Med (in bps)	$t < -t_{cv}$ N	$t < -t_{cv}$ Min (in bps)	$t < -t_{cv}$ Max (in bps)	$-t_{cv} \leq t \leq t_{cv}$ N	$-t_{cv} \leq t \leq t_{cv}$ Min (in bps)	$-t_{cv} \leq t \leq t_{cv}$ Max (in bps)	$t_{cv} < t$ N	$t_{cv} < t$ Min (in bps)	$t_{cv} < t$ 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

Table 7: 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 Table 5, $\mathbf{L}_{it}$ 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 $\beta_{k,j}$, $k = 0, \ldots, 2$. 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 yield curve, $\Delta_4^{dl}(x)$, 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 1,000 families separately by m using the DiD yield curve, $\Delta_4^{dl}(x)$. 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
No. 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
No. of sample couplets				1,954		1,818		1,210		486	

* At least one treated and one control bond in one-year maturity bucket.

Table 8: The fully flexible yield-curve DiD specification to eliminate mismeasured treatment effects.

This table shows treatment effects estimated with the fully flexible yield-curve DiD specification (specification and notation are as in Table 7). There are no treatment-unrelated idiosyncratic yield-curve effects in the data but the treatment effect varies along 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 underlying treatment effects across the two types of yield-curve treatment effects and the 1,000 families separately by m using the DiD yield curve, $\Delta_4^{dl}(x)$. Standard errors are clustered at the bond level and calculated using the delta method.

Residual maturity (in years)	Idiosyn-cratic 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
Panel A: All sample couplets											
1	0	5.87	-6.23	-0.01	0.01	-0.01	0.02	-0.05	0.03	-0.21	0.18
2	0	3.75	-2.97	-0.00	0.00	-0.00	0.00	-0.01	0.01	-0.08	0.06
3	0	1.58	-1.39	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.03	0.02
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
No. of sample couplets				2,000		2,000		2,000		2,000	
*Panel B: Good sample couplets**											
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
No. of sample couplets				1,954		1,818		1,210		486	

* At least one treated and one control bond in one-year maturity bucket.

Table 9: Modeling idiosyncratic term-structure effects combined with treatment effects.

To model the term structure we employ Diebold and Li (2006)'s 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.

Panel A: Idiosyncratic short-end effect, treatment yield-curve twist

i) Parameter values	γ_0	γ_1	γ_2	λ
Pre-curve	4.000	-2.000	0.000	0.7308
Post-curve controls	4.140	-2.650	-0.800	0.7308
Post-curve treated	4.030	-2.470	-0.620	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 controls	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	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: Idiosyncratic long-end effect, treatment yield-curve twist

i) Parameter values	γ_0	γ_1	γ_2	λ
Pre-curve	4.000	-2.000	0.000	0.7308
Post-curve controls	3.350	-1.350	1.000	0.7308
Post-curve treated	3.240	-1.170	1.180	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 controls	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	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: Idiosyncratic short-end effect, treatment 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 controls	4.140	-2.650	-0.800	0.7308
Post-curve treated	4.141	-2.781	-0.670	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 controls	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	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: Idiosyncratic long-end effect, treatment 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 controls	3.350	-1.350	1.000	0.7308
Post-curve treated	3.351	-1.481	1.130	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 controls	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	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

Table 10: The fully flexible yield-curve DiD specification to eliminate both false and mismeasured treatment effects. This table shows treatment effects estimated with the fully flexible yield-curve DiD specification (specification and notation are as in Table 7). There are both treatment-unrelated idiosyncratic yield-curve as well as yield-curve treatment effects in the data and both vary along 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 underlying treatment effects across the four combinations of idiosyncratic effects (at short- or long-end) and treatment effects (twist or short-end) and the 1,000 families separately by m using the DiD yield curve, $\Delta_4^{dl}(x)$. 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
Panel A: All sample couplets												
1	-50.35	3.91	5.87	-6.23	-0.01	0.01	-0.01	0.01	-0.02	0.02	-0.14	0.15
2	-43.65	-1.47	3.75	-2.97	-0.00	0.00	-0.01	0.01	-0.01	0.01	-0.06	0.06
3	-35.82	-9.31	1.58	-1.39	-0.01	0.00	-0.01	0.00	-0.01	0.01	-0.03	0.02
5	-22.58	-23.60	-1.87	-0.26	-0.00	0.00	-0.01	0.00	-0.00	0.00	-0.01	0.01
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	-0.01	0.00	-0.01	0.01	-0.01	0.01	-0.00	0.01
No. of sample couplets					4,000		4,000		4,000		4,000	
*Panel B: Good sample couplets**												
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
No. of sample couplets					3,908		3,636		2,420		972	

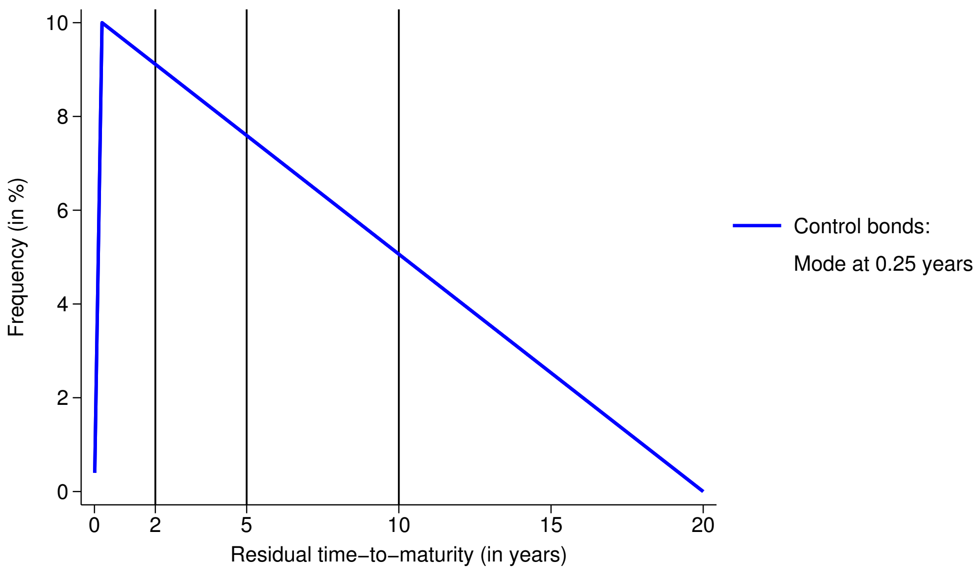
* At least one treated and one control bond in one-year maturity bucket.

Table 11: Semi-matching separately for individual maturity buckets.

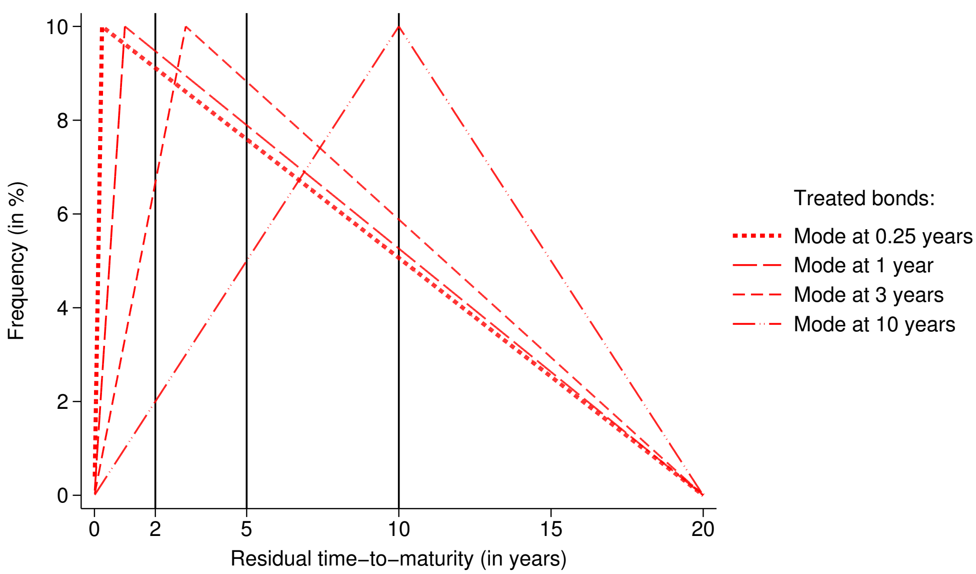
This table shows the results from applying semi-matching separately by maturity buckets using the data that exhibit both idiosyncratic and treatment effects simultaneously (as in Section 6), NLS with $\lambda_{Seed} = 1$ to estimate the Diebold-Li yield curve given in (2), and the specification $yield_{it}^{DiD} = \beta_{DiD} \times C + \varepsilon_{it}$ with $yield_{it}^{DiD}$ the bond-level DiDs in yields, C a constant, and β_{DiD} the treatment effect. Panel A (B) covers the case of a yield-curve treatment twist (short-end effect). Separately for each maturity bucket and idiosyncratic yield-curve short- and long-end effects, each panel shows the number of estimated treatment effects, minimum and maximum number of bonds involved in the estimations, the true treatment effects at maturity-range start and end, and minimum and maximum of the estimated treatment effects across the four $m \in \{0.25, 1, 3, 10\}$ and the 1,000 families of sample couplets (which is a total of 4,000 regressions per maturity bucket).

Maturity	Number of			True effects*		Distributions of $\widehat{\beta}_{DiD}$ (in bps)			
range	estim.	treat. bonds		(in bps)		GE: Short-end		GE: Long-end	
(in years)	coeff.	Min	Max	start	end	Min	Max	Min	Max
Panel A: Treatment yield-curve twist									
$[0, 2]$	3,614	1	20	7.00	3.75	3.75	7.00	3.75	7.00
$(2, 5]$	3,998	1	24	3.75	-1.87	-1.74	3.34	-1.73	3.34
$(5, 10]$	4,000	6	31	-1.87	-6.09	-5.30	-3.14	-5.30	-3.13
$(10, 20]$	4,000	4	36	-6.09	-8.54	-7.98	-6.61	-7.98	-6.61
Panel B: Treatment yield-curve short-end effect									
$[0, 2]$	3,614	1	20	-13.00	-2.97	-12.40	-2.97	-12.40	-2.97
$(2, 5]$	3,998	1	24	-2.97	-0.26	-2.60	-0.29	-2.60	-0.28
$(5, 10]$	4,000	6	31	-0.26	0.08	-0.10	0.05	-0.10	0.05
$(10, 20]$	4,000	4	36	0.08	0.09	0.08	0.09	0.08	0.09

* True treatment effects are given for start and end of maturity range in first column.

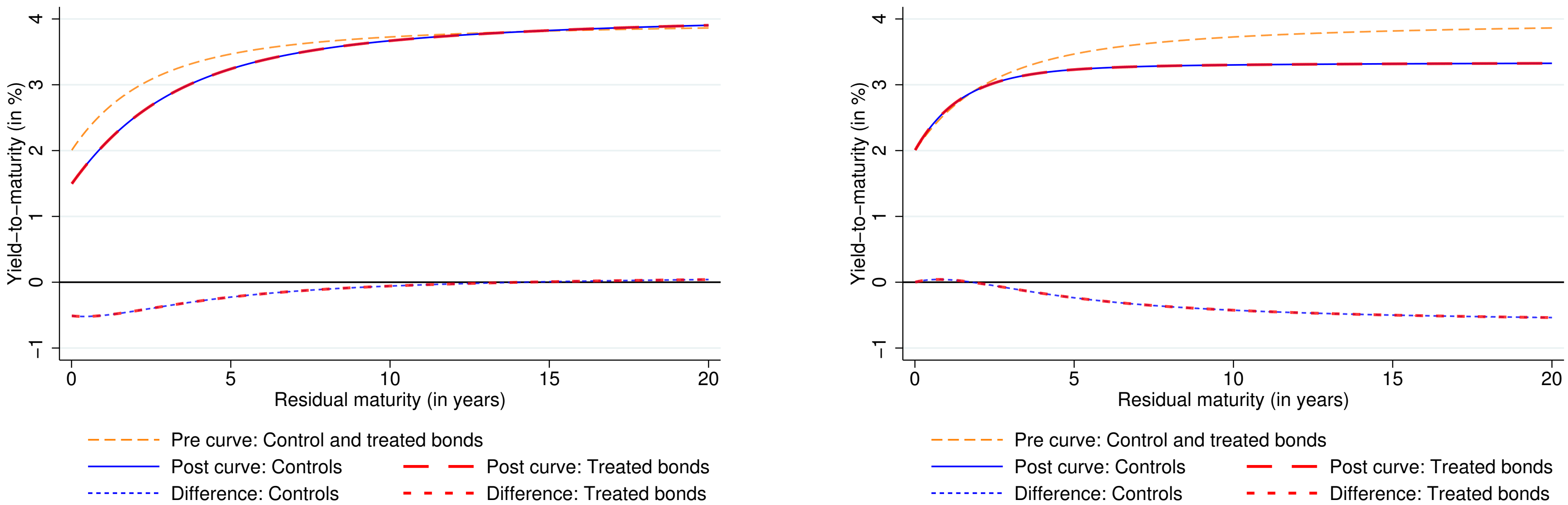


a) Control bonds



b) Treated bonds

Figure 1: Triangular probability density functions (pdfs) with different modes m. This figure shows the triangular pdfs 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).



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

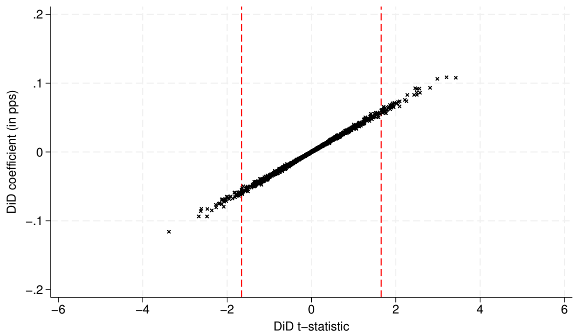
b) Idiosyncratic yield-curve long-end effect:

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

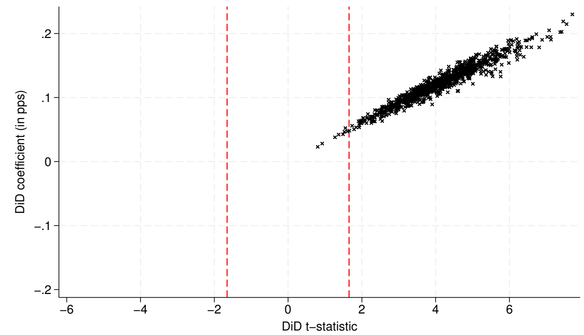
ii) Yields (in %) and differences (in pps) at selected maturities:

	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

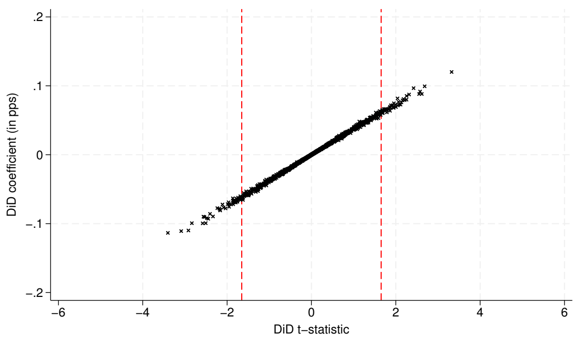
Figure 2: Modeling idiosyncratic effects in the term structure of interest rates.
To model the term structure we employ Diebold and Li (2006)'s yield curve specification. The mini table underneath each plot shows the parameter values to create the true underlying term structures as well as the resulting yield levels and yield differences at selected maturities. Figures 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.



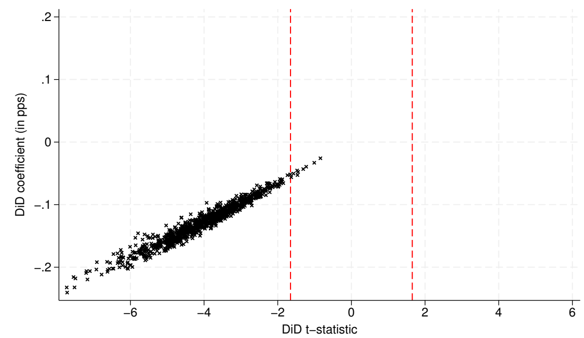
a) Idiosyncratic short-end effect: $m = 0.25$



b) Idiosyncratic short-end effect: $m = 10$



c) Idiosyncratic long-end effect: $m = 0.25$

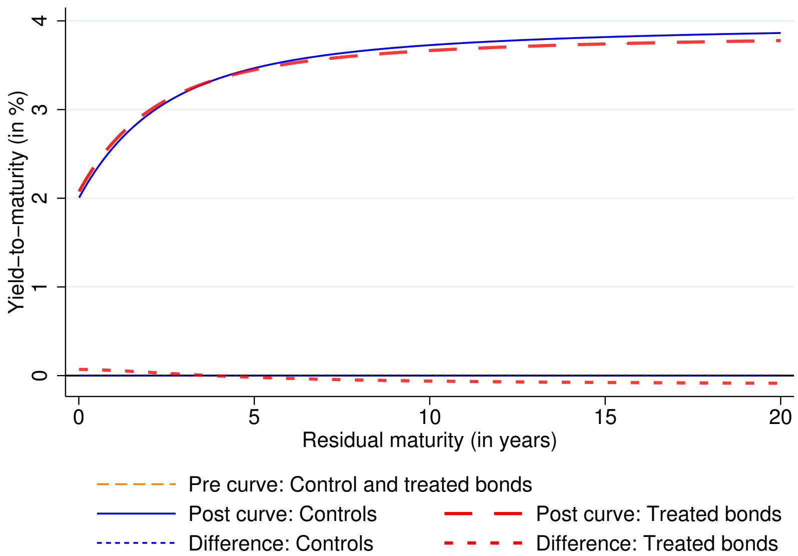


d) Idiosyncratic long-end effect: $m = 10$

								No. of coefficients	
Idiosyncratic effect		mode m	Mean	SD	Med	Min	Max	All	$\|t\| > 1.653$
a)	Short-end	0.25 years	-0.15	3.40	-0.14	-11.59	10.85	1,000	91
b)		10 years	12.24	3.04	12.26	2.30	22.99	1,000	992
c)	Long-end	0.25 years	0.15	3.60	0.13	-11.34	12.01	1,000	88
d)		10 years	-13.05	3.25	-13.08	-24.06	-2.59	1,000	991

Figure 3: False treatment effects graphically.

This figure shows estimated treatment effects based on the 1,000 families of sample couplets when the modeled term structure exhibits heterogeneous idiosyncratic effects across maturity but no true treatment effect present in the data. The specification is the same as in Table 5 and estimated with OLS. The (black) crosses in each plot show the 1,000 estimated DiD coefficients against the corresponding t-statistics. The vertical dashed (red) lines mark the values of ± 1.653, which correspond to two-sided confidence bands using a significance level of 10%. Subplots on the left (right) show the estimates when maturity of treated bonds is drawn from the triangular pdf with $m = 0.25$ ($m = 10$) years. The first (second) row of plots covers the idiosyncratic short-end (long-end) effect. The t-statistics are based on standard errors clustered at the bond level.

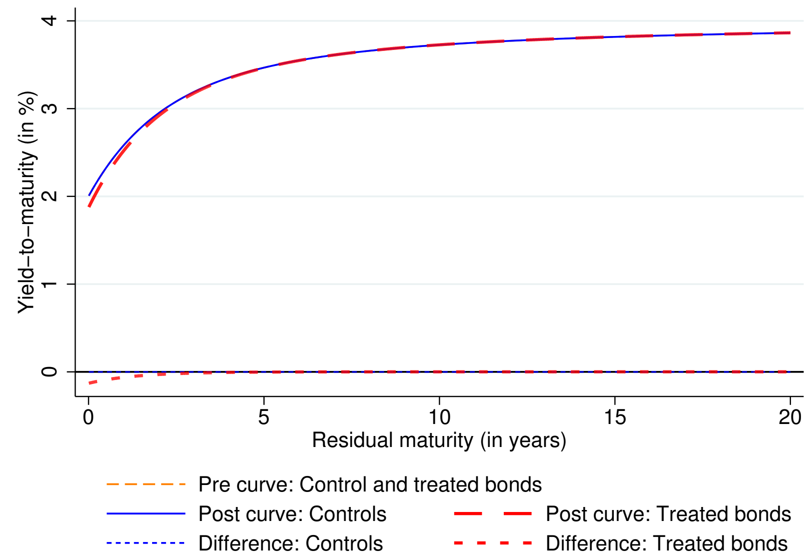


a) Yield-curve treatment twist:

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:

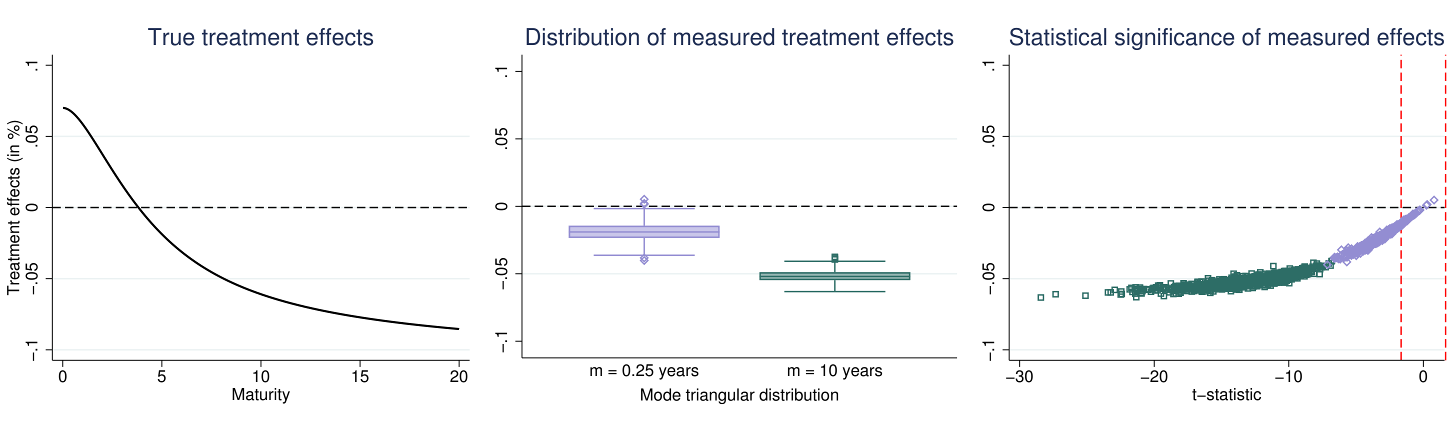
i) Parameter values	γ_0	γ_1	γ_2	λ
Pre-curve treated	4.000	-2.000	0.000	0.7308
Post-curve treated	4.001	-2.131	0.130	0.7308

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

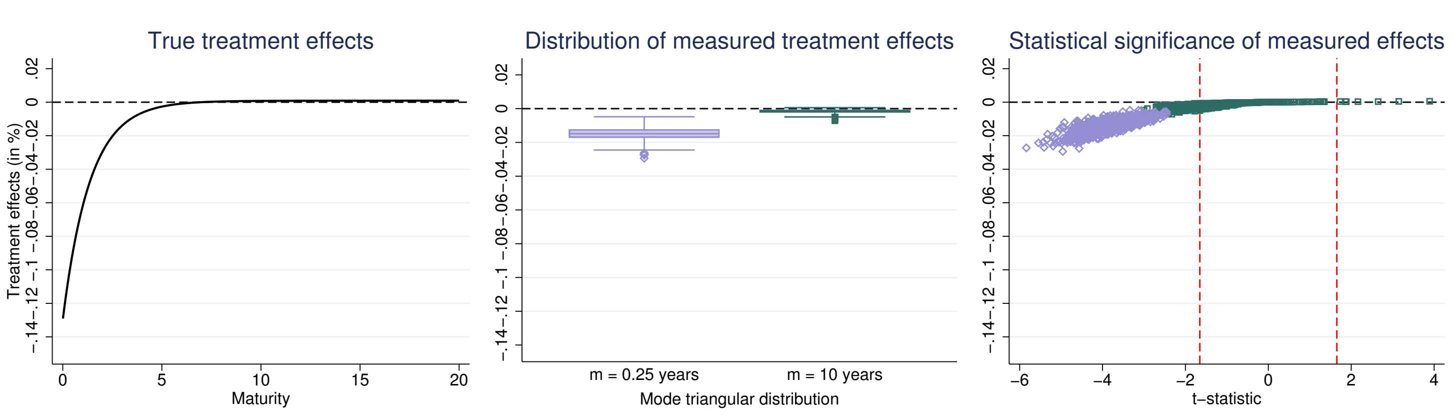
Figure 4: Modeling term-structure treatment effects.

To model the term structure we employ Diebold and Li (2006)'s 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. Figures 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

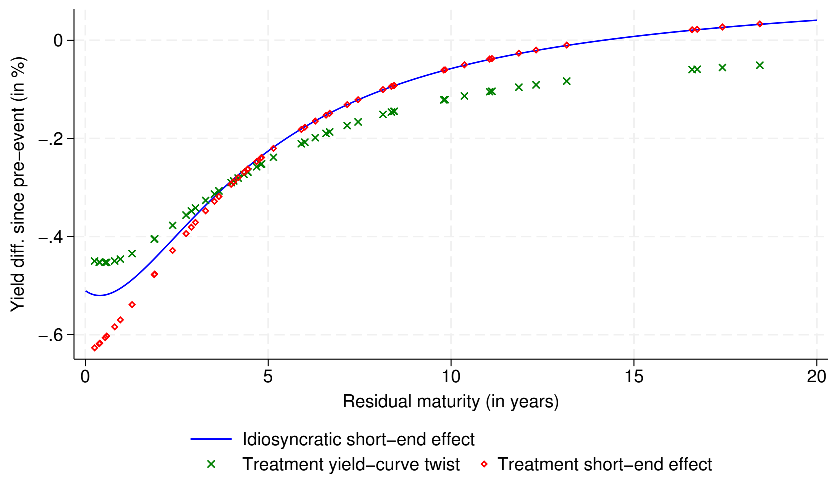


b) Treatment short-end effect

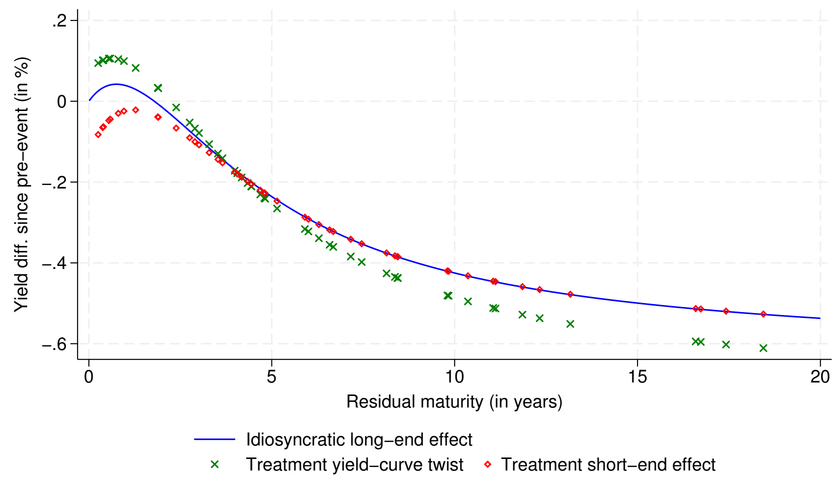
$m = 0.25$ years: $\widehat{\beta}_{DiD} \in [-2.94, -0.48]$ bps, $|t| > 1.653$: 1000 $m = 10$ years: $\widehat{\beta}_{DiD} \in [-0.72, +0.06]$ bps, $|t| > 1.653$: 301

Figure 5: Mismeasured treatment effects graphically.

Figures 5a and 5b show true and measured treatment effects on the 1,000 families of sample couplets for yield-curve treatment twist and treatment short-end effect, respectively, using OLS to estimate the same specification as in Table 5. 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.



a) Idiosyncratic short-end effect



b) Idiosyncratic long-end effect

Figure 6: Illustration of semi-matching.

This illustration is based on a random sample couplet when $m = 0.25$ years for both control and treated bonds. Figures 6a and 6b provide graphical illustrations for semi-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 at the short-end.

Internet Appendix

Robust difference-in-differences analysis when there is a term structure[1]

Kjell G. Nyborg

University of Zurich,
Swiss Finance Institute,
and CEPR

Jiri Woschitz

BI Norwegian Business School

August 2023

[1]Nyborg: Department of Banking and Finance, University of Zurich, Plattenstrasse 14, CH-8032 Zurich, Switzerland. email: kjell.nyborg@bf.uzh.ch.
Woschitz: Department of Finance, BI Norwegian Business School, B4y, NO-0442 Oslo, Norway. email: jiri.woschitz@bi.no.

A.1. Appendix

A.1.1 Fully flexible yield-curve DiD: Example

To illustrate the functioning of the fully flexible yield-curve DiD specification, in this appendix we apply it to one sample couplet j of the simulated data for the eight combinations of idiosyncratic effects (no effect, short-end effect, long-end effect) and treatment effects (no effect, twist, short-end effect) leaving out the no effect/no effect combination.

To generate the true underlying yield curves in the simulated data of the paper's main body, we have chosen values for the parameters γ_0, γ_1, γ_2, and λ and have plugged them into Diebold and Li's (2006) spot curve in Specification (2). Table A.1, Panel A, collects these parameter values of the spot curves from Figures 2 and 4 as well as Table 9 for the eight combinations of yield-curve movements. However, while the spot curve parameters of the control bonds prior to the treatment, $\{\beta_{k,1}\}_{k=0}^{2}$, measure the same quantity as the gammas, γ_0, γ_1, and γ_2, in Specification (6), the $\{\beta_{k,j}\}_{k=0}^{2}$ for $j = 2, \ldots, 4$ (for treated pre, control post, and treated post) represent *differential curves* and are therefore quantities that differ from the corresponding gammas in Specification (2). In Table A.1, Panel A, since we want to compare estimated values to the true underlying parameter values, the γ-representation from Specification (2) is transformed into the β-representation in Specification (6).

Insert Table A.1 here.

In Panel A, except for λ, parameter values that are not zero are highlighted in bold. Panel B shows the result of estimating Specification (6) using NLS for the eight combinations of yield-curve effects using family couplet j of the simulated data. In Panel B, coefficients that are statistically significantly different from zero at a significance level of at least 1% are also marked in bold. Comparing Panel A, which provides the true underlying values for the βs in Specification (6), with Panel B, providing the estimated coefficients for sample couplet j, shows that the bold non-zero values in Panel A form the same pattern as the bold significant coefficients in Panel B. Measured in percentage points, the parameter values and coefficients in bold in the two panels are the same (up to at least the third decimal digit after the comma). These results show the feasibility of identifying the true underlying parameter values using a simple but well specified regression model.

Interesting are a few exceptions, where the parameter estimates seem to be slightly different from the true parameter values. These incidences appear, on the one hand, with the curvature factor of the control bonds prior to treatment, $\widehat{\beta}_{2,1}$, and, on the other hand, with the decay parameter, $\widehat{\lambda}$, at the bottom of the panel. As explained by Diebold and Li (2006), the decay parameter determines the point where the loading of the curvature factor obtains its maximum. Hence, these two parameters have more multicollinearity with each other than each of them has with the other parameters, level and slope. This relationship can, for example, be seen by looking at the case of a short-end treatment effect. The more *downward* and away-tilted the estimated lambda, $\widehat{\lambda} = 0.7302$, from the true value, $\lambda = 0.7308$, the more *upward* and away-tilted is the pre-treatment control-bond curvature estimate, $\widehat{\beta}_{2,1} = 0.001$, is from the true parameter value, $\beta_{2,1} = 0.000$. This shows that the estimation might be exposed to multicollinearity between the yield-curve parameters and estimates might, therefore, be confounded to a certain extent. We will discuss this further shortly below.

A different question, however, is whether this is the right quantity to consider. By looking, for example, at the case of an idiosyncratic yield-curve short-end effect and a yield-curve treatment twist in the fifth regression in Panels A and B of Table A.1, a researcher learns: First, the differential curve of treated compared to control bonds prior to the treatment is zero. Second, level, slope, and curvature factors of the control bonds change by 0.140, -0.650, and -0.800, respectively, from pre to post treatment (which represents a short-end effect). Third, compared to the post-curve of control bonds, the level factor of the curve of treated bonds is -0.110 smaller and the slope and curvature factors each 0.180 larger (the additional yield-curve treatment twist). Clearly it is very difficult to grasp what this information economically means. Panels C and D in Table A.1 provide an alternative to presenting the same results in a more readable and intuitive manner.

Table A.1, Panel C, shows the true underlying treatment effects of treated bonds from pre to post treatment, controlling for movements in the yield curve of treated compared to control bonds and movements in the yield curve from pre to post treatment. The DiD is a function of maturity and, hence, varies across different maturities as long as the DiD is not a pure level-shift. Panel C shows that the DiD at selected maturities, 1, 2, 3, 5, 7, 10, and 15 years, are the same across the three cases of no idiosyncratic effect, an idiosyncratic short-end, and an idiosyncratic long-end effect. That is how the true underlying effects are modeled and is therefore correct.

In Panel D we present the results if we estimate the DiD by using the estimation results from Panel B and predicting the DiD at the same selected maturities. We use the delta method to calculate standard errors, which are also clustered on the bond level. Marginal effects that are statistically significantly different from zero (all at the significance level of at least 1%) are marked in bold. To help visualize the similarities between Panels C and D, the true underlying non-zero marginal effects in Panel C are as well highlighted in bold. The results show that measuring the DiD in percentage points, the true and estimated numbers are the same up to the third decimal digit after the comma. With respect to multicollinearity between the regressors as touched upon above, this shows that presenting the results this way is not impacted by multicollinearity anymore.

Furthermore, the measured quantity is intuitive to understand. For example, the fifth regression in Table A.1, Panel D, the one for which we tried to describe the results already above, shows that in case of a yield-curve short-end effect of the control bonds and an additional yield-curve treatment twist of the treated bonds, the treatment effect corresponds to +5.87, +3.75, +1.58, −1.87, −4.11, −6.09, and −7.71 bps at maturities of 1, 2, 3, 5, 7, 10, and 15 years, respectively.

While this appendix section illustrated how to estimate a meaningful quantity using a random sample couplet j of simulated data, the paper applies the method to all simulated sample couplets and shows its power to eliminate both false and mismeasured treatment effects in fixed-income settings.

Table A.1: Illustration of the fully flexible yield-curve DiD specification.

This table illustrates the fully flexible yield-curve DiD specification by applying it to one sample draw of simulated data. This is done for the eight combinations of idiosyncratic (no, short-end, long-end effect) and treatment effects (no, twist, short-end effect), leaving out the no effect/no effect combination. The specification is $yield_{it} = \mathbf{B}_1' \, \mathbf{L}_{it} + \mathbf{B}_2' \, \mathbf{L}_{it} \, \mathbb{1}_{Treated,i} + \mathbf{B}_3' \, \mathbf{L}_{it} \, \mathbb{1}_{Post,t} + \mathbf{B}_4' \, \mathbf{L}_{it} \, \mathbb{1}_{Treated,i} \times \mathbb{1}_{Post,t} + \varepsilon_{it}$, where notation is as in Table 5 and, additionally, $\mathbf{B}_j$ are three-dimensional vectors of coefficients, with individual elements $\beta_{k,j}$, $k = 0, \ldots, 2$. They measure 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 ($j = 4$). $\mathbf{B}_4$ is the DiD estimator and depends on maturity x. Panel A collects the parameter values of γ_k, $k = 0, ..., 2$, and λ of the spot curves from Figures 2 and 4 and Table 9 and transforms the γ-representation in those tables to the β-representation. In Panel A, except for λ, parameter values that are not zero are highlighted in bold. Panel B shows the result of estimating the specification using NLS for the eight combinations of yield-curve developments using the one sample draw of simulated data. The decay parameter λ is assumed to be time invariant and the same for treated and control bonds. In Panel B, coefficients that are statistically significantly different from zero (all at the significance level of at least 1%) are marked in bold. Panel C shows $\mathbf{B}_4$, the true underlying DiD of treated bonds post-treatment, for selected maturities. In Panel C, parameter values that are not zero are highlighted in bold. Panel D shows $\widehat{\mathbf{B}}_4$, estimated using the estimation results from Panel B and predicting the DiD at the same selected maturities as in Panel C. Standard errors are clustered at the bond level and calculated using the delta method. Coefficients that are statistically significantly different from zero (all at the significance level of at least 1%) are marked in bold. In Panels B and D, a, b, and c denote significance (two-sided) at the levels of 1%, 5%, and 10%, respectively.

Panel A: True parameter values

Idiosyncratic effect	Short-end	Long-end	–		Short-end		Long-end	
Treatment effect	–	–	Twist	Short-end	Twist	Short-end	Twist	Short-end
$\beta_{0,1}$	**4.000**	**4.000**	**4.000**	**4.000**	**4.000**	**4.000**	**4.000**	**4.000**
$\beta_{1,1}$	**-2.000**	**-2.000**	**-2.000**	**-2.000**	**-2.000**	**-2.000**	**-2.000**	**-2.000**
$\beta_{2,1}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{0,2}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{1,2}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{2,2}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{0,3}$	**0.140**	**-0.650**	0.000	0.000	**0.140**	**0.140**	**-0.650**	**-0.650**
$\beta_{1,3}$	**-0.650**	**0.650**	0.000	0.000	**-0.650**	**-0.650**	**0.650**	**0.650**
$\beta_{2,3}$	**-0.800**	**1.000**	0.000	0.000	**-0.800**	**-0.800**	**1.000**	**1.000**
$\beta_{0,4}$	0.000	0.000	**-0.110**	**0.001**	**-0.110**	**0.001**	**-0.110**	**0.001**
$\beta_{1,4}$	0.000	0.000	**0.180**	**-0.131**	**0.180**	**-0.131**	**0.180**	**-0.131**
$\beta_{2,4}$	0.000	0.000	**0.180**	**0.130**	**0.180**	**0.130**	**0.180**	**0.130**
λ	0.7308	0.7308	0.7308	0.7308	0.7308	0.7308	0.7308	0.7308

Table A.1 – *continued*

Panel B: Estimated parameter values for family j of ordered sample couplets

Idiosyncratic effect	Short-end	Long-end	–		Short-end		Long-end	
Treatment effect	–	–	Twist	Short-end	Twist	Short-end	Twist	Short-end
$\widehat{\beta}_{0,1}$	**4.000**^{a}	**4.000**^{a}	**4.000**^{a}	**4.000**^{a}	**4.000**^{a}	**4.000**^{a}	**4.000**^{a}	**4.000**^{a}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\beta}_{1,1}$	**-2.000**^{a}	**-2.000**^{a}	**-2.000**^{a}	**-2.000**^{a}	**-2.000**^{a}	**-2.000**^{a}	**-2.000**^{a}	**-2.000**^{a}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\beta}_{2,1}$	-0.000	-0.000	-0.000	0.001	0.000	-0.000^{b}	-0.000	-0.000^{c}
	(0.317)	(0.148)	(0.889)	(0.333)	(0.908)	(0.036)	(0.384)	(0.068)
$\widehat{\beta}_{0,2}$	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.170)	(0.170)	(0.171)	(0.170)	(0.170)	(0.170)	(0.170)	(0.170)
$\widehat{\beta}_{1,2}$	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.466)	(0.466)	(0.467)	(0.463)	(0.466)	(0.466)	(0.466)	(0.466)
$\widehat{\beta}_{2,2}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.117)	(0.117)	(0.117)	(0.116)	(0.117)	(0.117)	(0.117)	(0.117)
$\widehat{\beta}_{0,3}$	**0.140**^{a}	**-0.650**^{a}	0.000	0.000	**0.140**^{a}	**0.140**^{a}	**-0.650**^{a}	**-0.650**^{a}
	(0.000)	(0.000)	(0.858)	(0.980)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\beta}_{1,3}$	**-0.650**^{a}	**0.650**^{a}	0.000	-0.000	**-0.650**^{a}	**-0.650**^{a}	**0.650**^{a}	**0.650**^{a}
	(0.000)	(0.000)	(0.953)	(0.845)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\beta}_{2,3}$	**-0.800**^{a}	**1.000**^{a}	-0.000	0.000	**-0.800**^{a}	**-0.800**^{a}	**1.000**^{a}	**1.000**^{a}
	(0.000)	(0.000)	(0.877)	(0.972)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\beta}_{0,4}$	0.000^{c}	0.000	**-0.110**^{a}	**0.001**^{a}	**-0.110**^{a}	**0.001**^{a}	**-0.110**^{a}	**0.001**^{a}
	(0.094)	(0.227)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\beta}_{1,4}$	0.000	-0.000	**0.180**^{a}	**-0.131**^{a}	**0.180**^{a}	**-0.131**^{a}	**0.180**^{a}	**-0.131**^{a}
	(0.576)	(0.941)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\beta}_{2,4}$	-0.000^{c}	-0.000	**0.180**^{a}	**0.130**^{a}	**0.180**^{a}	**0.130**^{a}	**0.180**^{a}	**0.130**^{a}
	(0.050)	(0.411)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\lambda}$	0.7308^{a}	0.7308^{a}	0.7308^{a}	0.7302^{a}	0.7307^{a}	0.7309^{a}	0.7308^{a}	0.7309^{a}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Num. of obs.	200	200	200	200	200	200	200	200
Adjusted R^2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RMSE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table A.1 – *continued*

Panel C: True difference-in-differences at selected maturities									
Idiosyncratic effect		Short-end	Long-end	–		Short-end		Long-end	
Treatment effect		–	–	Twist	Short-end	Twist	Short-end	Twist	Short-end
Maturity	1	0.0000	0.0000	**0.0587**	**-0.0623**	**0.0587**	**-0.0623**	**0.0587**	**-0.0623**
(in years)	2	0.0000	0.0000	**0.0375**	**-0.0297**	**0.0375**	**-0.0297**	**0.0375**	**-0.0297**
	3	0.0000	0.0000	**0.0158**	**-0.0139**	**0.0158**	**-0.0139**	**0.0158**	**-0.0139**
	5	0.0000	0.0000	**-0.0187**	**-0.0026**	**-0.0187**	**-0.0026**	**-0.0187**	**-0.0026**
	7	0.0000	0.0000	**-0.0411**	**0.0000**	**-0.0411**	**0.0000**	**-0.0411**	**0.0000**
	10	0.0000	0.0000	**-0.0609**	**0.0008**	**-0.0609**	**0.0008**	**-0.0609**	**0.0008**
	15	0.0000	0.0000	**-0.0772**	**0.0009**	**-0.0772**	**0.0009**	**-0.0772**	**0.0009**
Panel D: Estimated difference-in-differences at selected maturities for family j of ordered sample couplets									
Maturity	1	0.0000	0.0000	**0.0587**[a]	**-0.0623**[a]	**0.0587**[a]	**-0.0623**[a]	**0.0587**[a]	**-0.0623**[a]
(in years)		(0.895)	(0.706)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	2	-0.0000	-0.0000	**0.0375**[a]	**-0.0297**[a]	**0.0375**[a]	**-0.0297**[a]	**0.0374**[a]	**-0.0297**[a]
		(0.231)	(0.958)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	3	-0.0000	-0.0000	**0.0158**[a]	**-0.0139**[a]	**0.0158**[a]	**-0.0139**[a]	**0.0158**[a]	**-0.0139**[a]
		(0.139)	(0.974)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	5	-0.0000	0.0000	**-0.0187**[a]	**-0.0026**[a]	**-0.0187**[a]	**-0.0026**[a]	**-0.0187**[a]	**-0.0026**[a]
		(0.280)	(0.607)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	7	-0.0000	0.0000	**-0.0411**[a]	**0.0000**[a]	**-0.0411**[a]	**0.0000**[a]	**-0.0411**[a]	**0.0000**[a]
		(1.000)	(0.231)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
	10	0.0000	0.0000	**-0.0609**[a]	**0.0008**[a]	**-0.0609**[a]	**0.0008**[a]	**-0.0609**[a]	**0.0008**[a]
		(0.323)	(0.157)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	15	0.0000	0.0000	**-0.0772**[a]	**0.0009**[a]	**-0.0771**[a]	**0.0009**[a]	**-0.0771**[a]	**0.0009**[a]
		(0.163)	(0.174)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Num. of obs.		200	200	200	200	200	200	200	200
Adjusted R^2		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RMSE		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000