TI 2025-037/III
Tinbergen Institute Discussion Paper

Measuring and Explaining the CDS-Bond Basis Term-Structure Shape and Dynamics

Yonas Khanna[1]
André Lucas[2]
Norman Seeger[3]

1 ING Bank, Vrije Universiteit Amsterdam

2 Vrije Universiteit Amsterdam, Tinbergen Institute

3 Vrije Universiteit Amsterdam, Tinbergen Institute

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Measuring and Explaining the CDS-Bond Basis Term-Structure Shape and Dynamics*

Yonas Khanna[1,2,†], André Lucas[2,3], Norman Seeger[2,3]

[1] *ING Bank, The Netherlands*
[2] *Vrije Universiteit Amsterdam, The Netherlands*
[3] *Tinbergen Institute, The Netherlands*

This version: May 26, 2025

Abstract

The CDS-bond basis quantifies the difference in risk premia between credit default swap (CDS) and bond markets. It is hard to measure at the individual firm level given substantial missing-value problems (30%-100%) in either or both markets, even for highly liquid blue-chip financial firms. We propose a novel imputation approach to obtain full historical firm-level basis term-structures across all maturities. Our approach can accommodate different term-structure interpolation methods, including Nearest-Neighbor, spline, and Nelson-Siegel interpolation. Using the new methodology, we construct the full history of the 2011-2021 JP Morgan (JPM) basis term-structure and use it to analyze its empirical determinants. We find that factors like market liquidity, funding liquidity, counterparty risk, and the default premium all impact the basis term-structure, though not all at the same moment in time. All factors are statistically significant during the Covid-19 pandemic. The various empirical limits-to-arbitrage proxies correlate differently with different parts of the basis term-structure, stressing the need to model the full basis term-structure rather than assuming it to be flat. The results are robust for other blue-chip financials, each time requiring the full basis term-structure imputation approach as proposed in this paper.

Keywords: CDS-bond basis, missing value imputation, high-dimensional panel data, multi-curve modeling, time-varying spline interpolation, dynamic Nelson-Siegel, Kalman filter.

JEL classifications: C32, C33, C58, G12, G32.

*We thank conference and seminar participants at the 2024 Sveriges Riksbank Research Seminar (Stockholm), 2024 Netherlands Econometric Study Group Seminar (Maastricht University), 2024 Annual Conference of the International Association for Applied Econometrics (Xiamen University) and 76th European Meeting of the Econometric Society (Erasmus University Rotterdam).

†Khanna is grateful for the financial support provided by ING Bank. The views expressed in this paper are those of the authors and do not necessarily reflect the views ING Bank.

1 Introduction

For a given issuer and time-to-maturity, the *CDS-bond basis* captures the difference in credit risk premia between credit default swap (CDS) and bond markets. It is often taken as a proxy for credit market conditions by both policy makers and market participants, as a high basis implies a less liquity credit market (see, e.g., Boyarchenko et al., 2022). Theoretically, no-arbitrage relationships should lead to a CDS-bond basis of zero (see Duffie, 1999; Hull and White, 2000). Empirically, however, a zero basis is hardly ever observed. More often, credit spreads in bond markets as measured by so-called zero-volatility spreads or Z-spreads exceed credit spreads in CDS markets; see Neftci, 2008; Green, 2015.

Several studies have tried to relate the existence of a nonzero basis to market frictions or limits-to-arbitrage factors, including liquidity costs, funding costs, and counterparty risk differences (see de Wit, 2006; Bai and Collin-Dufresne, 2018; Augustin and Schnitzler, 2020). Most of these studies, however, focus on the cross-sectional properties of the basis, concentrating on a single (typically most liquid) maturity level (e.g., 5-years; see Bai and Collin-Dufresne, 2018; Augustin and Schnitzler, 2020). Much less attention has been paid to the shape of the basis term-structure across maturities at the firm-level, nor to the time-series properties of the basis term-structure as a whole. This is a substantial gap in the literature given the basis term-structure's importance in financial markets and regulation. In trading, understanding the time-series dynamics of the full basis term-structure is of significance importance for institutional traders, as their trading strategies often focus on fluctuations in the bases rather than on its absolute level (Risk, 2024). In derivatives markets, the firm-level basis is a key proxy for funding costs in for instance Funding Value Adjustment (FVA) calculations. For FVA, a funding spread (funding liquidity premium), that can be interpreted as a cost component, is required for the funding valuation frameworks that are proposed in Morini and Prampolini (2011), Burgard and Kjaer (2012) and Sommer et al. (2013) and see Green (2015) for a non-exhaustive overview. These costs are a standard part of derivative prices to account for counterparty default risk during the contract's lifetime; see, e.g., Gregory (2020) and Green (2015). Also from a regulatory perspective the basis is an important ingredient for risk calculations, in particular when quantifying banks' Basel III (BCBS, 2010) capital requirements for Over-the-Counter (OTC) transaction profits and losses. Jointly, the size of the derivatives and OTC markets is enormous, underlining the relevance of good basis

measurements. Moreover, since funding costs depend on the maturity of the underlying contract, it is also essential to have the basis term-structure across the whole range of possible maturities rather than at a single maturity level only. This stands in sharp contrast to current practices, where ad-hoc assumptions are often made, such as the existence of a flat basis term-structure.

The lack of attention for the full shape of the basis term-structure originates in a severe missing data problem. In particular, we are faced with a lack of corporate bond data and/or CDS data to accurately measure the basis. This impedes a more thorough understanding of (i) the shape of the basis term-structure at any moment in time, (ii) the time-series behavior of the basis term-structure, and (iii) the relationship between *the different maturity domains* of the basis term-structure and those variables that explain the occurrence of limits-to-arbitrage for situations where individual institutions want to trade on widening firm level spreads. In particular, it remains unclear, for instance, whether the same limits-to-arbitrage factors drive the basis at the short, middle, and long end of the curve, or whether basis levels are affected similarly across the entire, say 30-year, maturity spectrum.

In this paper, we propose a way forward to these challenges by treating the shortage of corporate bond and possibly CDS spread data as a missing value problem. For this, we introduce a novel firm-level state-space model that imputes credit spread data in both CDS and bond markets simultaneously using any available parametric or non-parametric term-structure interpolation mechanism. To facilitate imputation we model the shape of the basis curve in the following way. We first start with a level component, subsequently add a CDS-bond basis component which is either jointly or individually driven by the credit and/or bond market. Furthermore, we include components which explain the term-structure shape of the CDS-bond basis and by doing so we account for the CDS-bond basis different quantities with varying maturities. By modeling the CDS-bond basis directly we circumvent modeling the CDS spread and bond spread term-structure individually which would be problematic because of the scarce data problem. This approach automatically gives rise to measurements of the CDS-bond basis term-structure across all maturities and at the same time exploits the historical strong comovement between CDS and bond spreads. For evidence on comovement between CDS and bond spreads see, e.g., Lange et al. (2017) and de Wit (2006). Our imputation approach is fully empirically driven and does not impose any curve related no-arbitrage conditions. That is, we impose as little as possible modeling structure on the spread curves which, let the empirical data drive

results as much as possible and by that force the imputed values to capture realized market price and basis curve movements as accurately as possible. The final models remains in the class of linear Gaussian state-space models and is therefore easily estimated using maximum likelihood methods and the Kalman filter (Durbin and Koopman, 2012).

Our model produces two relevant outputs that can be studied in more depth. First, the estimated model produces a full historical record of the basis term-structure at the firm-level. We can use this to explore the determinants of the basis for individual firms and maturities. Second, the model generates complete histories of CDS and Z-spread term-structures, along with complementary confidence bands. Both of these are relevant inputs for market participants in risk management applications. For instance, they can be used to capture credit risk based on historical value-at-risk calculations for defaultable bonds (BCBS, 2019, 2010; Sarig and Warga, 1989; Houweling et al., 2005), or for fair-value computations in derivative pricing under Credit Value Adjustments (CVAs) and FVAs (Green, 2015; Morini and Prampolini, 2010; Pallavicini et al., 2011). The model and its different outputs may thus be of considerable interest to both academics practitioners.

In our empirical analysis we proceed in two major parts: First, we construct the CDS bond basis out of the data and second, we analyze the time-series dynamics of the CDS bond basis.

In the first part, we illustrate the model to construct the basis curve dynamics for JPMorgan Chase & Co (JPM).[1] The data covers the period Jan 2011 to Dec 2021. The missing data problem is quite pronounced, with missing values in basis time-series ranging from 38%-100% across the 30-years maturity spectrum. This implies that even for a highly liquid blue-chip financial like JPM, the imputation challenge is substantial. We find that the introduction of long-term and short-term components in the model significantly improves the model's in-sample and out-of-sample performance for the basis. This suggests a different behavior of the long and short end of the basis curve, and therefore provides evidence of a (non-flat) basis term-structure. One of the key modeling choices for backing out the CDS bond basis is an interpolation technique used to filter the term structure of the basis. For robustness we implement several of the prime interpolation candidates and find that the Nelson-Siegel and the piece-wise spine interpolation scheme outperform other possible choices. Both the parametric and non-parametric interpolation methods consistently produce well-behaved shapes for the bases, even when credit

[1] We leave the development of a fully integrated multi-name model to a future paper.

curves are partially or completely missing. In particular, our dynamic multi-curve piecewise linear spline model maintains accuracy over time and correctly captures a variety of historically observed basis term-structure shapes, including flat, upward-sloping and downward-sloping basis term-structures.

In the second part of our analysis, after having recovered the JPM basis term-structure history, we proceed by investigating the basis time-series dynamics across maturities and its empirical determinants. The analysis is novel in the sense that we study the CDS bond basis dynamics and term structure for a single firm compared to a cross-sectional analysis over many firms with one single maturity that has been done in the literature so far. Here, the CDS bond basis dynamics are related to limits-to-arbitrage factors as studied in for instance Augustin and Schnitzler (2020) and Bai and Collin-Dufresne (2018). Employing rolling-window regressions, we learn that JPM basis changes are primarily driven by changes in funding costs and by JPM related sentiment. Particularly, during the Covid-19 pandemic, all factors were statistically significant. Such regressions, interestingly, also reveal differences in impact and significance across the maturity spectrum of the basis term-structure. Here we find that (i) counterparty risk briefly caused the short-term basis to steepen; (ii) market and funding liquidity relate to flight-to-quality effects in short-term, (iii) the default premium relates more to the short end of the basis curve; and (iv) idiosyncratic sentiment and funding costs pushed the medium-and long-term end of the basis term-structure steeper throughout the pandemic.

We stress again that these results are only made possible by the imputation modeling approach across all maturities as proposed in this paper as this provides the comprehensive basis term-structure history, which can then be analysed in a second stage. Again the findings are robust to the precise choice of interpolation technique (e.g., spline-based versus parametric) within the state-space framework, and to the precise financial institution studied. Results for Bank of America, Goldman Sachs, and Morgan Stanley, show very similar patterns compared to JPM. The relationships between Bank of America, Goldman Sachs, and Morgan Stanley bases term-structure and the limits-to-arbitrage factors were even somewhat more pronounced during credit crunch period, especially in terms of funding costs and idiosyncratic sentiment.

The remainder of the paper is organized as follows. Section 2 discusses the CDS-bond basis arbitrage mechanism, the empirical basis data and its potential drivers. Section 3 introduces the modeling framework for filtering basis term-structures. Section 4 presents the in-sample and out-of-sample performance for JPM and the empirical determinants of

its basis term-structure dynamics. Section 5 concludes.

2 CDS-bond basis framework: basis arbitrage, sparse curves, and basis determinants

2.1 The CDS-bond basis arbitrage

A CDS is an Over-the-Counter (OTC) financial derivative where the protection seller agrees to compensate the protection buyer if the reference entity defaults before contract maturity. The holder of the CDS contract pays the seller periodic premiums known as the CDS spread. Arbitrage-free theory indicates that the CDS spread must match the credit spread on the deliverable bond with same maturity (see, e.g., Duffie, 1999; Hull and White, 2000). Market frictions, however, can cause the two spreads to diverge, giving rise to the CDS-bond basis.

Unlike CDS spreads, bond credit spreads are not directly observable and have to be approximated, accounting for bond features (such as callability, collateral provisions, floating rates,[2] etc.). There are multiple ways bond credit spreads can be measured. To calculate the credit spreads we adopt the most widely used market standard, the so-called zero-volatility spreads (Z-spread), that only requires plain vanilla bonds for calculation (Green, 2015). The Z-spread reflects the credit spread as a constant top-up that needs to be applied to the risk-free rate curve, such that the prices of the vanilla bonds equal the present value of their cash flows (Neftci, 2008). In our paper, we use CDS spread $\left(y_t^{\text{CDS}}\right)$ and Z-spread $\left(y_t^{\text{Z}}\right)$ data to compute the CDS-bond basis:

$$b_t(\tau) = y_t^{\text{CDS}}(\tau) - y_t^{\text{Z}}(\tau), \tag{1}$$

where $b_t(\tau)$ is the basis of maturity τ at time t.

Although, the basis is typically small in normal times due to no-arbitrage conditions, there may be market circumstances and factors that restrict arbitrageurs to drive Z-spreads and CDS spreads closer together (Augustin and Schnitzler, 2020; Bai and Collin-Dufresne, 2018; Duffie, 1999). For example, when the CDS-bond basis is negative, arbitrageurs are expected to exploit a risk-free profit opportunity (i.e., basis) by purchas-

[2]The theoretical corporate bond spread is calculated as the difference between the yield of a floating-rate note and risk-free rate of the same maturity(Duffie, 1999). This ideal spread, however, is rarely observed due to low issuance volume (Bai and Collin-Dufresne, 2018).

ing the reference bond and credit protection on it. This trade is known as the *negative*[3] basis trade and profit and loss (PnL) of prospects of this trade are well studied in the literature. In particular, the dynamics of this trade's PnL are influenced by various market frictions that hinder arbitrageurs from fully capturing the credit pricing discrepancy between the two markets. These frictions include funding and liquidity costs, haircuts, margin calls, counterparty risk, and collateral quality (see Bai and Collin-Dufresne, 2018). As a result, the basis is hardly ever found to be zero in the empirical data.

The market frictions (or factors) mentioned above fall under the umbrella of the 'limits-to-arbitrage' theories outlined in detail by Augustin and Schnitzler (2020) and Bai and Collin-Dufresne (2018). They test the relationship between basis determinants and the cross-sectional variation in bases during the 2008 Financial Crisis using (unbalanced) panel data of U.S. firm bond-level bases across a large group of industries. Focusing on a 5 year maturity, they include multiple bonds for each firm as it is challenging to create a complete historical time series due to the lack of corporate bond data. It is precisely this missing data problem that leads most studies to use small-T unbalanced panels and a single, most liquid 5Y maturity to investigate which factors affect the basis. To the best of our knowledge, for individual firms there are no studies on the entire firm-level basis term-structure, nor on its dynamics and its determinants. Consequently, it remains unclear whether factors that explain the cross-section of the basis at a single maturity also explain the firm-level basis and its dynamics across the full 30Y maturity spectrum. This is precisely the focus of our current paper.

2.2 Missing data challenge: empirical bond and CDS data

In this section we highlight the missing data problem that arises when computing the daily (constant maturity) CDS-bond basis at the firm level. The typical root of the problem is the lack of corporate bond data, though also CDS data may be missing and require some form of interpolation; see Gregory (2020), BIS (2016), BCBS (2004). As a case study, we focus on the CDS-bond basis of JPMorgan Chase (JPM). JPM is a highly liquid blue-chip financial institution, issuing a range of corporate bonds at different maturities. At the same time, a range of CDS contracts is traded on JPM given its systemic importance.

[3]A positive basis trade involves short-selling the reference bond, while simultaneously selling credit protection on it. This trade is unlikely to be executed in practice, as its PnL is subject to systematic risk components, see for instance Augustin and Schnitzler (2020).

We take daily USD denominated CDS spread quotes on JPM from Markit^{TM}. Following Augustin and Schnitzler (2020), CDS contracts of Senior Unsecured debt with a Modified Restructuring Clause are selected to minimize the impact of the 'cheapest-to-deliver' option. The daily CDS term-structure consists of nine 'on-the-run', near constant maturity points at 6M, 1Y, 3Y, 5Y, 7Y, 10Y, 20Y and 30Y. The exact daily time-to-maturity of these series is subject to rollover dates; see Appendix A for technical details and for a schematic representation of the maturities through time. The daily observations span an 11 year period from January 3, 2011 to December 31, 2021 ($T = 2870$).

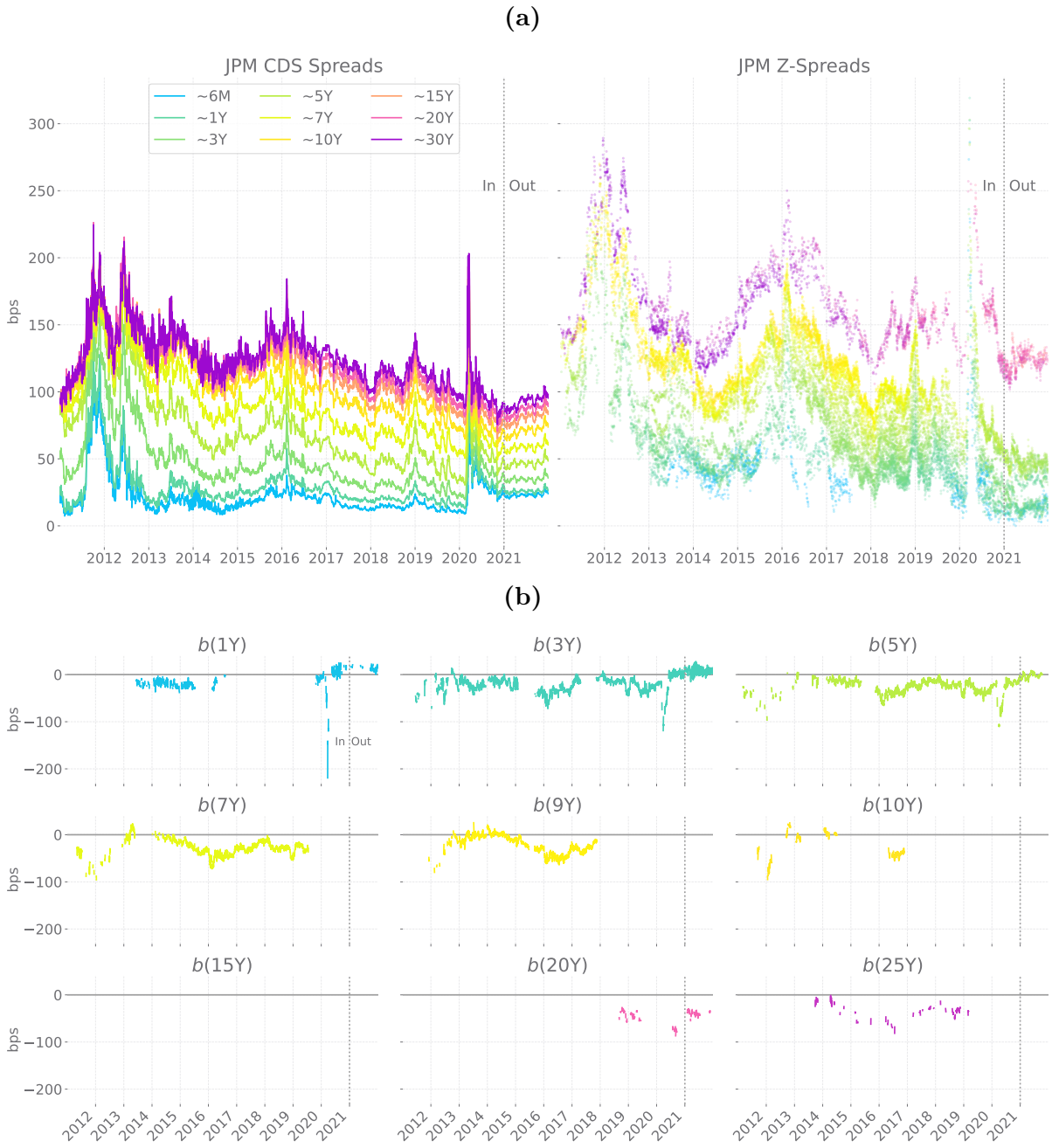
We also take daily bond composite Z-spread data from Markit^{TM}, which calculates spreads relative to the USD LIBOR curve Markit^{TM} aggregates bond prices from multiple sources to generate composite-level spreads. The number of price contributions serves as a liquidity measure, with higher contributions indicating lower pricing uncertainty. At each time point, we include a bond in our analysis only if it has at least five price contributions. In addition, we focus on Z-spreads derived from vanilla reference bonds that align with the CDS credit risk profile. Specifically, we select uncollateralized reference bonds consisting of Senior Unsecured debt with investment-grade ratings and remaining maturities between 6M and 30Y. We exclude option-adjusted spreads (OAS) for callable[4] bonds—the Z-spread equivalent for such bonds—because isolating the option risk premium introduces model risk associated with option pricing, making the resulting Z-spread less reliable. As a result of our filtering process, we obtain 31 unique Z-spread series with varying start dates across the full sample period. The complete list of bond ISIN identifiers are reported in Appendix A. The appendix also visualizes the time-to-maturities of the bonds over time.

To give some insights into the missing data challenge, even for a liquid name such as JPM, Figure 1 shows the daily CDS spread and Z-spread data of JPM together with the empirical basis estimates for various maturities. The daily constant maturity basis estimates are obtained non-parametrically via linear interpolation in the following way. First, we obtain constant maturity CDS spreads by linearly interpolating/extrapolating the daily CDS spread curves (separately). Hence, we recognize that CDS maturities are subject to rollover dates.

Second, constant maturity Z-spreads are obtained by linearly interpolating between

[4]Limiting analysis to a specific type of bond is common in the CDS-bond basis literature (e.g., Augustin and Schnitzler (2020) focus exclusively on floating-rate bonds) because differences between bond types make the basis difficult to compare (Bai and Collin-Dufresne, 2018).

Figure 1: CDS, Z-Spread and Basis data points



Notes: Subfigure 1a displays the daily CDS spread (left) and Z-spread observations (right) in basis points (bps), colored by their time-to-maturity, for JPM over the period 2011-2022. The black dotted line marks the split between our in-sample period (January 3, 2011 to December 31, 2021) and out-of-sample period (January 1, 2021 to December 31, 2021). Subfigure 1b displays the non-parametric constant maturity estimates for the CDS-bond basis for various tenors. The basis time-series are incomplete due to missing values arising from a shortage of bond Z-spread data at various maturities. The scarcity of bond Z-spread values is largest for (very) short-dated ($\leq$ 3Y) and long-dated ($\geq$ 10Y) underlying bonds.

the two closest Z-spreads surrounding the target tenor. The interpolation is only performed within a suitable range around the target tenor. The boundaries of the interpolation intervals are set to the two closest Markit^{TM}'s CDS tenors. For example, the 1Y Z-spread tenor is computed by interpolating samples with maturities within the range $[6M, 3Y)$. Thus, for the 5Y Z-spread we interpolate samples in $[3Y, 7Y)$, which is in line with the domain used in Augustin and Schnitzler (2020) and (Bai and Collin-Dufresne, 2018). Z-spreads at 6M and 30Y maturities are not computed, because they require extrapolation, which has poor performance due to a lack of data. Table 1 provides an overview of the interpolation ranges used for computing Z-spreads tenors.

The missing data problem for the CDS part of the basis is less pronounced: the CDS maturities are at most 6 months off from the corresponding Markit^{TM}'s CDS tenors, in contrast to the bond data, which is particularly scarce at the 6M and 30Y boundries. See Appendix A for more details.

Figure 1 clearly reflects the evolution of the U.S. economy and credit markets over the past decade, including the highly volatile periods of 2011–2012 and the Covid-19 pandemic in 2020. Interestingly, the upper end of JPM's CDS curve inverts during for instance 2012, with the 20Y yield exceeding the 30Y yield. This can be either a sign of illiquidity, or an inversion following market wide conditions at the time.[5] The phenomenon does not occur in the bond market at that time: there the Z-spread 30Y yield remains above the 20Y yield. Only during the Covid-19 pandemic (March 2020), the Z-spread curve also inverts shortly at the long end.

Overall, the (scattered) Z-spread data points appear to exceed the CDS spreads, underlining the existence of a (negative) CDS-bond basis on average across the maturity domain. This is confirmed by the CDS-bond basis time-series plotted in Figure 1b. The negative basis appears to even widen further during stressed periods like the 2011-2012 sovereign debt crisis, the 2015–2016 manufacturing recession, and the Covid-19 pandemic peak in March 2020.

It is also clear from Figure 1b that the basis has many missing values, making it challenging to analyze its dynamics across maturities. The challenge stems particularly from the lack of Z-spread data. Despite comparable numbers of observations for CDS

[5]For example, the U.S. CDS curve inverted during the start of the credit crunch (around August 2011) due to a downgrade in its credit rating (Financial Times, 2011; Reuters, 2011). The CDS curves of numerous U.S. banks subsequently followed this trend, such as Bank of America, Goldman Sachs and Morgan Stanley.

Table 1: Non-parametric interpolation intervals for constant maturity Z-spreads

	1Y	**3Y**	**5Y**	**7Y**	**9Y**	**10Y**	**15Y**	**20Y**	**25Y**
Interval	[6M, 3Y)	[1Y, 5Y)	[3Y, 7Y)	[5Y, 10Y)	[5Y, 10Y)	[7Y, 15Y)	[10Y, 20Y)	[15Y, 30Y)	[15Y, 30Y)

Notes: This table reports the interpolation ranges used for computing a selected set of constant maturity Z-spreads. Z-spreads tenors are obtained by linearly interpolating between the two closest Z-spreads surrounding the target maturity (in bold) within an interval whose boundaries are the closest Markit^{TM}'s CDS tenors: 6M, 1Y, 3Y, 5Y, 7Y, 10Y, 20Y and 30Y. For example, the 1Y Z-spread is computed by interpolating samples with maturities within the range [6M, 3Y).

Table 2: Available data points

	In-Sample ($T = 2609$)		**Out-of-Sample** ($T = 261$)		
Firm	CDS	Z	CDS	Z	Total
JPM ($p = 40$)	23,393	23,247	2,349	2,249	51,238

Notes: This table reports the total number of daily CDS spread and Z-Spread observations available over the in-sample period (January 3, 2011 to December 31, 2021) and out-of-sample period (January 1, 2021 to December 31, 2021) for JPM. The length of time-series (T) over these periods is reported in brackets. The total number of credit spread series (p) is also reported in brackets.

Table 3: CDS-bond basis descriptive statistics

Sample	**Statistic**	**1Y**	**3Y**	**5Y**	**7Y**	**9Y**	**10Y**	**15Y**	**20Y**	**25Y**
In	NA (%)	80	38	40	45	56	89	100	95	90
Out	NA (%)	69	7	37	100	100	100	100	62	100
Full	Avg. (bps)	-11.984	-17.538	-23.730	-27.614	-21.287	-24.683		-46.885	-37.556
	Std. (bps)	22.757	19.062	16.396	16.756	19.987	26.630		17.810	17.106
	$\rho(1)$	0.917	0.927	0.933	0.951	0.973	0.968		0.893	0.900
	$\rho(10)$	0.450	0.753	0.739	0.852	0.898	0.749		0.457	0.707

Notes: This table reports descriptive statistics of the empirically estimated constant maturity CDS-bond basis time-series displayed in Figure 1b.

and Z-spreads, the Z-spreads are unevenly distributed across maturities; see Table 2, and Figures A.1 and A.2 in the appendix. Most Z-spread maturities lie in the range of 1 to 10 years, with missing data problems occurring both at the short and long end of the term-structure, but particularly for long maturity senior investment grade Z-spreads. Also the number of observed Z-spreads varies significantly over time. There are fewer than 9, 6, and 3 observed Z-spreads 46%, 23% and 7% of the time, respectively. By contrast, the 9 CDS tenors hardly have any missing values.

The lack of Z-spread data directly affects the computability of the JPM basis curve. Even around the most liquid tenor of 5 years where we can compute the basis most often,

we still encounter missing values more than 37% the time; see Table 3. At the short and long end of the basis curve, more 80% of the values are even missing during certain episodes. As a result of this uneven distribution of Z-spreads across maturities, we have a hard time to construct a complete time-series of the constant maturity CDS-bond basis term-structure. Naturally, this also impedes a proper study of the drivers of the basis term-structure dynamics across maturities. Given the imputation challenge already is of such a magnitude for a liquid name like JPM, we can expect even more challenges for less liquid names with lower bond issuance.

2.3 Potential economic basis drivers

In this subsection, we briefly discuss the most commonly used determinants of the CDS-bond basis as used in the literature. Though earlier studies mostly focus on the most liquid (5Y) maturity, the underlying arguments are also applicable when studying the time-series behavior of the entire CDS-bond basis term-structure across all maturities.

2.3.1 Market liquidity

Executing a negative basis trade requires going long in the bond and in the CDS, thus incurring the associated trading costs. As highlighted in Bai and Collin-Dufresne (2018), trading costs play a role in determining the profitability of a (negative) CDS-bond basis trade. Arbitrageurs typically seek bonds with higher liquidity, i.e., with lower trading liquidity risk (Bai and Collin-Dufresne, 2018). If bond market liquidity is low or dries up, the basis is therefore expected to grow more negative (Augustin and Schnitzler, 2020; Bai and Collin-Dufresne, 2018; de Wit, 2006).

We use the Corporate Bond Market Distress Index (CMDI) developed by Boyarchenko et al. (2022) as a proxy for bond market liquidity. This index integrates diverse bond distress measures from both primary and secondary markets, such as trading volume, bid-ask spreads, and the Amihud metric; see Boyarchenko et al. (2022) for a detailed motivation. Boyarchenko et al. (2022) demonstrates that the CMDI accurately captures different uncertainty regimes in the bond market and possesses predictive power for the CDS-bond basis. This makes it a valuable explanatory factor for our study. The end-of-week values of the index are taken from the Federal Reserve Bank of New York.

2.3.2 Funding risk

Basis arbitrageurs require funding when buying a bond in the repo market (Augustin and Schnitzler, 2020; Bai and Collin-Dufresne, 2018). Higher interest rates generally raise the cost of holding a physical asset, such as a bond. Consequently, increased funding costs usually push the basis more negative, as arbitrageurs seek to be compensated for these costs. We proxy funding costs with the USD LIBOR curve, whilst acknowledging that investors obtain funding at rates that exceed the interbank offered rate (Ahmadian, 2015; de Wit, 2006). Our choice for LIBOR as a funding curve is further motivated by the fact that the Z-spreads in our data are also computed using this curve. The daily LIBOR rates are taken from Bloomberg.

Basis arbitrageurs also face funding liquidity risk, referring to difficulties in securing short-term financing (at favorable rates) for the bond-leg of the arbitrage deal. This can hinder their ability to participate in the repo market, restricting their capacity to secure arbitrage profits. During periods of low funding liquidity, funding constraints can become binding, forcing arbitrageurs to unwind their positions in basis trades. Increased funding illiquidity is expected to push the basis more negative. This happened for instance during the global financial crisis and (further) widened the basis at the time (see Augustin and Schnitzler, 2020).

Following Brunnermeier and Pedersen (2009), the TED spread, defined as the difference between the 3-month USD LIBOR and 3-month U.S. Treasury bill, is used as a proxy to capture funding risk. As argued by Augustin and Schnitzler (2020), a negative correlation between the TED spread and the basis supports the hypothesis of binding funding constraints and the liquidity spiral of Brunnermeier and Pedersen (2009), capturing a flight-to-quality effect. This effect is expected to be more pronounced during times of uncertainty. Daily TED spread values are downloaded from the Federal Reserve Bank of St. Louis.

2.3.3 Counterparty risk

Negative basis trades require hedging the credit risk in the long bond position by a long position in a CDS contract against the bond issuer. Counterparty risk, i.e., the risk that the CDS issuer fails to fulfill its commitments in case the reference entity defaults, decreases the CDS spread and thereby potentially also the CDS-bond basis. This risk is typically quantified by the correlation between the default risk of the protection (CDS)

issuer and the reference entity (bond issuer). This correlation is, however, notoriously hard to measure given that CDSs are OTC products, making it hard to impossible to identify their sellers.

We follow Bai and Collin-Dufresne (2018) who suggest to proxy for counterparty risk by the correlation between changes in CDS spreads of the reference entity and an artificial CDS index. This index is a stock market capitalization weighted average of the CDS spreads of representative CDS issuers, such as the primary dealers of the Federal Reserve Bank of New York. The index is constructed using Markit$^{\text{TM}}$ data. We only include those dealers as constituents that remained part of the primary dealers list throughout our 2011-2022 sample period. We exclude JPM from the list, as it also serves as our reference entity. Annual weights for the index are computed using each dealer's end-of-year market capitalization. The full list of index constituents and weights can be found in Appendix A.

Firm-level counterparty risk is finally computed as

$$\beta_t^{\text{CR}}(\tau) = \frac{\mathbb{C}\text{ov}[\Delta\text{CDS}_t(\tau), \Delta\text{CDS}_{\text{index},t}(\tau)]}{\mathbb{V}\text{ar}[\Delta\text{CDS}_{\text{index},t}(\tau)]}, \tag{2}$$

and is estimated over an 130-day rolling window (Bai and Collin-Dufresne, 2018). We expect $\beta_t^{\text{CR}}(\tau)$ to correlate negatively with the CDS-bond basis: a higher β_t^{CR} indicates a higher joint default probability of the reference entity and the protection seller and thus a higher counterparty risk.

2.3.4 Default premia and idiosyncratic sentiment

In contrast to the earlier limits-to-arbitrage related variables, previous research also relates basis dynamics to non-fundamental factors, such as the default premium and the economic environment. For instance, (Augustin and Schnitzler, 2020) argue that changes in bond markets' perceptions about credit risk may be reflected through a default discount, leading to larger bond credit spreads and thereby negatively affecting the basis. Similarly, Cai et al. (2020) demonstrate that a reduction in macroeconomic uncertainty pushes the basis towards zero as a result of an improved PnL of a basis trade. We therefore include these variables in our empirical analysis as further controls when explaining the CDS-bond basis term-structure. As in Augustin and Schnitzler (2020), we use the approach of Lin et al. (2011) and capture the corporate bond market's perception of credit risk as the difference between Moody's Baa and Aaa bond yield indices (as obtained from Federal Reserve Bank of St. Louis' website). Rather than using a general

market-wide measure for economic uncertainty such as the VIX (Cai et al., 2020), we take a firm-specific measure, namely JPM's intra-day realized volatility σ_t^{RV}, to proxy for its firm-specific economic uncertainty. The firm-level realized volatility was computed based on 5-min returns that are calculated utilizing quote data from NYSE TAQ database.

3 Dynamic basis curve models

3.1 A multi-curve state-space representation for spreads

We consider observed CDS spreads $y_{i,t}^{\text{CDS}}(\tau_{i,t}^{\text{CDS}})$ and bond Z-spreads $y_{j,t}^{\text{Z}}(\tau_{j,t}^{\text{Z}})$ at time t for $t = 1, \ldots, T$. The number of CDS contracts and corporate bonds available in the market, varies over time. As a result, the number of spreads also varies as $i = 1, \ldots, m_t$ for CDS spreads, and $j = 1, \ldots, n_t$ for the Z-spreads. Each spread corresponds to a particular maturity $\tau_{i,t}^{\text{CDS}}$ and $\tau_{j,t}^{\text{Z}}$ for CDS and Z-spreads, respectively.

We gather the CDS spreads and Z-spreads into a vector $y_t(\tau_t)$ and model it using a multi-curve state-space set-up as

$$y_t(\tau_t) = F_t(\tau_t) f_t + \varepsilon_t(\tau_t), \qquad f_{t+1} = c + \Phi f_t + \eta_t, \tag{3}$$

$$y_t(\tau_t)' = \left(y_{1,t}^{\text{CDS}}(\tau_{1,t}^{\text{CDS}}), \ldots, y_{m_t,t}^{\text{CDS}}(\tau_{m_t,t}^{\text{CDS}}) \ , \ y_{1,t}^{\text{Z}}(\tau_{1,t}^{\text{Z}}), \ldots, y_{n_t,t}^{\text{Z}}(\tau_{n_t,t}^{\text{Z}})\right)',$$

$$\begin{pmatrix} \varepsilon_t(\tau_t) \\ \eta_t \end{pmatrix} \overset{\text{iid}}{\sim} \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} , \begin{pmatrix} \Sigma_{\varepsilon,t}(\tau_t) & 0 \\ 0 & \Sigma_\eta \end{pmatrix} \right).$$

The total dimension $p_t = m_t + n_t$ of $y_t(\tau_t)$ and τ_t varies over time as bonds and/or CDSs enter or leave the sample. We also note that the elements of τ_t change over time: when moving from t to $t+1$, the maturities of the underlying CDS and bond instruments may either decrease deterministically due to the lapse of time, or increase according to rollover schedules. The model thus describes spreads with time-varying rather than constant maturities.

The vector of spreads $y_t(\tau_t)$ in Eq. (3) depends on a vector of unobserved stochastic factors $f_t \in \mathbb{R}^k$ via a design matrix $F_t(\tau_t) \in \mathbb{R}^{p_t \times k}$. One can think of f_t as for example risk factors like the level, slope and curvature of the term-structure in both the CDS and the bond market. The design matrix $F_t(\tau_t)$ depends in a fixed, deterministic way on τ_t and is, in that sense, pre-determined given that τ_t is known at the start of period t. We elaborate on the possible functional form of $F_t(\tau_t)$ later on, but as an overly simplistic

example, $F_t(\tau_t)$ might be just a column of ones, yielding a model with a single common level factor f_t for both CDS and bond markets. The measurement noise $\varepsilon_t(\tau_t) \in \mathbb{R}^{p_t}$ in Eq. (3) is characterized by a time-varying covariance matrix $\Sigma_{\varepsilon,t}(\tau_t) \in \mathbb{R}^{p_t \times p_t}$, whose possible functional form will also be described later.

The latent factor f_t has an intercept $c \in \mathbb{R}^k$ and autoregressive dynamics characterized by the matrix $\Phi \in \mathbb{R}^{k \times k}$. We assume that f_t is stationary, such that we can initialize the state-space model in a standard stationary way by assuming $f_1 \sim \mathrm{N}(\mu_f, \Sigma_f)$, where $\mu_f = (\mathrm{I}_k - \Phi)^{-1} c$ is the unconditional mean of f_t, and where Σ_f is the unconditional covariance matrix of f_t. The latter can be obtained from the relation $\Sigma_f = \Phi \, \Sigma_f \, \Phi' + \Sigma_\eta$, where $\Sigma_\eta \in \mathbb{R}^{k \times k}$ is the time-invariant, diagonal covariance matrix of the state innovations $\eta_t \in \mathbb{R}^k$.

The model in Eq. (3) simultaneously describes the CDS and Z-spread curves. It is parameter-driven in the classification of (Cox, 1981) and falls in the class of linear Gaussian state-space models; see Durbin and Koopman (2012). Its static parameters such as c, Φ, Σ_η, etc., can be estimated using standard maximum likelihood methods and the Kalman filter (see Kalman, 1960; Durbin and Koopman, 2012). The model is particularly suited for the data at hand. It allows for a time-varying number of observed CDS contracts and bonds via its time-varying dimension p_t. It can also easily cope with missing values given its state-space set-up, and can accommodate the curves of two (or more) different markets in one homogeneous framework. To complete the model, we need to specify the design matrix $F_t(\tau_t)$ and the covariance matrix $\Sigma_{\varepsilon,t}(\tau_t)$. This is done in the next sections.

3.2 Decomposition and interpolation for multiple curves

We first introduce our baseline decomposition of the CDS and bond market credit spread curves into a common and a basis-related level component. Next, we make the specification more realistic by including more risk factors in f_t and allowing for various curve interpolation techniques as used in the literature.

There are various ways to decompose $y_t(\tau_t)$ into the latent factors f_t via the design matrix $F_t(\tau_t)$. We use the following baseline specification, which we extend later on with

further relevant factors:

$$\underbrace{\begin{bmatrix} \mu_{1,t}^{\text{CDS}}(\tau_{1,t}) \\ \vdots \\ \mu_{m,t}^{\text{CDS}}(\tau_{m_t,t}) \\ \mu_{1,t}^{\text{Z}}(\tau_{1,t}) \\ \vdots \\ \mu_{n,t}^{\text{Z}}(\tau_{n_t,t}) \end{bmatrix}}_{\mu_t \ (p_t \times 1)} = \underbrace{\begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{bmatrix}}_{F_t = F \ (p_t \times 2)} \underbrace{\begin{bmatrix} f_{c,t} \\ \frac{1}{2} f_{b,t} \end{bmatrix}}_{f_t \ (2 \times 1)}. \tag{4}$$

The expected CDS spread level now equals $\mu_{i,t}^{\text{CDS}}(\tau) = f_{c,t} + \frac{1}{2} f_{b,t}$ and the expected Z-spread level $\mu_{j,t}^{\text{Z}}(\tau) = f_{c,t} - \frac{1}{2} f_{b,t}$, for any τ. In this simple baseline specification, the credit spread curves in the CDS and bond markets are thus decomposed into a common component $f_{c,t}$ and a basis component $f_{b,t}$. Note that the basis component is easily retrieved by considering the difference between the expected CDS and Z-spread levels,

$$\mu_t^b(\tau) = \mu_t^{\text{CDS}}(\tau) - \mu_t^{\text{Z}}(\tau) = \left(f_{c,t} + \tfrac{1}{2} f_{b,t}\right) - \left(f_{c,t} - \tfrac{1}{2} f_{b,t}\right) = f_{b,t}, \tag{5}$$

for any maturity τ.

The above so-called effect-coded dummy specification has been well studied in the literature; see for instance Harvey (1990) and Durbin and Koopman (2012) in the context of time-varying seasonality modeling. Its advantage in the current context is that both $f_{c,t}$ and the basis component $f_{b,t}$ are now common stochastic drivers of both curves. From an imputation point of view, this is an advantage: the entire cross-section now contributes to the estimation of the basis as both curves across all maturities are jointly modeled. Even if (many) observations are missing in one of the two markets in a specific cross-section, the model can still filter the signals about common component $f_{c,t}$ and the basis $f_{b,t}$ from the (possibly few) non-missing data points.

The specification of F_t in Eq. (4) is of course too simplistic for fitting the entire term-structure in two markets simultaneously. To improve the specification, additional factors are needed, either in the form of levels of the curves at different maturity points, or alternatively in the form of factors like slope and curvature of the term-structure. Either of these approaches calls for an interpolation mechanism to obtain the predicted curve levels at the observed, time-varying maturities $\tau_{i,t}^{\text{CDS}}$ and $\tau_{j,t}^{\text{Z}}$ from a limited number of risk factors f_t. Several of these mechanisms have been proposed in the literature, including nearest neighbor interpolation, bucketing, linear and spline interpolation, polynomial regressions, and more; see Hagan and West (2006) for a range of techniques widely used

by practitioners. The remainder of this section outlines how to extend the specification in Eq. (4) to include various popular curve interpolation mechanisms, and how to embed these in the state-space model in Eq. (3) via a suitable specification of the design matrix $F_t(\tau_t)$ in order to fit multiple curves simultaneously through time.

3.2.1 Nearest-neighbor and bucketing

The Nearest Neighbor (NN) and bucketing algorithm for interpolating term-structures are two of the simplest and most widely used approaches in practice.[6] Both algorithms specify a term-structure curve at a predefined number of K knots. Between these knots, the curve is interpolated by taking either the value at the closest knot (NN) or the (constant) value between two adjacent knots (bucketing). Both approaches can be seen as a form of piecewise constant interpolation (0th-order spline).

To cast the approaches into state-space format, we first define the maturity knots τ_k for $k = 1, \ldots, K$. The curve values $f_{\tau_k,t}$ at time t and at each knot τ_k are treated as latent factors in Eq. (3). For the NN method, this can be expressed as an expected curve level at maturity τ of

$$h_t^{\text{NN}}(\tau) = \sum_{k=1}^{K} f_{\tau_k,t} \; \mathbb{1}_{[k=k^\star(\tau)]}, \tag{6}$$

where $\mathbb{1}_{[.]}$ denotes the indicator function, and $k^\star(\tau) = \arg\min_{k=1,\ldots,K} |\tau_k - \tau|$ gives the index of the closest knot. Similarly, the bucketing approach can be written as

$$h_t^{\text{PB}}(\tau) = \sum_{k=1}^{K} f_{\tau_k,t} \; \mathbb{1}_{[\tau\in[\tau_k,\tau_{k+1})]}. \tag{7}$$

The indicator functions in Eqs. (6) and (7) act as dummy variables and are known for all CDS and bond spreads for all maturities at all times. They can therefore be directly embedded in the design matrix $F_t(\tau_t)$ in Eq. (3).

Given our baseline specification already contains a separate common ($f_{c,t}$) and basis ($f_{b,t}$) level component, we face an identification issue if the factors $f_{\tau_k,t}$ in Eq. (6) and (7) are left unrestriced. To resolve the problem, we impose the restriction that the $f_{\tau_k,t}$ sum to zero across the different knots. This is easily done by setting $f_{\tau_{\tilde{k}},t} = -\sum_{k\neq\tilde{k}} f_{\tau_k,t}$ for one specific reference maturity $\tau_{\tilde{k}}$ such as the 5Y tenor, and accounting for this in the specification of $F_t(\tau_t)$. We give an illustration in Figure 2 for a cross-section of 6 spreads

[6] See Beretta and Santaniello (2016) for a critical assessment of the Nearest Neighbor algorithm in the context of imputation.

with maturities ranging between 3Y to 8Y, and choosing the (most liquid) 5Y tenor as the reference point $\tilde{k}$.

3.2.2 Linear and spline interpolation

While simple, the disadvantage of the above two piecewise constant approaches is that they typically require many factors f_t to fit the dynamics of the two term-structures in y_t sufficiently well. A straightforward improvement is obtained by replacing constant (0th order) by linear (1st order) or by higher order (2nd or 3rd) spline interpolation. Each of these again specifies a series of levels $f_{\tau_k,t}$ at fixed knot positions τ_k for $k = 1, \ldots, K$. For linear interpolation, the knots are interpolated as

$$h^{\text{PL}}(\tau) = \frac{\tau_{k+1} - \tau}{\tau_{k+1} - \tau_k} f_{\tau_k,t} + \frac{\tau - \tau_k}{\tau_{k+1} - \tau_k} f_{\tau_{k+1},t}, \qquad \text{for} \quad \tau_k \leq \tau < \tau_{k+1}. \tag{8}$$

As in Eqs. (6) and (7), the linear interpolation scheme can directly be cast into a design matrix $F_t(\tau_t)$ that is known for all time points t. For identification, we again impose the restriction that the knots should sum to zero, i.e., $f_{\tau_{\tilde{k}},t} = -\sum_{k \neq \tilde{k}} f_{\tau_k,t}$ for some reference knot position $\tau_{\tilde{k}}$ such as the 5Y tenor. See Figure 2 for an example. For higher order splines the interpolation in Eq. (8) is somewhat more involved, but still straightforward algorithmically; see for instance Poirier (1973).

3.2.3 Nelson-Siegel interpolation

A last type of interpolation mechanism that we include in our study is the well-known empirical term-structure model of Diebold and Li (2006), also known as the dynamic Nelson-Siegel model. The Nelson-Siegel model is a parsimonious yield curve model that can produce a variety of shapes observed in financial term-structure data. It has many extensions, including arbirage-free versions (Christensen et al., 2011); see De Pooter (2007) for a non-exhaustive list of single curve specifications. A multi-curve extension was introduced by Diebold et al. (2008) for extracting common and idiosyncratic curve dynamics in an international yield curve modeling context.

The core interpolation mechanism of the Nelson-Siegel model is driven by two components, namely by a slope and curvature factor, each with its unique loadings. The specification is given by

$$h^{\text{NS}}(\tau) = \frac{1 - e^{-\lambda\tau}}{\lambda\tau} f_{1,t} + \left(\frac{1 - e^{-\lambda\tau_t}}{\lambda\tau} - e^{-\lambda\tau} \right) f_{2,t}, \tag{9}$$

where λ is the so-called decay parameter. The time-varying factors $f_{1,t}$ and $f_{2,t}$ govern

the slope respectively the curvature of the spread curves. Their interpretation follows directly from the factor loadings. For $f_{1,t}$, the loading decreases monotonically from 1 to 0 as the maturity τ increases. For $f_{2,t}$, by contrast, the loading starts at 0, then increases and eventually converges to 0 for large τ. As a result, the level factors $f_{c,t}$ and $f_{b,t}$ in the Nelson-Siegel model specify the long-term expect level of the CDS/Z-spread and CDS-bond basis curve, respectively. As a result, the Nelson-Siegel approach does not require a similar identification restriction on $f_{1,t}$ and $f_{2,t}$ as the previous interpolation methods. The specification in Eq. (9) can easily be incorporated in $F_t(\tau_t)$. This is illustrated in Figure 2.

The parametric interpolation with the Nelson-Siegel model has two advantages over the earlier non-parametric interpolation methods based on splines. First, the number of required risk factors in f_t is typically much smaller than for the spline methods. Second, the spline methods with many knots have a tendency to overfit the data, particularly at times when one of the markets only has few data points. This may result in counterintuitive patterns of the credit spread curves across maturities. We have found the Nelson-Siegel model to be much less susceptible to such effects. Of course, the assumption of a specific parametric form comes at the cost of model mis-specification risk. Given the flexible form of the Nelson-Siegel model and the substantial track-record of the model as an adequate empirical description of term-structure shapes, we think the model is a good choice in the current context.

Figure 2: Maturity-dependent designs for $F_t(\tau_t)$

(a) Nearest Neighbor (NN)

$$F_t(\tau_t) = \left[\begin{array}{c|c} F^{\text{Base}} & \begin{array}{ccc} \overbrace{1}^{f_{3\text{Y},t}} & & \overbrace{}^{f_{7\text{Y},t}} \\ -1 & -1 & -1 \\ & 1 & \\ 1 & & \\ -1 & -1 & -1 \\ & 1 & \end{array} \end{array}\right] \underbrace{\qquad\qquad}_{=: F_t^{\text{NN}}(\tau_t)} \overbrace{}^{f_{10\text{Y},t}}$$

(b) Piecewise Buckets (PB)

$$\left[\begin{array}{c|c} F^{\text{Base}} & \begin{array}{ccc} \overbrace{1}^{f_{[2\text{Y},4\text{Y}),t}} & \overbrace{}^{f_{[6\text{Y},8\text{Y}),t}} & \overbrace{}^{f_{[8\text{Y},10\text{Y}),t}} \\ -1 & -1 & -1 \\ & 1 & \\ 1 & & \\ -1 & -1 & -1 \\ & & 1 \end{array} \end{array}\right] \begin{array}{l} \} \, y_{1,t}^{\text{CDS}}(3\text{Y+8D}) \\ \} \, y_{2,t}^{\text{CDS}}(5\text{Y+10D}) \\ \} \, y_{3,t}^{\text{CDS}}(7\text{Y+15D}) \\ \} \, y_{1,t}^{\text{Z}}(2\text{Y+11M}) \\ \} \, y_{2,t}^{\text{Z}}(5\text{Y+7M}) \\ \} \, y_{3,t}^{\text{Z}}(8\text{Y+9D}) \end{array} \underbrace{\qquad\qquad}_{=: F_t^{\text{PB}}(\tau_t)}$$

(c) Piecewise Linear (PL)

$$F_t(\tau_t) = \left[\begin{array}{c|c} F^{\text{Base}} & \begin{array}{ccc} \overbrace{\frac{88}{90}}^{f_{3\text{Y},t}} & -\frac{1}{90} & -\frac{1}{90} \\ -\frac{71}{72} & -\frac{70}{72} & -\frac{71}{72} \\ & \frac{71}{72} & \frac{1}{72} \\ \frac{98}{60} & \frac{19}{60} & \frac{19}{60} \\ -\frac{53}{60} & -\frac{46}{60} & -\frac{53}{60} \\ & \frac{79}{120} & \frac{41}{120} \end{array} \end{array}\right] \underbrace{\qquad\qquad}_{=: F_t^{\text{PL}}(\tau_t)} \overbrace{}^{f_{7\text{Y},t}} \overbrace{}^{f_{10\text{Y},t}}$$

(d) Nelson-Siegel (NS)

$$\left[\begin{array}{c|c} F^{\text{Base}} & \begin{array}{cc} \overbrace{\frac{1-e^{-\lambda\times 3\text{Y8D}}}{\lambda\times 3\text{Y8D}}}^{f_{1,t}\text{ (slope)}} & \overbrace{\frac{1-e^{-\lambda\times 3\text{Y8D}}}{\lambda\times 3\text{Y8D}} - e^{\lambda\times 3\text{Y8D}}}^{f_{2,t}\text{ (curvature)}} \\ \frac{1-e^{-\lambda\times 5\text{Y10D}}}{\lambda\times 5\text{Y10D}} & \frac{1-e^{-\lambda\times 5\text{Y10D}}}{\lambda\times 5\text{Y10D}} - e^{\lambda\times 5\text{Y10D}} \\ \frac{1-e^{-\lambda\times 7\text{Y15D}}}{\lambda\times 7\text{Y15D}} & \frac{1-e^{-\lambda\times 7\text{Y15D}}}{\lambda\times 7\text{Y15D}} - e^{\lambda\times 7\text{Y15D}} \\ \frac{1-e^{-\lambda\times 2\text{Y11M}}}{\lambda\times 2\text{Y11M}} & \frac{1-e^{-\lambda\times 2\text{Y11M}}}{\lambda\times 2\text{Y11M}} - e^{\lambda\times 2\text{Y11M}} \\ \frac{1-e^{-\lambda\times 5\text{Y7M}}}{\lambda\times 5\text{Y7M}} & \frac{1-e^{-\lambda\times 5\text{Y7M}}}{\lambda\times 5\text{Y7M}} - e^{\lambda\times 5\text{Y7M}} \\ \frac{1-e^{-\lambda\times 8\text{Y9D}}}{\lambda\times 8\text{Y9D}} & \frac{1-e^{-\lambda\times 8\text{Y9D}}}{\lambda\times 8\text{Y9D}} - e^{-\lambda\times 8\text{Y9D}} \end{array} \end{array}\right] \begin{array}{l} \} \, y_{1,t}^{\text{CDS}}(3\text{Y+8D}) \\ \} \, y_{2,t}^{\text{CDS}}(5\text{Y+10D}) \\ \} \, y_{3,t}^{\text{CDS}}(7\text{Y+15D}) \\ \} \, y_{1,t}^{\text{Z}}(2\text{Y+11M}) \\ \} \, y_{2,t}^{\text{Z}}(5\text{Y+7M}) \\ \} \, y_{3,t}^{\text{Z}}(8\text{Y+9D}) \end{array} \underbrace{\qquad\qquad}_{=: F_t^{\text{NS}}(\tau_t)}$$

Notes: An illustration of maturity-dependent loading matrices for various interpolation methods. Each subfigure displays a matrix $F_t(\tau_t)$ consisting of two sub-matrices, with the first one being a time-invariant a fixed component, F^{Base}, that specifies the common level and basis level effect of Eq. (4). The second matrix in each plot is maturity-dependent whose loadings are implied from either the Nearest Neighbor, piecewise buckets, piecewise linear and Nelson-Siegel. The loadings in all designs are constructed for the exact same credit-and bond spread time-series (6 in total), with a 30/360 day count convention. All factors are treated as common effects for CDS and Z-spreads. Where possible, effect-coding is always applied on factors modeling the 5Y tenor.

3.3 Model extensions

3.3.1 Idiosyncratic curve effects

So far, the models from Section 3.2 only account for common joint tenor effects via the additional factors on top of $f_{c,t}$ and $f_{b,t}$. Such common tenor effects for both curves can be restrictive in practice. A straightforward solution is to model the basis term-structure $b_t(\tau)$ by additional curve-specific factors in the design matrices $F_t(\tau_t)$. For example, the *common* piecewise linear or first-order spline can be extended with a separate idiosyncratic (ID) effect-coded first-order spline for the *basis* term-structure:

$$h^{\text{PL-ID}}(\tau) = \begin{cases} \frac{\tau_{k+1}-\tau}{\tau_{k+1}-\tau_k} f^c_{\tau_k,t} + \frac{\tau-\tau_k}{\tau_{k+1}-\tau_k} f^c_{\tau_{k+1},t} + \frac{\tau_{k+1}-\tau}{\tau_{k+1}-\tau_k} f^b_{\tau_k,t} + \frac{\tau-\tau_k}{\tau_{k+1}-\tau_k} f^b_{\tau_{k+1},t} & \text{for} \quad y_t^{\text{CDS}}, \\ \frac{\tau_{k+1}-\tau}{\tau_{k+1}-\tau_k} f^c_{\tau_k,t} + \frac{\tau-\tau_k}{\tau_{k+1}-\tau_k} f^c_{\tau_{k+1},t} - \frac{\tau_{k+1}-\tau}{\tau_{k+1}-\tau_k} f^b_{\tau_k,t} - \frac{\tau-\tau_k}{\tau_{k+1}-\tau_k} f^b_{\tau_{k+1},t} & \text{for} \quad y_t^{\text{Z}}, \end{cases} \tag{10}$$

for $\tau_k < \tau < \tau_{k+1}$. Here, the $f^c_{\tau_k,t}$s span a spline that is common between the two credit spread curves, while the $f^b_{\tau_k,k}$s span an effect-coded idiosyncratic spline for the basis curve. Again, the extension is easily incorporated into the design matrix $F_t(\tau_t)$. Similar extensions of course hold for all the other interpolation techniques from Section 3.2. We can even choose for different numbers of factors $f^c_{\tau_k,t}$ versus $f^b_{\tau_k,t}$. For instance, in the Nelson-Siegel specification we can opt for a common level and slope component, augmented with only a basis level component, both characterized by different parameters λ^c and λ^b. This may be helpful if otherwise the model is at risk of overfitting the data, particularly during episodes where many data points are missing in one of the two markets.

3.3.2 Including volatility term-structures

The functional form of the observation variance matrix $\Sigma_{\varepsilon,t}(\tau_t)$ in (3) is thus far left unspecified. In its simplest form, $\Sigma_{\varepsilon,t}(\tau_t)$ can be a time-invariant diagonal matrix holding the constant conditional variances of each of the spread observations. The elements of y_t, however, relate to observations with a time-varying maturity. It is therefore unlikely that they will be empirically subject to the same level of noise as time progresses. Moreover, if a new element is added to the observation vector y_t, it is unclear what variance should be chosen for it.

In line with the motivation and results in Longstaff and Schwartz (1992), Bianchi et al. (2009), Viceira (2012), Fonseca and Gottschalk (2013), and Koopman et al. (2010), we therefore also consider a model specification where the CDS spread and Z-spread volatilities are directly a function of maturity,

$$\begin{aligned}\Sigma_{\varepsilon,t}(\tau_t) &= \text{diag}\left(\sigma_t^{2,\text{CDS}}(\tau_t)', \sigma_t^{2,\text{Z}}(\tau_t)'\right) \\ &= \text{diag}\left(\sigma_{1,t}^{2,\text{CDS}}(\tau_{1,t}), \ldots, \sigma_{m,t}^{2,\text{CDS}}(\tau_{m,t}) \quad , \quad \sigma_{1,t}^{2,\text{Z}}(\tau_{1,t}), \ldots, \sigma_{n,t}^{2,\text{Z}}(\tau_{n,t})\right),\end{aligned}$$

where we adopt a so-called local-volatility model of Dupire (1994) for fitting variance term-structures. A local-volatility term-structure is a non-parametric model for spanning a curve along a set of pre-specified variance knots, just like for the spread curve interpolation methods of Section 3.2. We use a piecewise linear interpolation, though other interpolation techniques can of course also be used:

$$\sigma_t^2(\tau) = \frac{\tau_{k+1} - \tau}{\tau_{k+1} - \tau_k}\sigma_{\tau_k}^2 + \frac{\tau - \tau_k}{\tau_{k+1} - \tau_k}\sigma_{\tau_{k+1}}^2, \qquad \text{for} \quad \tau_k < \tau < \tau_{k+1}, \tag{11}$$

for a number of volatility levels $\sigma_{\tau_k}^2$ over a grid of knots τ_k, $k = 1, \ldots, K$. As the maturities of the observations in y_t change over time, the covariance matrix $\Sigma_t(\tau_t)$ becomes time-varying even if the values of $\sigma_{\tau_k}^2$ at the knots remain constant over time. Also, a new element of y_t with a new, specific value of τ can easily be accommodated by the model using the specification of $F_t(\tau_t)$ as in Section 3.2 and the volatility term-structure as in Eq. (11).

Empirically, we typically only need a few variance knots $\sigma_{\tau_k}^2$ to adequately model the local-volatility term-structure. This only results in a few extra parameters to be estimated. Therefore, the use of variance curves is a parsimonious solution to limit the number of unknown parameters, especially in a high-dimensional set-up like ours.

3.4 Filtering and maximum likelihood estimation

The formulation of the time-varying basis curve model via Eq. (3) falls in the class of linear Gaussian state-space models. This makes it straightforward to estimate the model's static parameters, despite the model's flexibility and complexity. The static parameter vector ψ contains the free elements of c, Φ, Σ_η and $\Sigma_{\varepsilon,t}(\tau_t)$, and also the elements of $F_t(\tau_t)$, if any, that need to be estimated, such as the λ parameter of the

Nelson-Siegel specification. This is done as follows. Given ψ, the Kalman filter can be applied to recursively estimate the latent state vector f_t by evaluating its conditional mean and variance $f_{t|t-1} = \mathbb{E}_{t|t-1}[f_t]$ and $P_{t|t-1} = \mathbb{V}\text{ar}_{t|t-1}[f_t]$ at each point in time. The Kalman filter then uses these predictions and the prediction error decomposition to produce the log-likelihood value. This log-likelihood can subsequently be maximized numerically to produce the maximum likelihood estimator $\hat{\psi}^{\text{ML}}$ and associated standard errors; see Durbin and Koopman (2012) for a detailed discussion.

We obtain the predicted CDS-bond basis by evaluating the Kalman filter at the optimal $\hat{\psi}^{\text{ML}}$ as

$$\mathbb{E}_{t|t-1}[b_t(\tau)] = \mathbb{E}_{t|t-1}\left[y_t^{\text{CDS}}(\tau) - y_t^{\text{Z}}(\tau)\right] = \left(\begin{array}{c:c} \text{I} & -\text{I} \end{array}\right) F_t(\tau_t) f_{t|t-1}.$$

Similarly, the Kalman filter produces a conditional variance estimate for the basis, namely

$$\mathbb{V}\text{ar}_{t|t-1}[b_t(\tau)] = \left(\begin{array}{c:c} \text{I} & -\text{I} \end{array}\right) F_t(\tau_t) P_{t|t-1} F_t(\tau_t)' \left(\begin{array}{c} \text{I} \\ \hdashline -\text{I} \end{array}\right) + \Sigma_{\varepsilon,t}(\tau_t).$$

The predicted value of the basis and its one-period-ahead variance can be used to compute the out-of-sample risk of the basis in terms of for instance Value-at-Risk (VaR) or Expected Loss (EL).

The Kalman filter can also be used to obtain filtered and smoothed estimates of the basis and its variance based on the information $\{y_1, \ldots, y_t\}$ and $\{y_1, \ldots, y_T\}$, respectively. These can be useful for an expost analysis of the basis given all available data.

Whichever of these estimates of the basis is used (predicted, filtered, smoothed), the advantage of the Kalman filter and the current state-space set-up is that they can easily deal with missing data. At every time point, the filter uses all the available data and the structure embedded in $F_t(\tau_t)$ to obtain the best possible estimate of the basis. We exploit this in the next section to obtain estimates of the entire basis term-structure at the individual firm level, which we can then subsequently analyze to determine its economic drivers.

4 Empirical results

In this section, we apply the new imputation methodology for the basis to recover the firm-level basis term structure history for JP Morgan Chase & Co (JPM) and analyze its empirical economic determinants. We do so in two steps. First, in Section 4.2, we estimate the state-space model from Eq. (3) and assess the performance of several of the interpolation mechanisms for fitting CDS and Z-spread data to uncover JPM's CDS-bond basis term-structure. We then proceed in Section 4.3 by relating the constructed basis term structure to its potential economic determinants.

4.1 Empirical modeling choices

We consider an in-sample period of Jan 3, 2011 to Dec 31, 2020. To check whether the robustness of the results also holds up out-of-sample, we also evaluate the fit of the estimated model over the out-of-sample period Jan 1, 2021 to Dec 31, 2021.

We consider a total of 9 curve interpolation models. The simplest `Base` configuration of F_t has two common level components $f_{c,t}$ and $f_{b,t}$. The remaining models augment these two baseline factors with additional factors characterizing either the knots or parametric curve loadings for the interpolation mechanisms discussed in Section 3.

For the nearest neighbor (`NN`), bucketing (`PB`), and piecewise linear (`PL`) models, we include 8 free knots: $f_{6\text{M}}$, $f_{1\text{Y}}$, $f_{3\text{Y}}$, $f_{7\text{Y}}$, $f_{10\text{Y}}$, $f_{15\text{Y}}$, $f_{20\text{Y}}$ and $f_{30\text{Y}}$. These models use a common spline for the CDS spread and Z-spread data. We assign the most liquid maturity knot ($f_{5\text{Y}}$) to be effect-coded.[7] The `PL-ID` model includes the same common curve components as the `PL` model, but extends it with tenor-wise effect-coded knots for the basis at 1Y, 4Y, 8Y and 20Y. In this way, it allows for a non-flat basis term-structure. The choice of knots for the basis is in-between the common spline knots to reduce the risk of multicollinearity and model instability when data is scarce. We only allow for a limited set of four basis tenor factors given the scarcity of the Z-spread data.

As a prime example of a parametric model, we use the Nelson-Siegel (`NS`) specification with only common factors, or the Nelson-Siegel specification augmented with basis specific components (`NS-ID`); see Section 3.2.3. The parametric `NS-ID` model avoids

[7] As the bucketing model (`PB`) uses intervals rather than knots, one knot position is redundant and can be discarded. We use $f_{[5\text{Y},\ 7\text{Y})}$ interval as the effect-coded one for `PB`.

many of the missing data problem that the earlier interpolation methods have when data are scarce, as it imposes a tighter parameteric shape onto the CDS and Z-spread data.

Finally, for the two models with idiosyncratic basis curve dynamics, we also consider the variance curve specification of Section 3.3.2. We refer to these two models as `PL-ID-VC` and `NS-ID-VC`. The variance knots are positioned at 6M, 1Y, 5Y, 10Y and 30Y and form a piecewise-linear spline.

4.2 Basis estimation results

In this subsection, we first compare the different models in terms of their aggregate in-sample and out-of-sample performance before selecting the candidate models that are used to explain the economic drivers of the basis term structure in Section 2.3. Table 4 presents the total log-likelihood (LL) values, Akaike information criteria (AIC), and mean absolute errors (MAEs) of the different models, where the MAEs are split out over the CDS spreads and the Z-spread data. The results show that there is a significant improvement in log-likelihood when credit spreads are interpolated with both common and basis-specific effects: the log-likelihood increases by more than 18,000 points from `PL` to `PL-ID`, and by more than 6,500 points from `NS` to `NS-ID`. The improvement in fit is mirrored in a similar reduction in the MAE level. Overall, all models fit the CDS data better than the Z-spread data in terms of MSE.

The linear spline model with idiosyncratic curve components (`PL-ID`) performs best in terms of both in-sample and out-of-sample statistics. Moreover, we find that imposing the variance curve structure (`PL-ID-VC`) decreases the in-sample log-likelihood value and even increases the AIC, despite a reduction of 30 parameters. Out-of-sample, however, the `PL-ID-VC` model behaves slightly better than the `PL-ID` model in terms of MSE for both CDS and Z-spread observations. A similar pattern can be seen for the Nelson-Siegel models, `NS-ID` versus `NS-ID-VC`, but now only for the Z-spread component of the model.

In Table B.2, we split out the fit of the different models across the maturity spectrum. Each cell holds the one-step-ahead MAE for a specific model and maturity. Below each MAE, we report the model confidence set (MCS) p-values per maturity across models. The models that are part of the 95% model confidence set for a specific maturity are

Table 4: Aggregate model fit statistics

Sample	Statistic	Base	NN	PB	PL	NS	PL-ID	NS-ID	PL-ID-VC	NS-ID-VC
	#Factors	2	10	9	10	4	14	5	14	5
	#Params	46	70	67	70	53	82	57	52	27
	Est. Time	4.404	14.695	10.723	18.94	8.161	38.311	6.809	32.049	3.696
	LL	-213,942	-145,360	-149,996	-141,798	-153,930	-123,387	-147,209	-125,529	-150,816
	AIC	427,977	290,861	300,127	283,737	307,966	246,938	294,532	251,162	301,686
In	MAE CDS	38.975	2.544	2.956	2.576	5.107	2.478	4.483	2.457	4.247
	MAE Z	30.488	10.691	10.573	9.647	8.074	4.322	6.188	4.239	5.897
Out	MAE CDS	26.653	0.729	1.527	0.791	2.687	0.711	1.580	0.693	1.706
	MAE Z	39.825	15.721	16.082	15.942	13.407	6.823	8.519	6.525	7.162

Notes: This table reports model aggregate model (performance) statistics, including the number of factors and parameters per model. Estimation was done in `Python` and `Numba` on a laptop with 32 GB RAM and a 2.90 GHz AMD Ryzen 5 PRO 6650U processor. The estimation time is reported in minutes. The table presents the log-likelihood contributions (LL), the Akaike information criterion (AIC), and mean absolute error (MAE) of one-step ahead predictions for CDS spread and Z-spread data (in basis points). The MAEs are reported for both the in-sample (January 3, 2011 to December 31, 2020) and out-of-sample period (January 1, 2021 to December 31, 2021).

marked with a star.

We observe significant variability in MAE across maturities for models with and without idiosyncratic curve dynamics. Models without such dynamics show MAEs ranging between 5 and 33 basis points), whereas models with such dynamics have a reduced variability range between approximately 4 to 10 basis points. Particularly the short-end and long-end parts of the basis curve appear challenging for the models without idiosyncratic components. For instance, models with short-run basis dynamics have roughly half of the MAE for short maturities, 25% lower MAE for medium maturities, and about 2-3 times lower MAEs for long maturities. Particularly the 20 year segment of the curve shows a major improvement when adding idiosyncratic components. The results are consistent for the in-sample and out-of-sample period; see also Figure B.1 in Appendix B.3.

Given the improved fit of the model with idiosyncratic components for most segments of the curve, the models in the 95% model confidence set for each maturity all have such components, with a single exception for the `NS` model at the 7Y maturity. Models with idiosyncratic short-run basis dynamics thus outperform models with only a long-run basis component. This points to the existence of a non-flat CDS-bond basis term-structure.

The `PL-ID-VC` model is part of the in-sample model confidence set for every maturity,

Table 5: In-sample and out-of-sample model fit on CDS-bond basis term-structure level for JPM

Model	1Y	3Y	5Y	7Y	9Y	20Y	25Y	Cross-Sect. Avg.
Panel A: In-sample MAE								
Base	37.76	44.298	50.352	45.597	33.056	98.833	64.504	43.503
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
NN	14.306	8.971	5.966	5.661	5.872	33.769	18.807	8.406
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
PB	13.413	9.792	8.572	8.459	9.853	33.459	13.09	10.446
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]
PL	15.555	9.4	6.135	4.98	4.674	33.31	18.008	8.233
	[0.000]	[0.000]	[0.000]	[0.01]	[0.042]	[0.000]	[0.000]	[0.000]
NS	16.752	10.889	5.267	4.307⋆	4.777	29.402	16.704	8.342
	[0.000]	[0.000]	[0.000]	[0.872]	[0.03]	[0.000]	[0.000]	[0.000]
PL-ID	7.455⋆	4.624⋆	4.366	4.419⋆	4.512	8.21	8.889⋆	5.104
	[1.000]	[0.459]	[0.046]	[0.188]	[0.042]	[0.002]	[0.084]	[0.000]
NS-ID	10.126	7.573	4.606	4.289⋆	4.425⋆	14.521	14.513	6.559
	[0.001]	[0.000]	[0.003]	[0.872]	[0.103]	[0.002]	[0.000]	[0.000]
PL-ID-VC	7.556⋆	4.552⋆	4.12⋆	4.253⋆	4.11⋆	5.816⋆	7.594⋆	4.832⋆
	[0.803]	[1.000]	[0.264]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
NS-ID-VC	9.115⋆	5.71	3.969⋆	4.584⋆	5.43	10.988	12.127	5.859
	[0.088]	[0.005]	[1.000]	[0.147]	[0.01]	[0.002]	[0.002]	[0.000]
Panel B: Out-of-sample MAE								
Base	53.623	56.029	67.541			104.529		65.673
	[0.000]	[0.000]	[0.000]			[0.000]		[0.000]
NN	11.183	9.194	4.72			41.655		12.366
	[0.000]	[0.000]	[0.003]			[0.000]		[0.000]
PB	12.886	8.987	7.259			41.843		13.225
	[0.000]	[0.000]	[0.002]			[0.000]		[0.000]
PL	10.722	8.414	5.872			42.868		12.455
	[0.000]	[0.000]	[0.000]			[0.000]		[0.000]
NS	17.719	13.839	3.63			36.582		14.504
	[0.000]	[0.000]	[0.003]			[0.000]		[0.000]
PL-ID	4.497⋆	6.53⋆	2.73⋆			8.875		5.433⋆
	[0.816]	[1.000]	[1.000]			[0.000]		[0.07]
NS-ID	5.129⋆	8.616	3.032⋆			19.287		7.994
	[0.215]	[0.000]	[0.299]			[0.000]		[0.000]
PL-ID-VC	4.684⋆	7.129	2.742⋆			4.706⋆		5.152⋆
	[0.652]	[0.000]	[0.97]			[1.000]		[1.000]
NS-ID-VC	4.419⋆	8.984	4.273			9.268		7.015
	[1.000]	[0.000]	[0.039]			[0.000]		[0.000]

Notes: This table contains the MAE (in bps) for the one-step ahead in-sample and out-of-sample CDS-bond basis predictions for JPM across 7 fixed constant maturities, using the estimated models from Table 4 and the basis observations from Figure 1b. For the last column, we target the cross-sectional average of the basis to compute the MAE. Empty cells correspond to periods where we do not have basis observations due to missing data; see also Table 3 for an overview. Model confidence set p-values are reported in squared brackets beneath the MAEs. MAEs that are part of the 95% model confidence set are marked by ⋆.

and for all but one maturity out-of-sample. The PL-ID-VC is joined by the PL-ID model for maturities up to 5 years, again both in-sample and out-of-sample, but not for the longer maturities. At the middle segment, also the NS-ID(-VC) incidentally enters the model confidence set. The inclusion of variance curves appears to slightly improve the models' fit in-sample, but not necessarily out-of-sample. These finding are also supported by the performance metrics and model confidence sets on the raw data CDS and Z-spread in Appendix B.2.

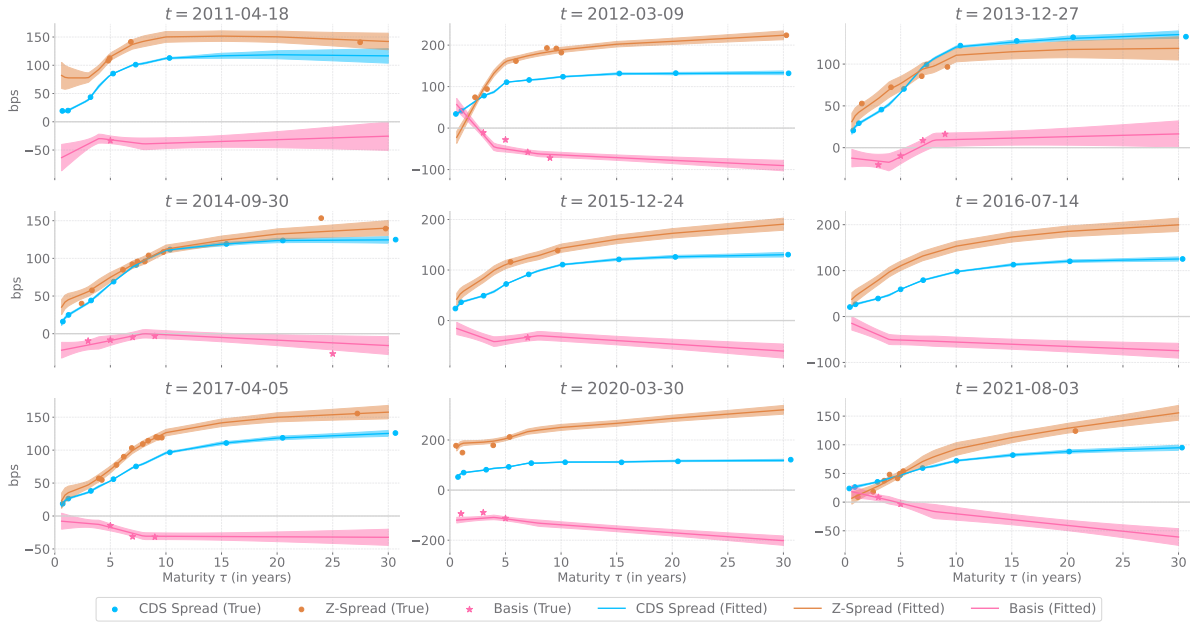
Summarizing, the PL-ID(-VC) model is the most promising model for constructing the basis curve histories at the firm level based on the limited CDS and Z-spread data available. This, therefore, will be our main model in the remainder of the analysis. As a possible second choice, the NS-ID(-VC) model can be considered. We consider this model in one of the robustness analyses later on.

Figure 3 shows the constructed JPM basis term-structure curves for a selection of nine different dates. The PL-ID-VC model indeed appears to provide an accurate description of the basis across a range of different historical basis term-structure shapes, including flat, upward-sloping or downward-sloping basis term-structures: the fitted curves are close to the observed data points. The model also performs consistently when CDS (blue dots) or Z-spread data (brown dots) are partially or completely missing.

The full time-series of the JPM PL-ID-VC model-implied basis at different maturities as well as the overall time-series average of the curve are presented in Figure 4. The basis curves display quite some persistence over time with, on average, a negative basis level and downward sloping pattern across maturities. The basis term-structure is often non-monotonic and hardly seems to stick to zero for a prolonged period of time. Its dynamics appear to reflect some of the major economic developments. For instance, major troughs in the basis appear after market-wide shocks, such as the credit crunch (2011-2012), the manufacturing crisis (2015-2016), and the Covid-19 pandemic (2020-2021). We study the relation between the basis term-structure and possible economic and limits-to-arbitrage factors in more detail in the next section.

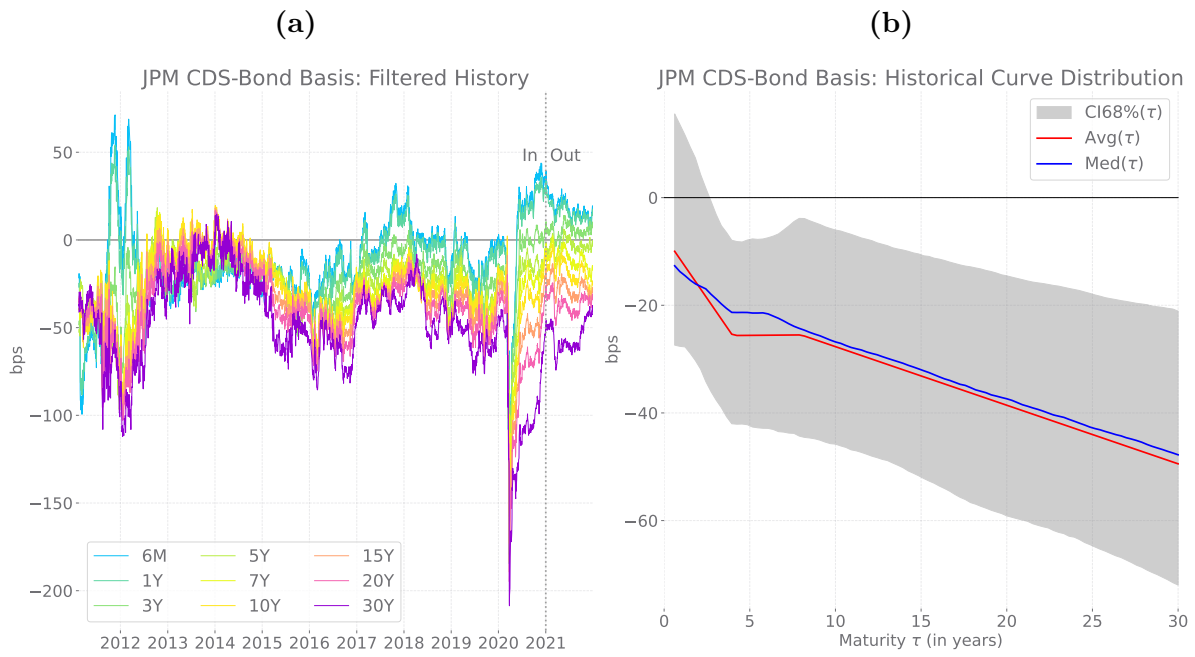
To provide some final support for the constructed basis term-structures given the scarcity of firm-level Z-spread data, we try to create an external benchmark using index data. As American iBoxx Z-spread investment grade indices for the financial sector do not exist, we use their European counterparts instead. We construct our benchmark

Figure 3: Fitted CDS, Z-spread, and basis term-structures for the PL-ID-VC model



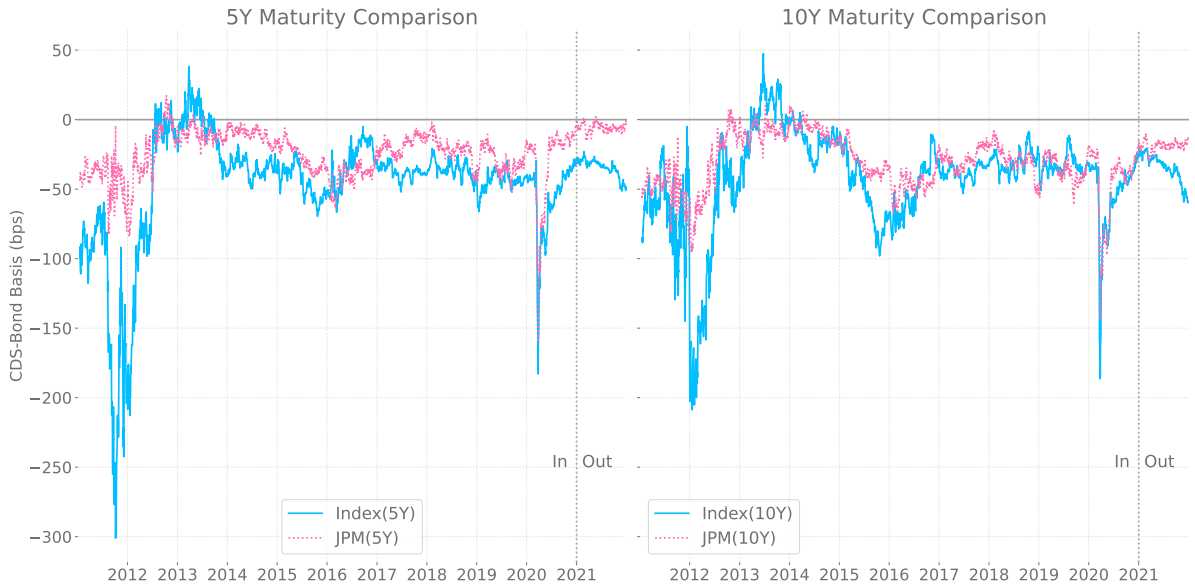
Notes: This figure displays fitted CDS, Z-spread and CDS-bond basis observations on 9 selected dates by the PL-ID-VC model. See caption of Figure B.1 for more details.

Figure 4: Filtered constant maturity CDS-bond basis term-structure of JPM by PL-ID-VC model



Notes: Figure 4a shows the constant maturity CDS-bond basis time-series (in bps) for JPM based on the PL-ID-VC model. The vertical line denotes the end of the in-sample period. Figure 4b shows the time-series average of the term-structure with its pointwise empiricical quantile bands for the historical series.

Figure 5: Index basis vs. `PL-ID-VC` model basis



Notes: This figure plots the constant maturity index CDS-bond basis time-series (blue) computed non-parametrically for various tenors. The CDS iTraxx and Z-spread iBoxx indices with tickers 'Eur Sr Finls' and 'Eur Financials A', respectively, were used to compute the index-based basis term-structures using piecewise-linear interpolation. Data are taken from Markit^{TM}. The (pink) dotted time-series correspond to the filtered basis estimates of JPM using the `PL-ID-VC` model. A vertical line denotes the end of the in-sample period.

index-based basis term-structure curve using simple piecewise-linear interpolation of the observed CDS iTraxx and iBoxx Z-spread indices for the financial sector as obtained from Markit. The results are shown in Figure 5. The figure reveals that despite the differences in our reference entity JPM compared to the European financial index constituents, the JPM firm-level term-structure time-series shows striking similarities with its index-based counterpart, both in terms of the levels at different maturities and in terms of the secular time-series dynamics. This provides further back-up that the model-based imputation at the firm-level is successful and provides empirically congruent patterns of the historical basis term-structure. We therefore proceed with analyzing the economic determinants of this term-structure in the next section.

4.3 Explaining basis term-structure dynamics

4.3.1 Setting up the regression framework

After having constructed the entire JPM's CDS-bond basis term-structure over the full sample period in Section 4.2, we now proceed by relating the basis term structure to

its potential economic and limits-to-arbitrage determinants; see Section 2.3. As our dependent variable, we use the filtered basis estimates $\widehat{\mu}^b_{t|t}(\tau)$ using the `PL-ID-VC` model from Subsection 2.3. Compared to their predicted counterparts $\widehat{\mu}^b_{t|t-1}(\tau)$, the filtered estimates use all data available up to time t.[8]

Figure 6a presents the time-series patterns of our explanatory variables. As both the basis time-series and the economic drivers are highly persistent at the daily frequency, we run all our regressions in differences to avoid any spurious results; see Figure 6b. The baseline model reads

$$\begin{aligned}\Delta B_{i,t} = \beta_{0,t} + \beta_{1,t}\Delta \mathrm{ML}_t + \beta_{2,t}\Delta \mathrm{FC}_{i,t} + \beta_{3,t}\Delta \mathrm{FL}_t + \beta_{4,t}\Delta \mathrm{CR}_{i,t} \\ + \beta_{5,t}\Delta \mathrm{DP}_t + \beta_{6,t}\Delta^5 \mathrm{IS}_t + \varepsilon_{i,t},\end{aligned} \tag{12}$$

where $\Delta B_{i,t}$ is the first difference of the CDS-bond basis at the ith maturity. Similarly, $\Delta \mathrm{ML}_t$, $\Delta \mathrm{FC}_{i,t}$, $\Delta \mathrm{FL}_t$, $\Delta \mathrm{CR}_{i,t}$, and $\Delta \mathrm{DP}_t$, represent the first differences in the explanatory variables: market liquidity,[9] funding costs, funding liquidity, counterparty risk, default premium. Only for idiosyncratic risk, a 5-day difference was taken: $\Delta^5 \mathrm{IS}_t$. We found that first differences of intra-day realized volatility are too noisy and do not correlate well with bases, but a smoother variant of the volatility, like the 5-day difference, does correlate well. To simplify some of the exposition, we also define the vector $B_t \in \mathbb{R}^9$ (without the subscript i), holding the stacked 9 basis tenors plotted in Figure 4a. Notice that the proxies for funding costs and counterparty risk are also maturity specific, but the remaining variables are fixed across the basis term structure.

We first present the pooled estimation results for the regression coefficients in Eq. (12), denoted as β_j(Full), obtained by pooling the observations for all segments of the basis term-structure B_t. Next, we split the different segments of the term-structure into short (6M, 1Y and 3Y; β_j(Short)), medium (5Y, 7Y and 10Y; β_j(Medium)) and long (15Y, 20Y and 30Y; β_j(Long)) basis tenors, and only use the observations corresponding to those observations to estimate segment-specific regression coefficients. In this way, we can test whether the limits-to-arbitrage factors affect different segments of the basis

[8] We explicitly refrain from using the smoothed estimates $\widehat{\mu}^b_{t|T}(\tau)$ to avoid using future data in the estimation of the basis.

[9] This variable is publicly available as an end-of-week variable. After taking first differences, $\Delta \mathrm{ML}_t$ was filled backward to obtain a time-series at daily frequency. Daily changes in the remaining variables were filled forward in case of missing values.

term-structure differently.

We also allow the β coefficients in Eq. (12) to vary over time, indicated by the index t. This follows the arguments in Augustin and Schnitzler (2020) and Bai and Collin-Dufresne (2018), who emphasize that the relation between the basis and its drivers may not be constant over different periods. To accommodate this, we adopt a 261-day rolling window regression set-up. Tho ensure comparability of the coefficients both across time and across the different regressors, we standardized all explanatory variables in each rolling regression. The βs can then be interpreted as sensitivities to a one (time-varying) standard-deviation shock to the regressor.

Figure 6 shows the levels and first differences of the six basis determinants used in Eq. (12). During the Covid-19 lockdown period in March 2020, large jumps can be seen in variables such as market liquidity, the default premium, funding liquidity, and idiosyncratic sentiment. Note that similar large shocks were observed in the basis itself; see Figure 4a. This poses a non-negligible robustness risk to the rolling least-squares regressions. To avoid that our results are too much driven by such an incidental period, we winsorize the first differences of all independent and dependent variables at the 2% and 98% level.[10] Figure 6a also shows that the proxies for funding costs and counterparty risk can take negative values. These relate to curve inversions. For funding costs for instance, short-term LIBOR rates were higher than long-term rates between 2019 and 2020. Similarly, short-term JPM counterparty risk betas exceeded their long-term counterparts during several sub-periods.

4.3.2 Pooled results

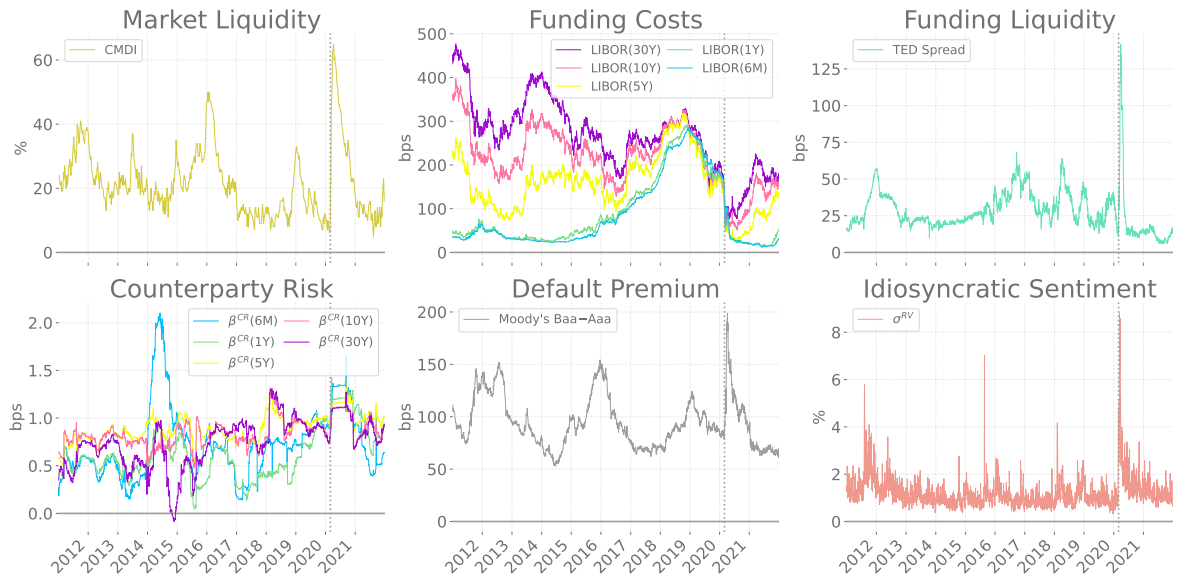
Figure 7 summarizes the results for the pooled panel regressions across all maturities. Each subfigure shows the rolling window estimates for one of the regression coefficients in Eq. 12 along with its 90% pointwise confidence band. Note that as the rolling window has around 261 or one trading year of daily observations, the plots start in 2012.

The first thing we notice is that there is quite some time-variation in the impact of the different determinants on the overall basis level and that all determinants appear

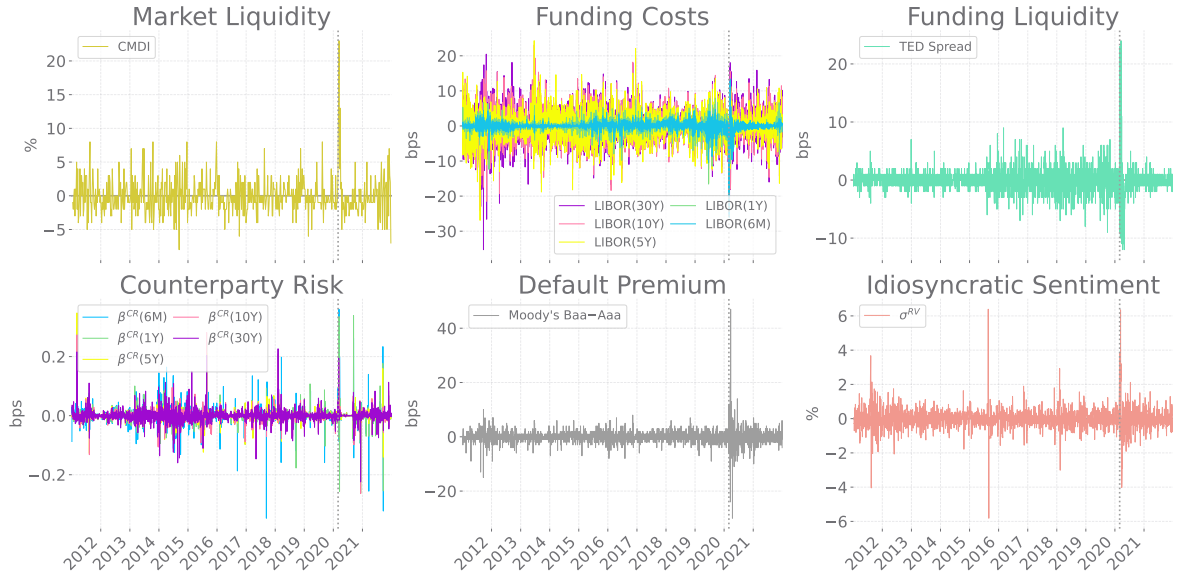
[10] This is a rather conservative choice compared to the 0.5th and 99.5th percentile clipping of the bases in Augustin and Schnitzler (2020) and Bai and Collin-Dufresne (2018).

Figure 6: Basis drivers for JPM

(a) Drivers in levels

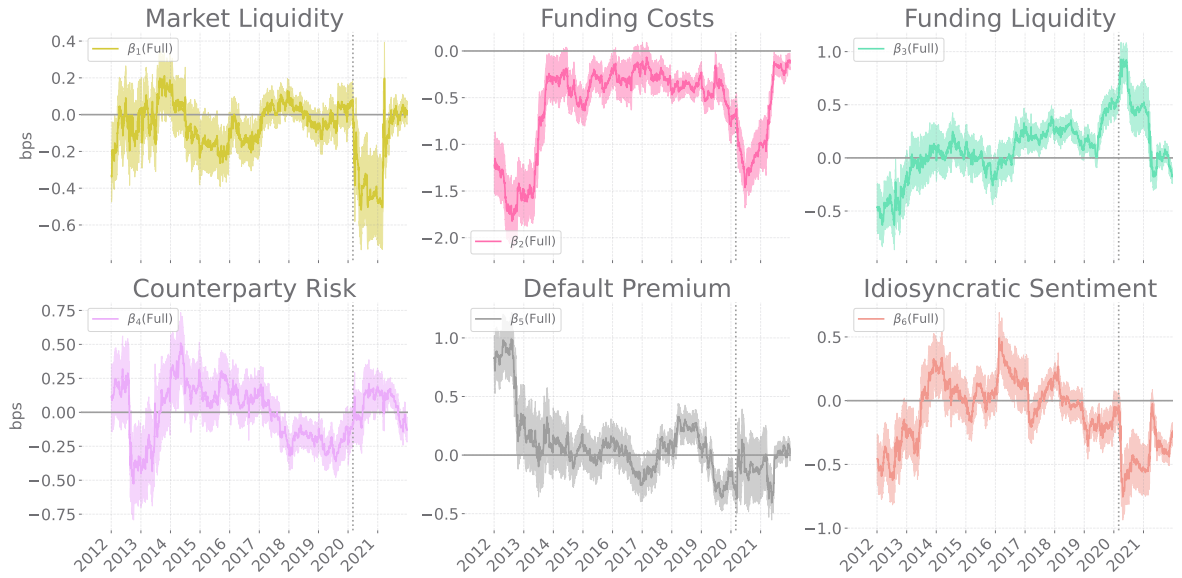


(b) Drivers in differences



Notes: This figure displays the proxies for risk factors (in levels and differences) for explaining dynamics of bases. A first order difference was taken for all variables, except for idiosyncratic sentiment, whose 5-day differences are plotted. All variables are measured in basis points, except the proxy for market liquidity, which is measured in percentages. The factors market liquidity, funding costs, funding liquidity and default premium are market-wide drivers, whereas counterparty risk and idiosyncratic sentiment are JPM specific. The dotted black vertical line marks the time point March 1, 2020, of the Covid-19 lockdowns.

Figure 7: Term-structure-level Rolling window regression coefficients



Notes: This figure displays the 261-day rolling-window regression coefficients (in bps) based on Eq. (12),

$$\Delta B_{i,t}(\text{Full}) = \beta_{0,t} + \beta_{1,t}\Delta\text{ML}_t + \beta_{2,t}\Delta\text{FC}_{i,t} + \beta_{3,t}\Delta\text{FL}_t + \beta_{4,t}\Delta\text{CR}_{i,t} + \beta_{5,t}\Delta\text{DP}_t + \beta_{6,t}\Delta^5\text{IS}_t + \varepsilon_{i,t},$$

for JPM bases computed by the `PL-ID-VC` model. The results are computed by pooling all maturities into a single regression with six explanatory variables: market liquidity (ML), funding costs (FC), funding liquidity (FL), counterparty risk (CR), default premium (DP) and idiosyncratic risk (IS), denoted as B(Full). Dependent and independent variables are winsorized at their 2nd and 98th percentiles. The shaded areas surrounding the estimates correspond to 90% confidence intervals based on the Newey-West adjusted covariance matrix. The dotted black vertical line marks March 1, 2020, the time around the Covid-19 lockdowns.

significant during some sub-period of the full 2011-2021 sample. Only funding costs seems to be a determinant with a persistent negative sign over the entire sample. This makes intuitive sense: higher funding costs make it more expensive to enter negative basis trades due to the need to finance the long position in the bond, thus widening a(n already on average) negative basis. All the other variables show a much more mixed pattern with periods of positive and negative impact. For instance, counterparty risk and the default premium seem to be particularly strong and positive during the aftermath of the 2011-2012 credit crunch. For most of the remaining sample period, their impact hovers around zero, with some incidental modest departures. Funding liquidity, like funding costs, has a negative impact at the start of the sample, as has idiosyncratic sentiment. The former is intuitive and in line with the results for funding costs. Also the latter is in line with earlier results by for instance Cai et al. (2020): reduced uncertainty makes basis trades more profitable and drive the basis towards zero.

The Covid-19 crisis marks several interesting patterns regarding the impact of the different economic determinants on the basis. First, we note that in the light of this crisis, the impact of the different limits-to-arbitrage factors changes rapidly over time. For some factors like funding costs, funding liquidity, default premium and idiosyncratic sentiment the shift away from zero already takes place before the heat of the crisis (indicated by the vertical line in the figure). For other factors, the shift only starts during the crisis (market liquidity) or hardly any shift is visible at all (counterparty risk). Only funding costs seem to have the expected negative impact on the basis. Other variables have less expected signs. For instance, funding liquidity becomes significant with an unexpected positive in the on-set and during the heat of the Covid-19 crisis. Similarly, market liquidity, the default premium, and idiosyncratic sentiment all start having unexpected negative signs around and during the Covid crisis. The results can probably be attributed to a combination of factors. First, the cheap funding made available by central banks at times when markets became illiquid might have reversed the relation temporarily between the determinants and the basis. Second, incidental government rescue and support packages for the economy at large during Covid times might also have dislocated the relation between the economic drivers and the basis. And finally, the impact between the economic drivers and the basis might be different across maturities, such that the current pooled estimates might be biased. For instance, we see in Figure 4a that the basis moves towards (and into) positive territory at the short end of the maturity spectrum when we get close to (into) the Covid crisis. After the 2020 lockdowns, the 6M basis becomes clearly positive, possibly reversing any relations with the economic drivers, whereas the basis remains substantially negative at the long end. To resolve the possible heterogeneity of the relationship between the economic drivers and the basis over time and across maturities, we therefore split out the regressions over the maturity spectrum in the next section.

Table 6 summarizes the results numerically, and also provides some indications of model fit. We consider the full sample, the first four months of the Covid-crisis following the lockdowns, and the subsequent six months. The left hand of the table provides the average of the rolling window regression estimates of Figure 7. Averaged over the entire period, we see that that only funding costs (FC) is again significantly negative. The other coefficients are insignificant, given their variation over time as visualized in Figure 7.

Table 6: Average rolling window regression results of pooled bases in Figure 7

Panel A: Regression coefficients (bps)				Panel B: Adj.R^2 contribution/regressor (%)			
Period:	**Jan 2011 - Dec 2021**	**Mar 2020 - Jun 2020**	**Jul 2020 - Dec 2020**	**Period:**	**Jan 2011 - Dec 2021**	**Mar 2020 - Jun 2020**	**Jul 2020 - Dec 2020**
Const (+/−)	0.051	0.172**	0.458***	All	3.647	10.503	9.314
	(0.059)	(0.078)	(0.086)				
ML (−)	-0.074	-0.291**	-0.421***	ML	0.080	0.414	0.749
	(0.071)	(0.113)	(0.134)				
FC (−)	-0.605***	-1.072***	-1.183***	FC	2.187	5.697	6.806
	(0.111)	(0.134)	(0.139)				
FL (−)	0.102	0.812***	0.476***	FL	0.473	3.283	0.996
	(0.089)	(0.092)	(0.117)				
CR (−)	0.000	-0.039	0.139	CR	0.185	-0.023	0.061
	(0.099)	(0.107)	(0.098)				
DP (−)	0.068	-0.082	-0.131	DP	0.249	0.113	0.046
	(0.101)	(0.146)	(0.157)				
IS (+/−)	-0.110	-0.527***	-0.503***	IS	0.427	1.479	1.210
	(0.089)	(0.106)	(0.118)				

Notes: This table reports the average regression results of the 261-day rolling window regression B(Full) (see Figure 7) based on Eq. (12) for JPM bases computed by the `PL-ID-VC` model. The averages are reported for different sample periods and basis maturity spectra. Panel A reports the average regression coefficients for each variable, with their Newey-West standard errors in brackets. The significance of the coefficients at the 10%, 5% and 1% is indicated by '*', '**', '***', respectively. Similarly, Panel B reports the average of the adjusted R^2 (%) of the regressions including all explanatory variables, but also their individual contributions (increase/decrease) in explanatory power via the leave-one-variable-out strategy. See caption of Figure 7 for more details.

At the start of the Covid period, in addition market liquidity, funding liquidity and idiosyncratic risk become significant, of which only market liquidity and idiosyncratic risk remain significant during the remainder of the covid period. The overall adjusted-R^2 of the regressions is quite low at almost 4%, going up to 10-11% in the Covid period. Looking at the leave-one-out decreases in adjusted-R^2 as reported in the last column, more than half of this is captured by funding liquidity, in line with the earlier results. In Figure C.1 in Appendix C.1 we present further rolling window adjusted-R^2 values, showing that the adjusted-R^2 can go up to 16% for some periods and regressors, and that jointly the limits to arbitrage can account for up to 20% of the 261-day time-series variation of the pooled bases, again with funding costs accounting for most of this.

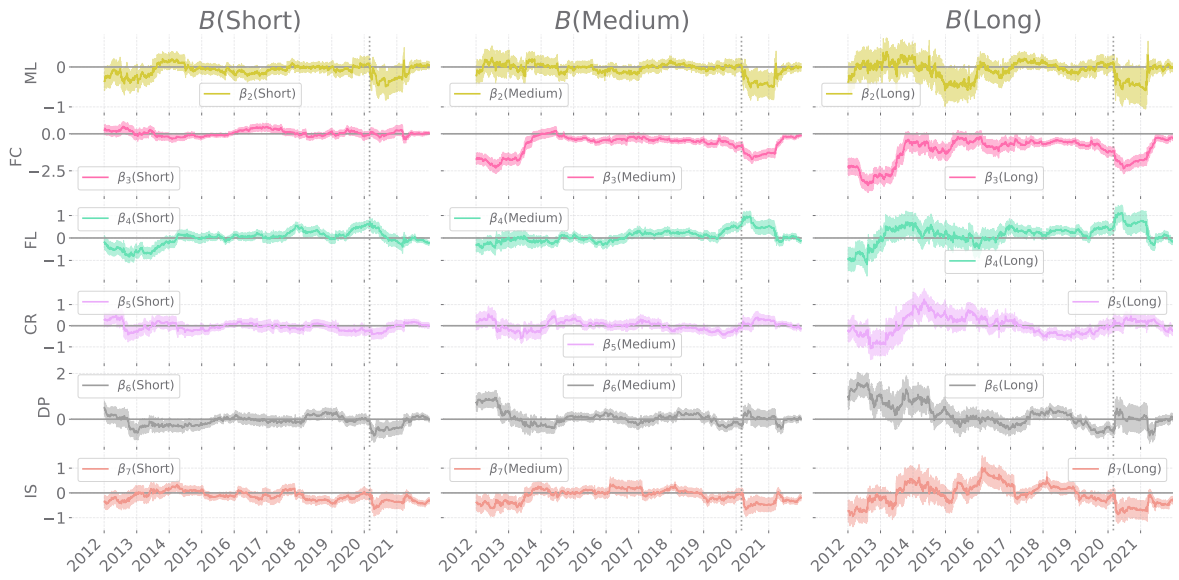
4.3.3 Maturity-segmented results

We now present the regression results different across the three maturity segments of JPM's basis term-structure. Figure 8 summarizes the results graphically. Broadly speaking, the results are in line with the pooled results in terms of secular dynamics and sign consistency. However, also some striking differences can be seen. For instance, the consistent negative sign of funding costs (FC) in the pooled regression has quite a distinct pattern across the three different maturity segments. At the short end of the curve, funding costs seem less of an issue and the magnitude of the coefficients is small, suggesting funding costs are not a particular driver of the basis in that part of the term-structure. At medium and long maturities, however, we the significant negative sign coming back in, as well as the significant large negative swings at the start of the sample (credit crunch and aftermath) and during Covid times. This is plausible: funding costs are likely to play more of a role in setting up long-maturity negative basis arbitrage trades than short-maturity ones given the long position in the bond has to be held (and thus funded) for a longer period of time.

Similar differences across the maturity spectrum are visible for other variables. For instance, the impact of counterparty risk and of the default premium are much more pronounced at the long compared to the short end of the basis curve. Again, this makes intuitive sense. For counterparty risk, for instance, we expect a positive sign as a long position in the CDS has to be taken. The positive impact, however, seems only relevant during the first part of the sample, and particularly at the long end of the basis curve, as there the CDS position and thus the counterparty risk relates to a longer period. By contrast, for funding liquidity, we see that most of the effects take place at the short and long end of the basis curve, whereas the medium maturities hardly show a relation with this determinant. Interestingly, the signs and dynamics of most of the determinants line up across maturities in times of stress such as during the Covid crisis, though their size and magnitude may differ over the different segments of the basis curve.

In Table 7 we test whether the coefficients are the same across maturities using a standard (generalized) Chow test (Dufour, 1982; Chow, 1960). We do so for each year in the sample, as well as for the entire sample. For each of the rolling samples in a one-year period, we compute the Chow-test. Panel A of the table reports the average of

Figure 8: Maturity-level rolling window regression coefficients



Notes: This figure displays the 261-day rolling-window regression coefficients (in bps) based on Eq. (12),

$$\Delta B_{i,t} = \beta_{0,t} + \beta_{1,t}\Delta \mathrm{ML}_t + \beta_{2,t}\Delta \mathrm{FC}_{i,t} + \beta_{3,t}\Delta \mathrm{FL}_t + \beta_{4,t}\Delta \mathrm{CR}_{i,t} + \beta_{5,t}\Delta \mathrm{DP}_t + \beta_{6,t}\Delta^5 \mathrm{IS}_t + \varepsilon_{i,t},$$

for JPM bases computed by the `PL-ID-VC` model. B(Full) pools all maturities into the regression, whereas B(Short), B(Medium) and B(Long) only include the maturities [6M, 1Y, 3Y], [5M, 7Y, 10Y], and [15M, 20Y, 30Y], respectively. The (standardized) explanatory variables (market liquidity, funding costs, funding liquidity, counterparty risk, default premium and idiosyncratic risk) are identical for all three regressions. The regression coefficients are color coded by variable name and the shaded areas surrounding the estimates correspond to 90% confidence intervals based on the Newey-West adjusted covariance matrix. The dotted black vertical line marks the time point March 1, 2020.

the test values, while panel B reports the number of times the test exceeds its standard critical value. The high rejection numbers in panel B clearly indicate that the impact of the limits-to-arbitrage factors differs across the basis maturity spectrum. Though there is an obvious multiple testing problem, the results also hold ground when the test is performed on 11 non-overlapping yearly periods, separately. The Chow test statistic for this set-up and the previous one are plotted in Figure 9. It underlines that different economic determinants may not only be relevant during different time periods, as argued in Augustin and Schnitzler (2020) and Bai and Collin-Dufresne (2018), but that the impact of different limits-to-arbitrage arguments may also differ across maturities. Again, it points to conclusion that the basis term-structure cannot be assumed to be flat, nor its dynamics identical across maturities.

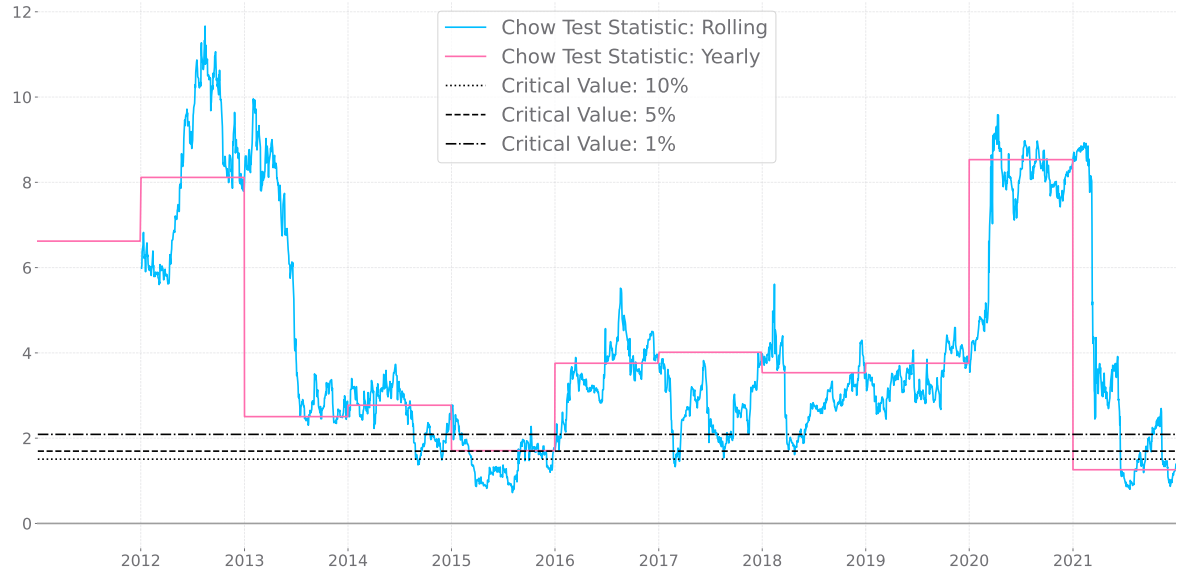
In Table 8, we summarize the results again numerically by averaging the rolling window coefficients over time, and reporting the adjusted-R^2 values for the full sample

Table 7: Chow test results for JPM bases computed by PL-ID-VC model

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Avg
Panel A: Number of days and average test statistic per year											
T	260	261	261	261	261	260	261	261	262	260	2608
CT	8.265	5.352	2.584	1.490	3.566	2.665	3.043	3.395	7.513	3.297	4.117***
Panel B: Number of null-hypothesis rejections per year (%)											
$\alpha = 10\%$	100	100	97.318	48.275	100	98.076	100	100	100	65.000	90.870
$\alpha = 5\%$	100	100	91.954	35.632	100	93.461	99.233	100	100	60.769	88.108
$\alpha = 1\%$	100	100	73.563	8.812	94.636	77.307	84.291	100	100	54.615	79.324

Notes: This table summarizes the Chow test results for the null-hypothesis $H_0 = \beta_{j,t}(\text{Full}) = \beta_{j,t}(\text{Short}) = \beta_{j,t}(\text{Medium}) = \beta_{j,t}(\text{Long}), \text{for}, j = 0, \ldots, 6$. The Chow test statistic (CT) considers the estimates in Figures 7 8 and for each rolling regression for JPM using the bases of the PL-ID-VC model. The test statistic is computed for each day, separately. The daily test statistic is CT $\sim F(d_1 = 14, d_2 \in [2193, 2328])$, with critical values (approximately) 1.508, 1.696 and 2.089, at the $\alpha = 10\%$ (*), $\alpha = 5\%$ (**), $\alpha = 1\%$ (***) significance level, respectively. The yearly average of the test statistic and the number of days per year are reported in Panel A. The number of null-hypothesis rejections (%), at the aforementioned significance levels, is reported in Panel B. The daily test statistics are plotted in Figure 9.

Figure 9: Chow test statistics for JPM bases computed by PL-ID-VC model



Notes: This figure displays the Chow test statistics for the null-hypothesis $H_0 = \beta_{j,t}(\text{Full}) = \beta_{j,t}(\text{Short}) = \beta_{j,t}(\text{Medium}) = \beta_{j,t}(\text{Long}), \text{for}, j = 0, \ldots, 6$. The test statistics is obtained based on a daily 261-day rolling-window (blue) and 11 yearly non-overlapping period (pink) regressions of the basis model in Eq. (12). The critical values at the $\alpha = 10\%$ (*), $\alpha = 5\%$ (**), $\alpha = 1\%$ (***) significance level are plotted in plotted (black). See caption of 7 for more details.

and the start and continuance of the Covid period. We confirm our earlier findings. Over the entire sample, only funding costs appear significant, but only at the medium and long end of the basis curve. At the short end, no significant effect is found. The

Table 8: Average rolling window regression results

Period:	**Jan 2011 - Dec 2021**			**Mar 2020 - Jun 2020**			**Jul 2020 - Dec 2020**		
Maturity domain:	**Short**	**Medium**	**Long**	**Short**	**Medium**	**Long**	**Short**	**Medium**	**Long**
Panel A: Regression coefficients (bps)									
Intercept (+/−)	0.042	0.062	0.048	0.107	0.198*	0.21	0.392***	0.492***	0.491***
	(0.092)	(0.093)	(0.133)	(0.118)	(0.118)	(0.159)	(0.137)	(0.126)	(0.172)
ML (−)	-0.088	-0.053	-0.079	-0.242	-0.276*	-0.324	-0.353*	-0.419**	-0.415
	(0.102)	(0.101)	(0.158)	(0.17)	(0.162)	(0.222)	(0.195)	(0.187)	(0.266)
FC (−)	0.037	-0.698***	-1.136***	-0.024	-1.329***	-1.849***	-0.005	-1.449***	-1.991***
	(0.135)	(0.144)	(0.229)	(0.172)	(0.184)	(0.226)	(0.183)	(0.161)	(0.214)
FL (−)	-0.012	0.087	0.176	0.485***	0.853***	0.944***	-0.004	0.491***	0.645***
	(0.116)	(0.117)	(0.203)	(0.131)	(0.131)	(0.199)	(0.144)	(0.164)	(0.245)
CR (−)	-0.041	0.004	0.019	-0.348**	0.067	0.098	-0.157	0.291**	0.278*
	(0.132)	(0.133)	(0.22)	(0.138)	(0.159)	(0.197)	(0.14)	(0.147)	(0.164)
DP (−)	-0.094	0.057	0.21	-0.489***	0.15	0.143	-0.428**	0.040	0.005
	(0.131)	(0.133)	(0.22)	(0.175)	(0.174)	(0.291)	(0.178)	(0.202)	(0.338)
IS (+/−)	-0.144	-0.117	-0.061	-0.450***	-0.457***	-0.625***	-0.326*	-0.459***	-0.655***
	(0.115)	(0.123)	(0.199)	(0.162)	(0.159)	(0.206)	(0.172)	(0.167)	(0.225)
Panel B: Adjusted R^2 contribution per regressor (%)									
All	1.837	6.166	6.716	5.407	17.644	16.42	3.527	17.834	15.091
ML	0.055	0.039	0.015	0.364	0.424	0.268	0.704	0.961	0.377
FC	0.205	4.471	4.488	-0.033	10.872	11.289	-0.035	13.801	12.131
FL	0.604	0.496	0.459	1.732	4.401	2.872	-0.033	1.38	1.099
CR	0.134	0.223	0.43	0.762	-0.051	-0.059	0.154	0.463	0.136
DP	0.293	0.301	0.54	1.674	0.243	0.261	1.173	-0.088	-0.103
IS	0.478	0.424	0.51	1.431	1.384	1.317	0.653	1.328	1.222

Notes: This table reports the average regression results of the 261-day rolling window regressions for different basis segments (short, medium and long) of the bases for JPM computed by `PL-ID-VC` model, see Figure 8. The averages are reported for different sample periods and basis maturity spectra. Panel A reports the average regression coefficients for each variable, with their Newey-West standard errors in brackets. The significance of the coefficients at the 10%, 5% and 1% is indicated by '*', '**', '***', respectively. Similarly, Panel B reports the average of the adjusted R^2 (%) of the regressions including all explanatory variables, but also their individual contributions (increase/decrease) in explanatory power via the leave-one-variable-out strategy. See caption of Figure 8 for more details.

other determinants appear insignificant over the full sample due to the time-variation in the sign of their impact on the basis.

At the start of and during the covid crisis, more determinants enter the model significantly with their expected signs. However, the determinants do not enter uniformly across all maturities, but affect different segments of the curve differently as explained before. For instance, at the start of the crisis the idiosyncratic sentiment becomes a

significant driver for the short, medium and long-run basis, in line with the arguments of Cai et al. (2020) that reduced uncertainty drives the (negative) basis closer to zero. In fact, the effect appears to be monotonically increasing for this determinant across maturities. For the remainder of the crisis, however, the effect remains significantly negative, but shrinks in impact for short-term bases. On the other hand, short-term counterparty risk was only found significant with the expected sign at the beginning of the crisis. Hence, similar to the findings in Augustin and Schnitzler (2020) and Bai and Collin-Dufresne (2018), our results suggest that the influence of limit-to-arbitrage factors (and explanatory power) outside volatile periods significantly decreases, but may also start to correlate in opposite directions. Finally, we see that the funding liquidity beta for short-term bases drops quickly from being positive before Covid and becomes negative during the heat of the crisis. This is in line with the flight-to-equality effects of Brunnermeier and Pedersen (2009), where funding liquidity matters most during market turmoil.

The model fit in terms of adjusted-R^2 also differs across the maturity spectrum. It appears more difficult to describe the dynamics of the short end of the basis curve with the given set of determinants. The adjusted-R^2 values are typically less than half those of the medium and long maturity regressions. The latter are also typically higher than when pooling all the maturities into one single regression. Again, we confirm that funding costs are the most important determinant of the basis, but that jointly the other regressors can still cause a substantial further increase in the explanatory power of the model.

4.4 Robustness checks

4.4.1 Alternative interpolation models

To examine the robustness of our findings from the prior subsections, we repeat the regression analysis using the CDS-bond basis curve time-series of the `NS-ID-VC` model. This parametric model imposes a more stringent structure on the curves, preventing overfitting of spreads, albeit with the trade-off of increased pricing errors (see the results in Table B.2). The Figures C.2 and C.3 in Appendix C.1 indicate that the `NS-ID-VC` model also maintains accuracy and captures the variety in observed bases shapes well.

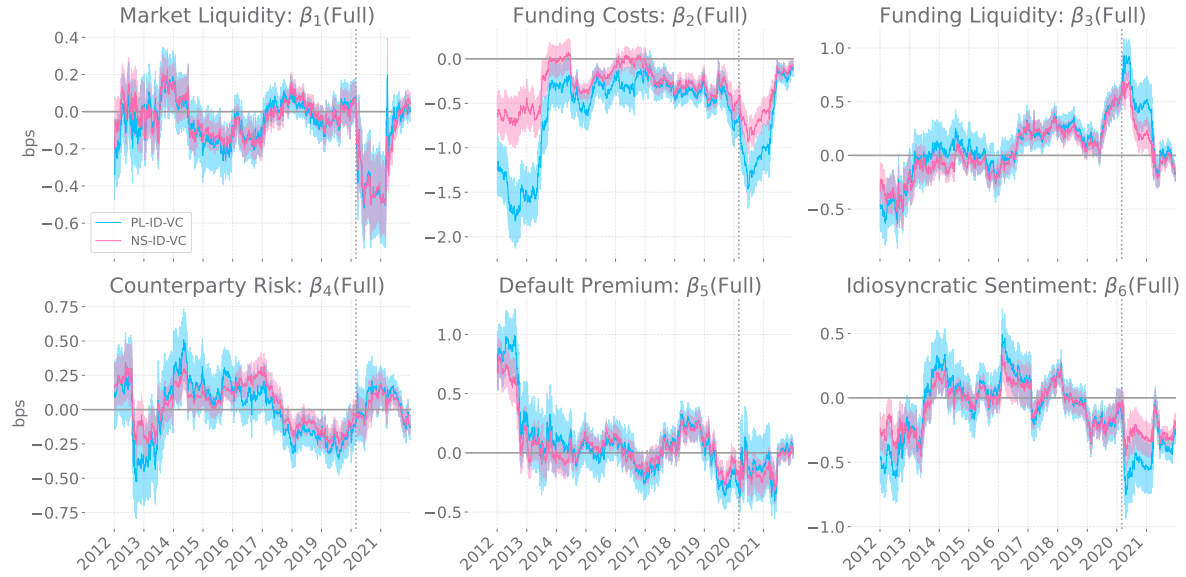
Our conclusion that the basis term-structure is not flat thus holds up for the NS-ID-VC model, given the importance of the idiosyncractic short-term components for the curve in the model.

The robustness of the results is further confirmed in Figure 10a. Panel (a) compares the pooled rolling regression coefficients of the two models based on Eq. (12). The patterns and magnitude of the pooled coefficients of the NS-ID-VC model closely resemble those of the PL-ID-VC model. Although, in the beginning of the sample a large difference in magnitude is observed between the two funding costs beta. It turns out, that the level differences in this period are due higher short-term funding costs betas for the NS-ID-VC.

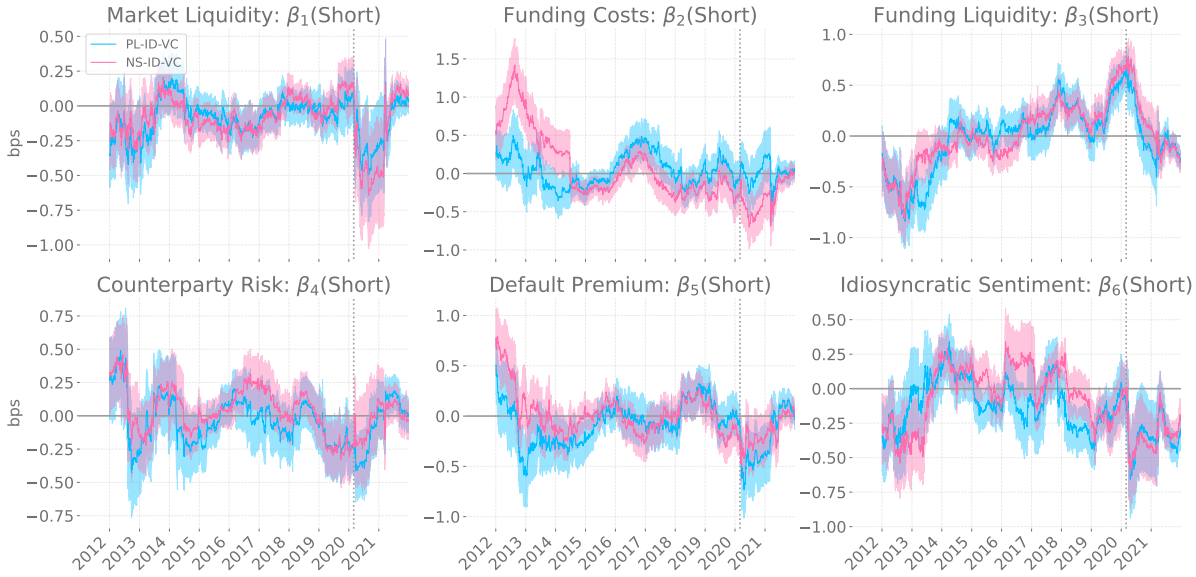
Panel (b) of Figure 10 presents the rolling regression results for both models, but now for the short end of the basis term-structure only. Full results for all segments can be found in Appendix C.1. Panel (b) shows that the secular movements of the coefficients for both models have many similarities. This is comforting and supports the robustness of the results. The panel also shows, however, that for the short-end part of the curve there are some marked differences between the two interpolation models. For example, the time-series pattern of the funding costs betas align, but the level of this beta for the NS-ID-VC model is dislocated up to Jun-2014. Further investigation reveals that this difference is caused by incidental spikes in 2013 in the 6M LIBOR, that still dominate after censoring this data. These spikes, however, do not affect the short-term funding cost beta of PL-ID-VC, which is likely due to its imputed bases being closer to the empericial data points, and hence being also noisier. In contrast, the betas for remaining determinants show much more consistency for the NS-ID-VC model and hardly change when comparing the pooled ones in panel (a) with short ones in panel (b). We attribute this to the tighter parametric form of the Nelson-Siegel specification, which allows less flexibility at the short end of the curve. Also at the medium and long end of the curve there are some differences between the models, albeit of a much smaller magnitude. In particular, the differences between medium and long end coefficients are less pronounced for the NS-ID-VC model. Again this is likely caused by its parametric specification: whereas the PL-ID-VC model linearly extrapolates at the upper end of the curve for any missing observations, thus causing larger differences between the coefficients for the medium and long end, the extrapolation of the NS-ID-VC model is dampened due its concave shape for long maturities.

Figure 10: `PL-ID-VC` **vs** `NS-ID-VC` **rolling window regression coefficients for JPM**

(a) Pooled



(b) Short segment



Notes: This figure displays the 261-day rolling-window regression coefficients (in bps) based on Eq. (12), for JPM bases computed by the `PL-ID-VC` model (blue) and `NL-ID-VC` model (pink). Subfigure 10a compares the regression coefficient obtained for B(Full) and Subfigure 10b compares B(Short). Shaded areas correspond to 90% confidence intervals based on Newey-West standard errors, see caption of Figure 7 and 8.

The smaller differences between the coefficients across the maturity spectrum for the `NS-ID-VC` model also has effects on the Chow test results from Table 9. In particular, the coefficients of the economic drivers are no longer significantly different across the different

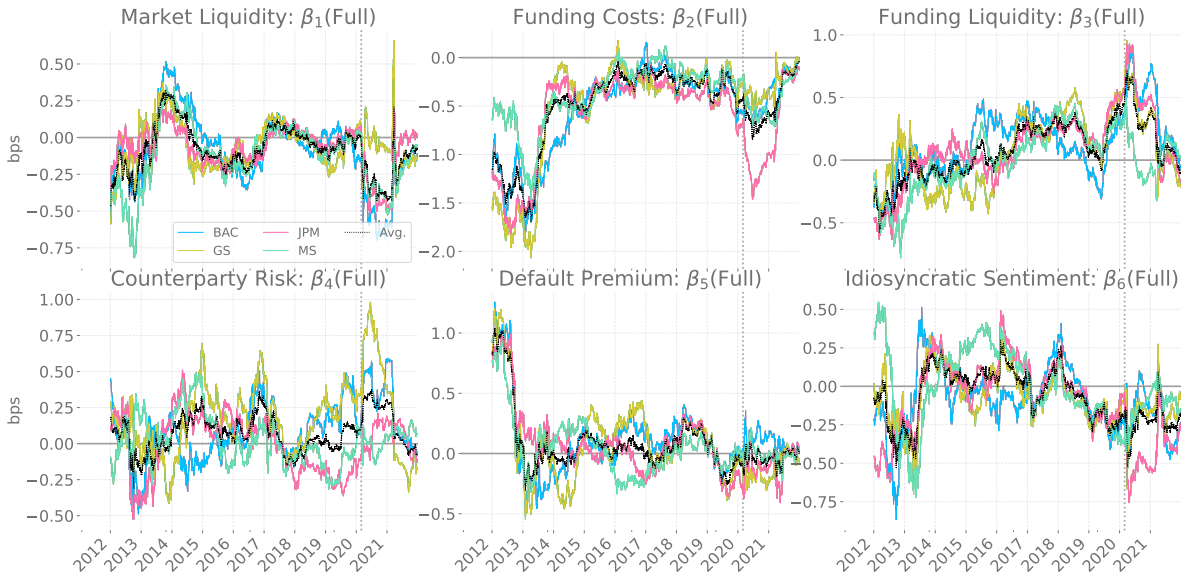
segments of the basis curve between 2013-2019 and 2021. For example, during 2020, the coefficients across different maturities were only significantly different 24% of the time at the 90% confidence level, whereas for the `PL-ID-VC` model this was always the case during Covid-19. This comes as no surprise in light of the earlier comments: the parametric form of the Nelson-Siegel model apparently may be too restrictive across maturities, resulting in a similar impact of the economic drivers for each maturity. Given the better fit to the data both in-sample and out-of-sample of the `PL-ID-VC` model, we put more emphasis on the results based on the latter. Also note that even for the `NS-ID-VC` this still leaves the conclusion on the existence of a non-flat basis term-structure intact, as the latter hinges on the need of the idiosyncratic short-end components in the state-space model (3). It is only the differential impact of the economic drivers across this non-flat basis term-structure that would fall away under the more restrictive parametric form of the `NS-ID-VC` model.

4.4.2 Results for other prime financial institutions

We further asses the robustness of our results by studying the dynamics of the CDS-bond basis term-structure for other blue-chip financials. For this, we consider the CDS and Z-spread data of Bank of America (BAC), Goldman Sachs (GS) and Morgan Stanley (MS). Just like JPM, these are prominent U.S. financial institutions. The empirical basis term-structures have to be constructed from data with similar missing data patterns as JPM. Just as we did for JPM, we use the `PL-ID-VC` state-space model to impute missing values and construct the historical basis term structures. We then analyze these and investigate their relation to the economic drivers using the same rolling regression framework as before.

Figure 11 shows the rolling window coefficients for each of the banks based on the full in-sample period. We see both striking similarities as well as some marked differences. First, the funding cost coefficient has a very similar pattern for all four banks. Its impact on the basis is negative across the entire time period, with particular troughs for all banks in the aftermath of the credit crunch (2011-2012) and in the Covid years 2020-2021. Also market liquidity is quite similar for MS, JPM, and BAC. Only GS shows a different pattern during the Covid-19 crisis, with market liquidity having hardly any effect for GS,

Figure 11: Rolling window regression coefficients for pooled bases across firms



Notes: This figure displays the 261-day rolling-window regression coefficients (in bps) based on Eq. (12),

$$\Delta B_{i,t}(\text{Full}) = \beta_{0,t} + \beta_{1,t}\Delta\text{ML}_t + \beta_{2,t}\Delta\text{FC}_{i,t} + \beta_{3,t}\Delta\text{FL}_t + \beta_{4,t}\Delta\text{CR}_{i,t} + \beta_{5,t}\Delta\text{DP}_t + \beta_{6,t}\Delta^5\text{IS}_t + \varepsilon_{i,t},$$

for BAC, GS, JPM and MS bases computed by the `PL-ID-VC` model. The B(Full) pools all maturities into a single dependent variable and is regressed on (standardized) 6 explanatory variables: market liquidity, funding costs, funding liquidity, counterparty risk, default premium and idiosyncratic risk. Only counterparty risk and idiosyncratic sentiment are firm-specific. The two economic variables: market liquidity and funding liquidity are market-wide factors and are exactly the same in all four regressions, whereas funding costs are maturity-dependant. The proxy for counterparty risk also differs per maturity. Each dependent and independent variable was winsorized at their 2nd and 98th percentiles. The regression coefficients are color coded by firm. The cross-sectional average of the rolling coefficient is plotted a black dotted time-series. The dotted black vertical line marks the time point March 1, 2020.

whereas it has an expected negative impact for the other institutions. In the 2011-2012 credit crunch, also the funding liquidity, default premium and idiosyncratic sentiment exhibit a consistent pattern across all institutions. This is rather surprising result for the latter determinant as it is a firm-specific covariate, implying that their realized volatility seems to co-move. However, the disparity across banks of the impact of counterparty risk is the largest, which also a firm-specific determinant. For this covariate, no consistent pattern emerges, neither over time nor in the cross-section of banks. This is also made clear by the cross-sectional average (in black), which remains close to zero throughout the sample.

The Chow tests reported in Table 9 confirm our earlier results for JPM. Also for BAC, GS, and MS, there is ample evidence that the impact of the limits-to-arbitrage factors differs over the different segments of the basis term-structure. This is also confirmed

Table 9: Chow test results for BAC, GS and MS bases computed by `PL-ID-VC` model

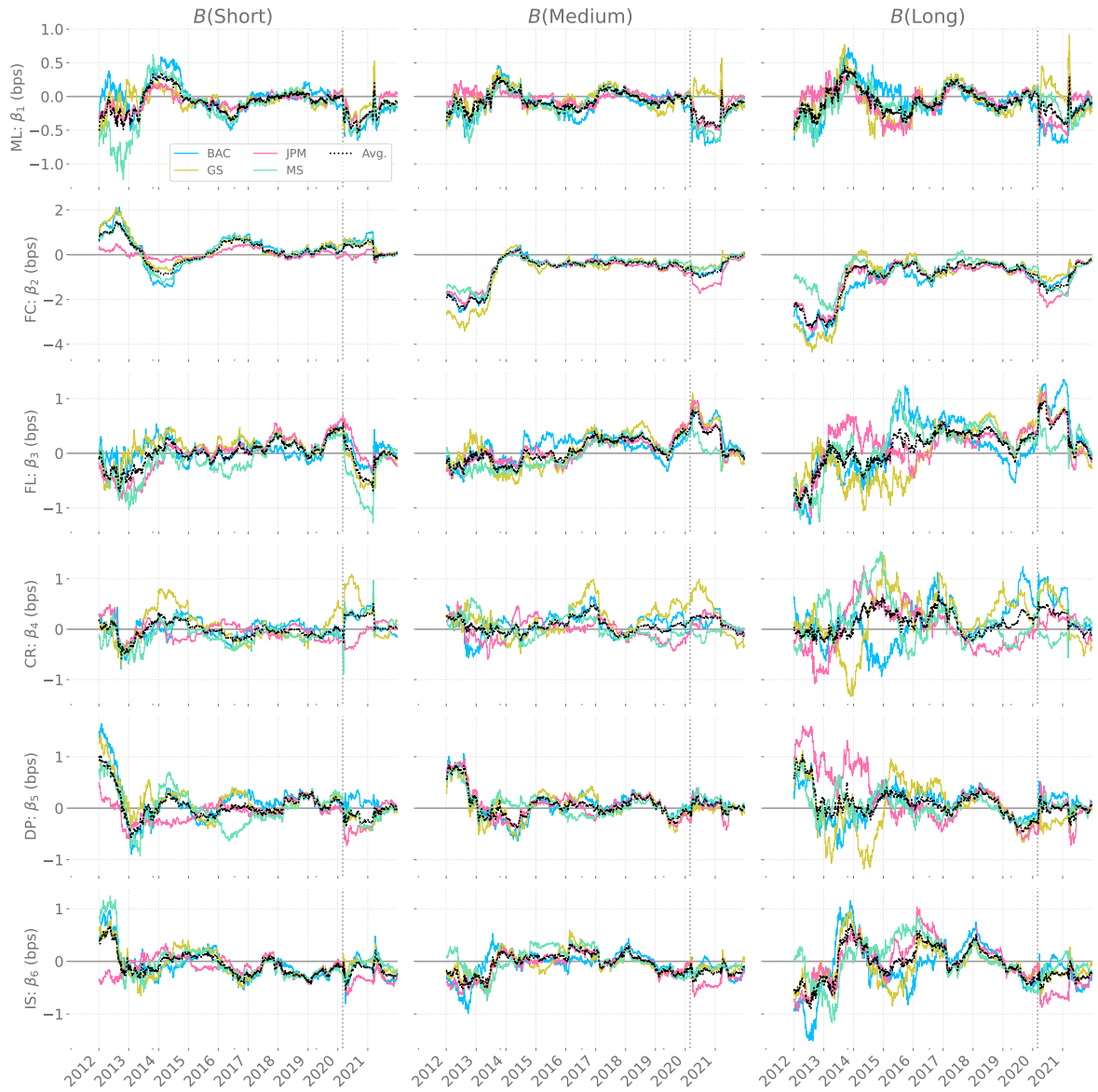
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Avg.
Panel A: Average Chow test statistic per year by firm											
BAC	10.184	5.339	2.299	1.888	6.424	3.554	1.987	5.779	5.545	2.625	4.562
GS	10.950	5.804	2.879	2.686	4.994	2.793	3.152	4.210	4.459	2.079	4.401
MS	5.243	3.133	2.590	3.264	6.142	1.957	2.802	3.452	6.931	4.053	3.957
Panel B: Number of null-hypothesis rejections per year (%) by firm											
BAC											
$\alpha = 10\%$	100	100	71.647	49.808	100	100	51.340	100.	100	64.230	83.641
$\alpha = 5\%$	100	100	66.283	47.126	100	100	47.892	100	100	48.461	80.908
$\alpha = 1\%$	100	100	57.854	39.846	100	98.076	39.463	100	100	34.615	76.905
GS											
$\alpha = 10\%$	100	100	100	88.122	88.505	53.461	81.992	100	100	55.769	86.804
$\alpha = 5\%$	100	100	98.467	80.076	88.505	52.307	80.842	100	100	50.	85.040
$\alpha = 1\%$	100	100	85.057	74.329	87.356	48.461	67.049	100	100	37.692	80.015
MS											
$\alpha = 10\%$	100	53.256	94.252	100	100	66.153	99.233	100	100	85.384	89.835
$\alpha = 5\%$	100	49.808	78.927	100	100	46.923	89.655	100	100	83.461	84.886
$\alpha = 1\%$	100	48.275	63.218	100	100	26.538	80.842	100	100	71.538	79.056

Notes: This stable summarizes the Chow test results for the null-hypothesis $H_0 = \beta_{j,t}(\text{Full}) = \beta_{j,t}(\text{Short}) = \beta_{j,t}(\text{Medium}) = \beta_{j,t}(\text{Long}), \text{for}, j = 0, \ldots, 6$. The Chow test statistic (CT) is considering the estimates for each rolling regression for BAC, GS and MS using the bases of the `PL-ID-VC` model. The test statistic is computed for each day, separately. The daily test statistic is CT $\sim F(d_1 = 14, d_2 \in [2193, 2328])$, with critical values (approximately) 1.508, 1.696 and 2.089, at the $\alpha = 10\%$, $\alpha = 5\%$, $\alpha = 1\%$ significance level, respectively.

by Figure 12, which shows the rolling window estimates per segment. Like before for JPM, funding costs appear to have a stronger impact at the medium and long end of the curve compared to the short end. The effects at the short end for BAC and MS, but particularly GS are stronger during 2011-2012 than for JPM, but still smaller than for the longer maturity segments. This again supports our conclusion that the economic drivers may have a different impact for different segments of the basis term-structure.

Also for the other determinants we find visual differences between the different maturity segments. Particularly the coefficients for the long maturities seem much more volatile, both over time and in the cross-section of banks. This holds for funding liquidity, counterparty risk, the default premium as well as for idiosyncratic sentiment. We should be careful to read too much into the differences between the banks as some of the rolling window coefficients are also characterized by large standard errors. In the end, mainly funding costs has a consistent pattern across time, banks, and maturities.

Figure 12: Maturity-level rolling window regression coefficients for each firm



Notes: This figure displays the 261-day rolling-window regression coefficients (in bps) based on Eq. (12),

$$\Delta B_{i,t} = \beta_{0,t} + \beta_{1,t}\Delta \mathrm{ML}_t + \beta_{2,t}\Delta \mathrm{FC}_{i,t} + \beta_{3,t}\Delta \mathrm{FL}_t + \beta_{4,t}\Delta \mathrm{CR}_{i,t} + \beta_{5,t}\Delta \mathrm{DP}_t + \beta_{6,t}\Delta \mathrm{IS}_t + \varepsilon_{i,t},$$

for BAC, GS, JPM and MS bases computed by the `PL-ID-VC` model. The dependent variables B(Short), B(Medium) and B(Long) pool include the maturities [6M, 1Y, 3Y], [5M, 7Y, 10Y], and [15M, 20Y, 30Y], respectively. The dependent variable is regressed on (standardized) 6 explanatory variables: market liquidity, funding costs, funding liquidity, counterparty risk, default premium and idiosyncratic risk. The regression coefficients are color coded by firm. The cross-sectional average of the rolling coefficient is plotted a black dotted time-series. The dotted black vertical line marks the time point March 1, 2020. See caption of Figure 11 for more details.

The main conclusion that remains is that the empirical basis dynamics are difficult to reconcile with the typical flat basis term-structure assumption: the dynamics are too

different between the short and long end, and the determinants of the basis at these different segments may differ or differ in magnitude, depending on the interpolation method used.

4.4.3 Basis sign dependence

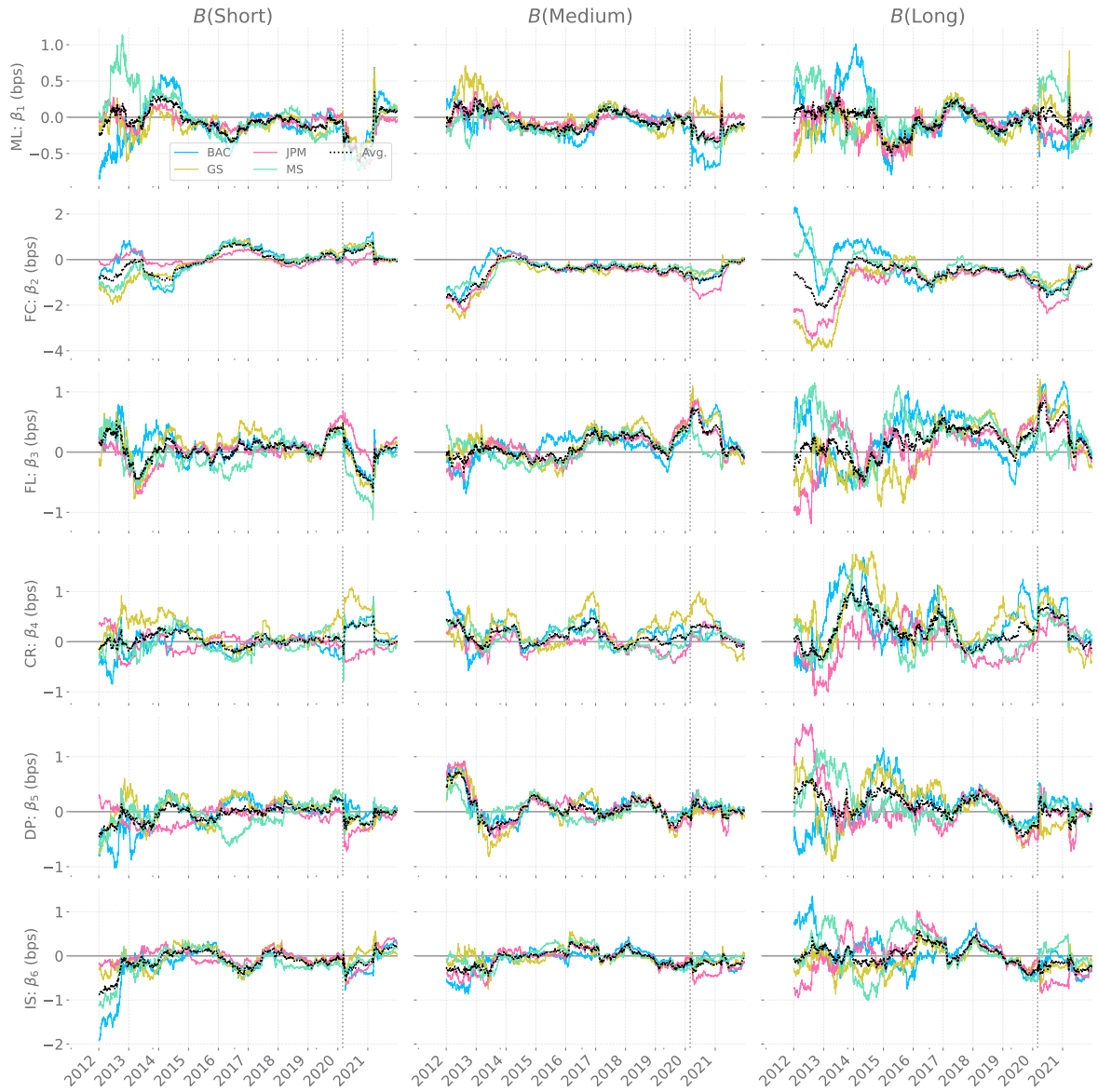
As a final robustness check, we analyze the relationship between basis determinants and the sign of the bases. Different segments of the CDS-bond basis term-structure may move in the opposite direction, giving rise to different basis signs across maturity, as evident by the troughs in Figure 4a for JPM. Different basis signs at different maturities, simultaneously, are also observed in the filtered bases of the other three prime companies. Positive basis signs are mainly observed for the short end of the curves, whereas the remaining parts of the curve remain negative. This potentially reverses the relationship between the bases and their economic drivers. The limit-to-arbitrage studies in Augustin and Schnitzler (2020) and Bai and Collin-Dufresne (2018) control for such effects in the regressions by including an additional basis sign-dependant control variate. We adopt a similar strategy to analyze the influence of the basis signs on the regression coefficients in Eq. (12). The baseline regression model with basis sign-adjustments now reads

$$\begin{aligned}\Delta B_{i,t} = \beta_{0,t} + \text{sign}(B_{i,t-1})\Big[&\beta_{1,t}\Delta\text{ML}_t + \beta_{2,t}\Delta\text{FC}_{i,t} + \beta_{3,t}\Delta\text{FL}_t + \beta_{4,t}\Delta\text{CR}_{i,t} \\ &+ \beta_{5,t}\Delta\text{DP}_t + \beta_{6,t}\Delta\text{IS}_t\Big] + \varepsilon_{i,t},\end{aligned} \tag{13}$$

where $\text{sign}(B_{i,t-1})$ takes the value 1 if $B_{i,t-1} < 0$ and -1 otherwise. The sign of each limit-to-arbitrage factor a time t is now controlled by the sign of the previous sign of the previous basis at the respective maturity. A negative sign will control for the potential reversal of the relationship between bases and their drivers when the basis becomes positive.

We now present the regression results for the different maturity segments of the four blue-chip company basis term-structure. The rolling regression parameters are identical as before in terms of window size, standardization and winsorization. Figure 13 (with sign control) summarizes the new results, similarly to previous Figure 12 (without sign control), across all segments and firms. The time-series dynamics of the betas of the two different regression are very similar, because the basis term-structures of the four firms are predominantly negative throughout 2011-2022. But some significant difference arise

Figure 13: Sign-dependant rolling window regression coefficients for each firm



Notes: This figure displays the 261-day rolling-window regression coefficients (in bps) based on the regression model with basis sign-adjustments Eq. (13),

$$\Delta B_{i,t} = \beta_{0,t} + \text{sign}(B_{i,t-1})\left[\beta_{1,t}\Delta\text{ML}_t + \beta_{2,t}\Delta\text{FC}_{i,t} + \beta_{3,t}\Delta\text{FL}_t + \beta_{4,t}\Delta\text{CR}_{i,t} + \beta_{5,t}\Delta\text{DP}_t + \beta_{6,t}\Delta\text{IS}_t\right] + \varepsilon_{i,t},$$

where $\text{sign}(B_{i,t-1})$ takes the value 1 if $B_{i,t-1} < 0$ and -1 otherwise, for BAC, GS, JPM and MS bases computed by the `PL-ID-VC` model. The dependent variables B(Short), B(Medium) and B(Long) pool include the maturities [6M, 1Y, 3Y], [5M, 7Y, 10Y], and [15M, 20Y, 30Y], respectively. The dependent variable is regressed on (standardized) 6 explanatory variables: market liquidity, funding costs, funding liquidity, counterparty risk, default premium and idiosyncratic risk. Only counterparty risk and idiosyncratic sentiment are firm-specific. The remaining variables are market-wide factors and are exactly the same in all four regressions. Each dependent and independent variable was winsorized at their 2nd and 98th percentiles. The regression coefficients are color coded by firm. The time-wise average of the rolling coefficient is plotted a black dotted time-series. The dotted black vertical line marks the time point March 1, 2020.

after controlling for the basis' signs. For example, the sign of the rolling coefficients for funding costs, default premium and idiosyncratic sentiment during 2011-2012 credit crisis and now take the expected negative sign for almost all firms. As a consequence, the cross-sectional estimate (black) now also signals the expected negative sign. Again, note that short basis tenors took a positive sign during this period. This result indeed confirms that economic basis drivers correlate differently when the basis is positive versus when it is negative. In addition, these results clearly suggest that when the influence of factors that limit arbitrage opportunities increases (e.g., during market turmoil), the basis curve further widens, regardless its sign.

Moreover, the medium term rolling coefficients for each firm in Figure 13 (with sign control) hardly differ from the previous regressions in Figure 12 (without sign control). However, this does not seem for the long segment coefficients. In particular, the ones for BAC and MS flip for almost all drivers throughout 2011-2012. This is an unexpected result as the bases are predominantly negative at the upper end of the curve, even the during the credit crisis, see e.g., Figure 4a for JPM. We attribute this unexpected result to the uncertainty in the extrapolation/imputation by the `P1-ID-VC` model. During 2011-2012 there were little to no liquid bond Z-spread observed with more than 10Y maturity. For the remaining history the beta patterns with sign adjustment align with the ones without sign adjustment.

5 Conclusion

In this paper, we addressed the challenge of the lack of corporate bond and CDS data when constructing the CDS-bond basis full term-structure histories at the firm level across maturities. To solve this challenge, we formulated a state-space model for the joint term structures in CDS and bond markets. The state-space approach allowed us to easily deal with many missing values in either market, and to filter the full basis term-structure from the remaining available data. The model set-up easily allowed for a range of curve interpolation techniques that could be selected by the analyst. By including both common risk factors for the CDS and Z-spreads, as well as basis-specific factors, we could describe firm-level basis curve dynamics without imposing the typical assumption of a flat basis term structure.

Empirical evidence revealed that a model with short-term basis dynamics significantly increases the model fit across maturities, both in-sample and out-of-sample. This points to the existence of a non-flat basis term-structure. Results are robust across different curve interpolation methods, with some methods like spline interpolation providing a better fit at the expense of less intuitive curve reversals at the short end, and others like the multi-curve extension of the dynamic Nelson-Siegel model providing robust and interpretable CDS-bond-basis term-structure histories at the expense of a somewhat reduced model fit.

We then used the estimated basis term-structure histories to perform rolling-window regressions to determine the basis determinants. We found that changes in the bases of major US financial institutions are primarily driven by changes in funding costs and by their firm-specific sentiment factors. While other factors occasionally impact the basis term-structure over the 2011-2021 sample period, it is particularly during the Covid-19 pandemic that all factors enter the model significantly during different times of the crisis. The statistical evidence reveals that the typical limits-to-arbitrage variables from the literature may not be significant at the same moment in time, nor that they uniformly have the same influence on the basis across the entire maturity spectrum, particularly in times of turmoil. In particular, different variables have a statistically significant impact on different parts of the basis term-structure. Again, these results based on our state-space modeling approach point to the existence of a non-flat, non-constant CDS-bond-basis term-structure.

References

Ahmadian, R. (2015). Complications in cds-bond basis analysis and modeling. Imperial College London Business School.

Augustin, P. and J. Schnitzler (2020). Disentangling types of liquidity and testing limits-to-arbitrage theories in the CDS-bond basis. *European Financial Management 27*(1), 120–146.

Bai, J. and P. Collin-Dufresne (2018). The CDS-bond basis. *Financial Management 48*(2), 417–439.

BCBS (2004). Basel II: Revised international capital framework. *Basel Committee on Banking Supervision*.

BCBS (2010). Basel III: A global regulatory framework for more resilient banks and banking systems. *Basel Committee on Banking Supervision*.

BCBS (2019). Minimum capital requirements for market risk. *Basel Committee on Banking Supervision*.

Beretta, L. and A. Santaniello (2016). Nearest neighbor imputation algorithms: a critical evaluation. *BMC Medical Informatics and Decision Making 16*(S3).

Bianchi, F., H. Mumtaz, and P. Surico (2009). The great moderation of the term structure of UK interest rates. *Journal of Monetary Economics 56*(6), 856–871.

BIS (2016). Minimum capital requirements for market risk. *Bank for International Settlements*.

Boyarchenko, N., R. K. Crump, A. Kovner, and O. Shachar (2022). Measuring corporate bond market dislocations. WorkingPaper 957, Federal Reserve Bank of New York.

Brunnermeier, M. K. and L. H. Pedersen (2009). Market liquidity and funding liquidity. *The review of financial studies 22*(6), 2201–2238.

Burgard, C. and M. Kjaer (2012). A generalised cva with funding and collateral. *SSRN Electronic Journal*.

Cai, C. X., X. Ye, and R. Zhao (2020). Informational friction, economic uncertainty and CDS-bond basis. *SSRN Electronic Journal*.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*(3), 591.

Christensen, J. H., F. X. Diebold, and G. D. Rudebusch (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics 164*(1), 4–20.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics 8*(2), 93–115.

De Pooter, M. (2007). Examining the nelson-siegel class of term structure models: In-sample fit versus out-of-sample forecasting performance. *Available at SSRN 992748*.

de Wit, J. (2006). Exploring the CDS-bond basis. *SSRN Electronic Journal*.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*(2), 337–364.

Diebold, F. X., C. Li, and V. Z. Yue (2008). Global yield curve dynamics and interactions: A dynamic nelson-siegel approach. *Journal of Econometrics 146*(2), 351–363.

Duffie, D. (1999). Credit swap valuation. *Financial Analysts Journal 55*(1), 73–87.

Dufour, J.-M. (1982). Generalized chow tests for structural change: A coordinate-free approach. *International Economic Review 23*(3), 565.

Dupire, B. (1994). Pricing with a smile. *Risk 7*(1), 18–20.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Financial Times (2011). US cds curve inverts for first time ever.

Fonseca, J. D. and K. Gottschalk (2013). A joint analysis of the term structure of credit default swap spreads and the implied volatility surface. *Journal of Futures Markets 33*(6), 494–517.

Green, A. (2015). *XVA: Credit, Funding and Capital Valuation Adjustments*. Wiley.

Gregory, J. (2020). *The xVA Challenge*. Wiley.

Hagan, P. S. and G. West (2006). Interpolation methods for curve construction. *Applied Mathematical Finance 13*(2), 89–129.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Houweling, P., A. Mentink, and T. Vorst (2005). Comparing possible proxies of corporate bond liquidity. *Journal of Banking & Finance 29*(6), 1331–1358.

Hull, J. C. and A. D. White (2000). Valuing credit default swaps I. *The Journal of Derivatives 8*(1), 29–40.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering 82*(1), 35–45.

Koopman, S. J., M. I. P. Mallee, and M. van der Wel (2010). Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters. *Journal of Business & Economic Statistics 28*(3), 329–343.

Lange, R.-J., A. Lucas, and A. Siegmann (2017). Score-driven systemic risk signaling for european sovereign bond yields and CDS spreads. pp. 129–150.

Lin, H., J. Wang, and C. Wu (2011). Liquidity risk and expected corporate bond returns☆. *Journal of Financial Economics 99*(3), 628–650.

Longstaff, F. A. and E. S. Schwartz (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. *The Journal of Finance 47*(4), 1259–1282.

Morini, M. and A. Prampolini (2010). Risky funding: A unified framework for counterparty and liquidity charges. *Risk Management eJournal*.

Morini, M. and A. Prampolini (2011). Risky funding: A unified framework for counterparty and liquidity charges. *Risk Magazine*.

Neftci, S. (2008). *Principles of Financial Engineering*. Elsevier.

Pallavicini, A., D. Perini, and D. Brigo (2011). Funding valuation adjustment: a consistent framework including CVA, DVA, collateral, netting rules and re-hypothecation.

Poirier, D. J. (1973). Piecewise regression using cubic splines. *Journal of the American Statistical Association 68*(343), 515–524.

Reuters (2011). United States loses prized AAA credit rating from S&P.

Risk (2024). Traders eye negative cds-bond basis.

Sarig, O. and A. Warga (1989). Bond price data and bond market liquidity. *The Journal of Financial and Quantitative Analysis 24*(3), 367.

Sommer, D., D. Todd, M. Peter, and H. Carstens (2013). FVA - putting into the equation.

Viceira, L. M. (2012). Bond risk, bond return volatility, and the term structure of interest rates. *International Journal of Forecasting 28*(1), 97–117.

Appendix

A Data

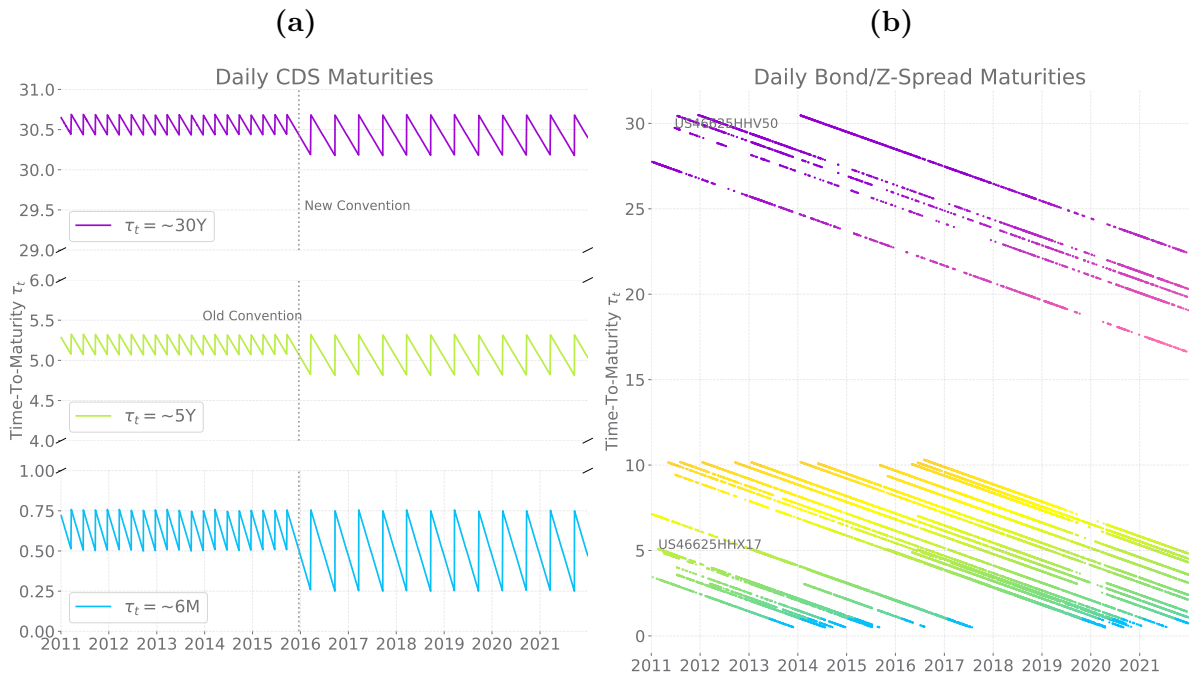
A.1 CDS and Z-Spread

Table A.1: JPM Bond ISIN codes

US46625HHF01	US46625HHN35	US46625HGY09	US46625HHX17	US46625HHW34	US46625HHZ64
US46625HHV50	US46625HHU77	US46625HHP82	US46625HHR49	US46625HJB78	US46625HJC51
US48126BAA17	US46625HJD35	US46623EJP51	US46625HJE18	US46625HJH49	US46625HJG65
US46625HJT86	US46625HJU59	US46623EJY68	US46625HJX98	US46625HMN79	US46625HKC33
US46625HNX43	US46625HQW33	US46625HQJ22	US46625HRL68	US46625HRS12	US46625HRT94
US46625HRV41					

Notes: This table reports 31 bond ISIN codes of the underlying Z-spreads of JPM used in this paper.

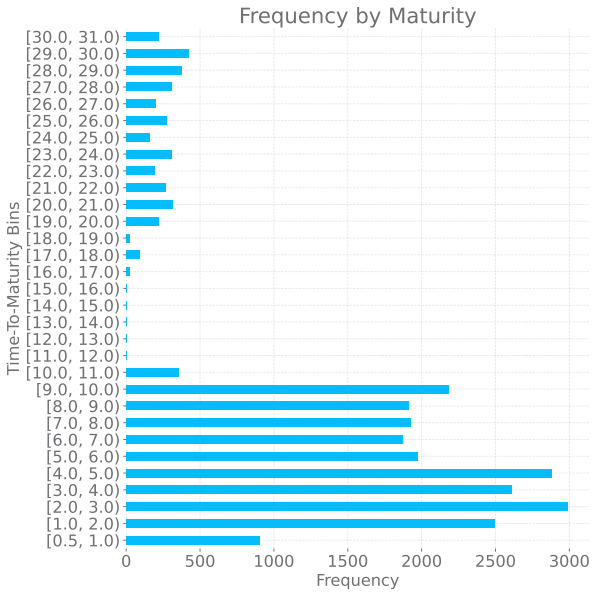
Figure A.1: Time-varying maturities of credit spread underliers



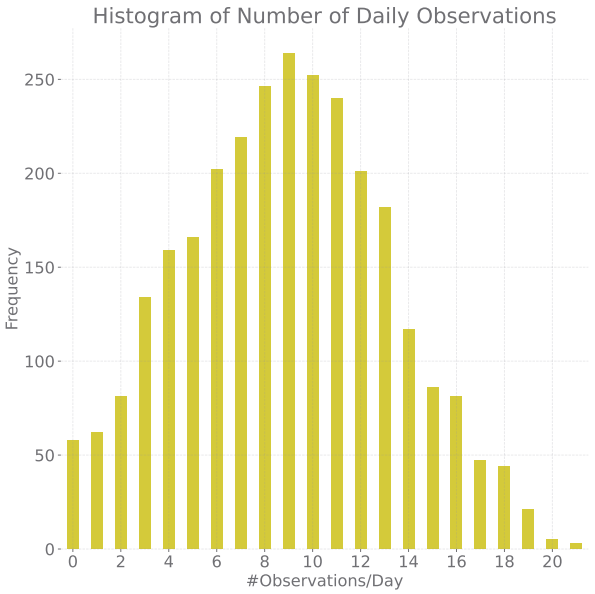
Notes: This figure display the deterministically time-varying maturities of underlying CDS and bond Z-spread instruments of JPM. The left panel displays the *actual* maturities of on-the-run CDS contracts with approximate Markit^{TM} tenors of '6M', '5Y' and '30Y'. The exact maturity computed based on the old and current IMM rollover dates used by ISDA, as described in this amendment: https://www.isda.org/a/vGiDE/amend-single-name-on-the-run-frequency-faq-revised-as-of-12-10.pdf. The right panel plots the running maturities of the 31 underlying bonds of JPM's Z-spreads based on a 30/360 day count convention.

Figure A.2: Z-spread frequency counts

(a) Frequency by maturity bins



(b) Histogram of #obs. per day



Notes: Subfigure A.2a of this figure displays frequency counts by maturity intervals of daily Z-spread observations for each bank. Subfigure A.2b displays the histogram of the number Z-spread data points observed per day.

A.2 Primary dealers list

Table A.2: Primary dealers and market capitalization

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
BNP Paribas	12.58	47.18	69.97	97.41	74.64	71.09	79.65	92.60	56.33	73.78	66.17	84.88
Barclays Bk plc	33.80	33.47	51.91	72.25	62.52	54.49	46.55	46.55	32.72	40.79	34.69	42.28
Bk Nova Scotia	59.54	54.24	68.41	75.41	69.49	48.51	66.97	77.43	61.30	68.28	65.37	85.38
Bk of America Corp	134.53	56.35	125.13	164.91	188.14	174.69	223.32	307.91	238.25	311.20	262.20	359.38
Citigroup Inc	137.44	76.92	119.82	157.85	163.62	152.83	169.35	196.74	123.30	168.89	128.37	119.82
Daiwa Secs Gp Inc	8.96	5.18	9.33	16.99	13.50	10.56	10.48	10.34	7.97	7.71	6.92	8.35
Deutsche Bk AG	47.79	34.28	40.49	48.80	41.89	33.95	25.14	39.15	16.46	16.00	22.72	26.13
Goldman Sachs Gp Inc	85.34	44.51	59.33	79.12	83.39	75.60	95.21	96.09	61.43	79.86	90.73	127.61
HSBC Hldgs plc	180.54	136.15	193.14	205.67	182.04	156.37	160.31	206.08	164.46	160.17	105.16	122.93
Mizuho Bk Ltd	40.51	32.02	43.96	52.57	41.42	50.25	45.75	45.96	39.13	39.19	32.00	32.22
Morgan Stanley	41.14	29.16	37.74	60.99	75.69	61.07	79.12	94.86	67.39	81.48	123.98	173.93
Nomura Hldgs Inc	22.75	10.97	21.66	28.64	21.05	20.32	21.06	20.20	12.61	15.99	16.15	13.14
UBS AG	62.45	44.32	58.98	71.84	62.72	74.07	58.09	68.14	45.84	45.62	51.21	61.06
Wells Fargo & Co	163.07	145.33	180.00	238.67	283.43	276.80	276.77	298.75	211.10	222.43	124.77	186.44

Notes: This table reports the primary dealers of the Federal Reserve Bank of New York (names are listed as presented Markit^{TM}'s CDS dataset and their end-of-year market capitalization (in billion USD). The CDS spreads of these firms is used to compute a daily CDS index based on a weighted average of CDS, with weights equal to their yearly market capitalization. The market capitalization data is sourced from https://companiesmarketcap.com/.

A.3 Correlations of basis determinants

Table A.3: Yearly correlations (%)

Pair		2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Avg.
ML	FC	-8.064	5.000	-3.608	2.041	1.543	-0.823	-4.395	-5.092	-2.445	-2.264	-7.856	-2.360
ML	FL	-0.538	3.720	5.194	4.353	6.890	-2.197	-4.465	6.221	15.889	25.794	1.206	5.642
ML	CR	11.962	-4.631	-0.443	-5.497	17.693	4.911	-0.356	-2.082	-2.893	1.182	5.865	2.337
ML	DP	0.165	0.066	7.653	-1.406	-1.078	-1.601	-4.217	7.670	10.666	30.137	-4.223	3.985
ML	IS	-1.394	-0.477	4.104	5.380	-2.749	12.015	5.111	7.426	-1.507	17.511	6.145	4.688
FC	FL	-13.790	-5.287	5.946	0.194	-5.057	-5.132	1.819	-17.287	-26.734	-34.550	-7.512	-9.763
FC	CR	-0.063	8.760	-10.032	-1.901	-1.986	13.058	4.167	-1.136	1.386	0.631	-8.095	0.435
FC	DP	-6.866	-0.399	-0.438	0.970	-4.245	11.316	0.601	4.923	16.762	1.390	-4.435	1.780
FC	IS	2.860	5.419	10.421	12.500	-4.240	7.087	-5.819	-6.830	2.599	14.438	-0.294	3.467
FL	CR	-5.656	16.515	-1.591	-0.219	5.417	4.497	-6.500	-11.992	-5.247	0.445	-1.611	-0.540
FL	DP	-7.263	0.602	-3.032	-4.973	3.418	-1.621	-2.352	4.060	-2.449	1.397	4.004	-0.746
FL	IS	-1.739	-12.857	4.625	-10.900	-6.078	1.090	-3.294	-8.244	3.509	8.592	-7.889	-3.017
CR	DP	-6.267	-9.990	1.168	4.057	1.956	1.473	-2.120	1.057	-4.472	1.210	-1.214	-1.195
CR	IS	6.589	3.522	-2.360	5.738	-3.184	-3.803	-3.173	6.073	12.166	7.573	6.827	3.270
DP	IS	-2.780	16.135	7.885	-6.491	3.423	10.284	-10.605	2.819	3.418	0.118	9.955	3.106

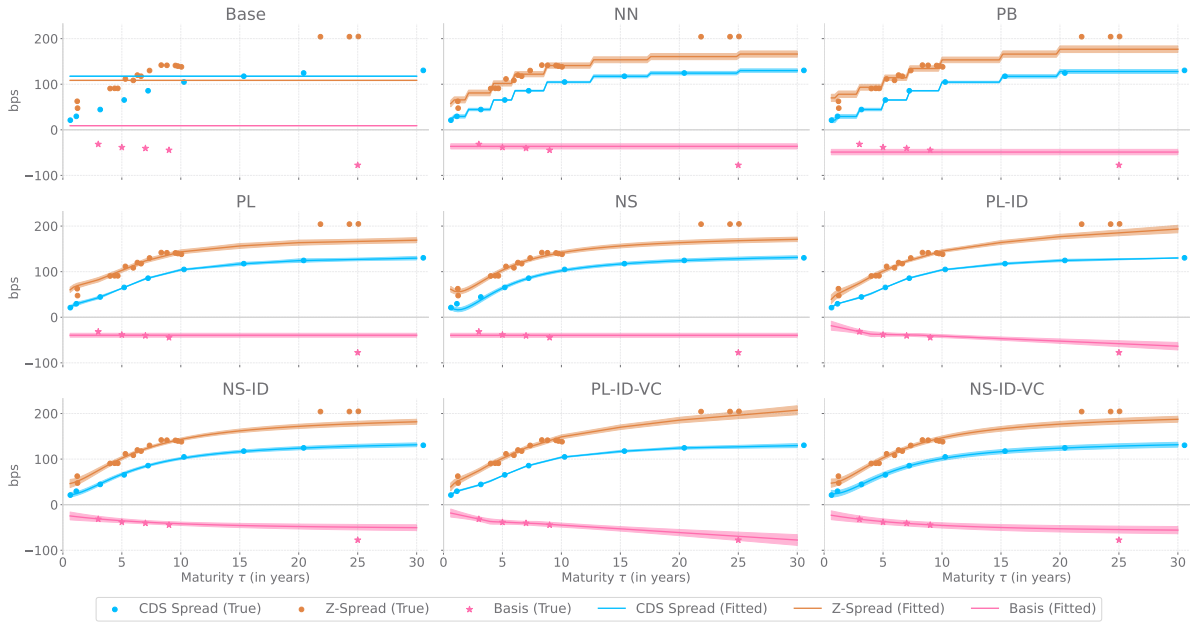
Notes: This table reports the correlations among the daily differenced explanatory variables in (12). The correlations are expressed in percentages and are reported for each year, separately, throughout our sample period.

B Dynamic basis model estimates

B.1 Single day illustration for all interpolation models

Figure B.1 shows the fitted basis term-structures on $t =$ Nov 3, 2016, for all models under consideration. Overall, models without curve-specific components exhibit is similar fitted curves and yield a comparable basis (around -50 bps), except for the `Base` model, which shows a slightly positive basis. This suggest a failure in capturing the basis when term-structures are not interpolated. Empirical basis estimates reveal a slightly downward-sloping term-structure, which is more accurately captured by models incorporating short-term basis dynamics. The `PL-ID(-VC)` model clearly sits closer to actual data points.

Figure B.1: CDS Spread and Z-Spread fit on $t = 2016\text{-}11\text{-}03$



Notes: This figure plots the observed and fitted CDS spread, Z-spread and empirical (non-parametric) basis observations (in bps) for the estimated models in Table 4 on $t = 2016\text{-}11\text{-}03$. The fitted curves are based on nowcasted state values. The the shaded areas (orange or blue) surrounding the fit represent 99% confidence intervals, whereas the raster (pink) corresponds to the model implied basis term-structure.

B.2 Model fit and confidence sets for JPM's CDS and Z-spread data

Table B.1: In-sample and out-of-sample model fit on CDS term-structure level for JPM

Model	6M	1Y	3Y	5Y	7Y	10Y	20Y	25Y	30Y	Avg.
Panel A: In-sample MAE										
Base	94.859	87.291	67.771	44.267	25.64	10.299	3.359	6.372	9.872	38.995
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]
NN	2.095	1.819	1.834	2.032⋆	2.361⋆	2.584⋆	3.195⋆	3.428⋆	3.576⋆	2.549
	[0.000]	[0.000]	[0.000]	[1.000]	[0.662]	[1.000]	[1.000]	[1.000]	[1.000]	[0.000]
PB	2.987	2.191	1.87	2.137	2.45	2.633	3.27	5.014	4.101	2.958
	[0.000]	[0.000]	[0.000]	[0.001]	[0.022]	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]
PL	2.086	2.031	1.866	2.041⋆	2.359⋆	2.59	3.21	3.437⋆	3.588⋆	2.58
	[0.000]	[0.000]	[0.000]	[0.172]	[1.000]	[0.018]	[0.003]	[0.157]	[0.119]	[0.000]
NS	1.955	9.825	10.335	4.755	2.639	3.81	3.615	4.041	4.956	5.117
	[0.02]	[0.000]	[0.000]	[0.000]	[0.022]	[0.000]	[0.002]	[0.001]	[0.000]	[0.000]
PL-ID	1.593	1.415⋆	1.659⋆	2.045⋆	2.364⋆	2.615	3.238	3.522	3.887	2.482⋆
	[0.02]	[1.000]	[1.000]	[0.142]	[0.662]	[0.000]	[0.003]	[0.001]	[0.000]	[0.091]
NS-ID	1.807	6.107	8.392	5.623	2.448⋆	4.012	3.4	3.737	4.793	4.495
	[0.000]	[0.000]	[0.000]	[0.000]	[0.147]	[0.000]	[0.002]	[0.001]	[0.000]	[0.000]
PL-ID-VC	1.557⋆	1.419⋆	1.666⋆	2.052⋆	2.398⋆	2.597	3.214	3.457⋆	3.782	2.461⋆
	[1.000]	[0.759]	[0.052]	[0.109]	[0.484]	[0.018]	[0.003]	[0.151]	[0.000]	[1.000]
NS-ID-VC	2.618	5.394	5.321	3.883	3.668	4.824	3.493	3.829	5.181	4.258
	[0.000]	[0.000]	[0.000]	[0.000]	[0.004]	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]
Panel B: Out-of-sample MAE										
Base	57.061	54.204	45.259	33.422	21.317	9.246	0.827⋆	6.112	12.427	26.653
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.971]	[0.000]	[0.000]	[0.000]
NN	0.554	0.751	0.584	0.546⋆	0.597⋆	0.707⋆	0.839⋆	1.017	0.967⋆	0.729⋆
	[0.000]	[0.000]	[0.000]	[1.000]	[0.657]	[0.27]	[0.461]	[0.001]	[1.000]	[0.053]
PB	1.668	1.507	0.816	0.675	0.692⋆	0.839	0.988	4.549	2.009	1.527
	[0.000]	[0.000]	[0.000]	[0.029]	[0.076]	[0.029]	[0.028]	[0.000]	[0.000]	[0.000]
PL	0.63	0.897	0.998	0.551⋆	0.591⋆	0.698⋆	0.825⋆	0.963⋆	0.967⋆	0.791
	[0.000]	[0.000]	[0.000]	[0.92]	[1.000]	[1.000]	[1.000]	[0.051]	[0.985]	[0.000]
NS	0.635	7.79	8.605	1.691	0.612⋆	0.815	1.081	1.018	1.941	2.687
	[0.000]	[0.000]	[0.000]	[0.000]	[0.228]	[0.000]	[0.000]	[0.012]	[0.000]	[0.000]
PL-ID	0.466	0.407⋆	0.452⋆	0.554⋆	0.613⋆	0.703⋆	0.829⋆	1.034	1.344	0.711⋆
	[0.000]	[0.056]	[0.727]	[0.729]	[0.328]	[0.59]	[0.971]	[0.001]	[0.000]	[0.264]
NS-ID	0.816	3.989	3.536	0.862	0.633⋆	0.836	1.109	0.912⋆	1.529	1.58
	[0.000]	[0.000]	[0.000]	[0.001]	[0.179]	[0.000]	[0.000]	[0.474]	[0.000]	[0.000]
PL-ID-VC	0.409⋆	0.398⋆	0.45⋆	0.55⋆	0.615⋆	0.723⋆	0.834⋆	0.968⋆	1.288	0.693⋆
	[1.000]	[1.000]	[1.000]	[0.92]	[0.228]	[0.101]	[0.937]	[0.067]	[0.000]	[1.000]
NS-ID-VC	1.724	3.738	3.51	0.834	0.814	0.854	1.254	0.893⋆	1.733	1.706
	[0.000]	[0.000]	[0.000]	[0.000]	[0.007]	[0.000]	[0.000]	[1.000]	[0.000]	[0.000]

Notes: This table contains the MAE (in bps) for the one-step ahead in-sample and out-of-sample CDS predictions for JPM across 9, near constant IIM-driven maturities, using the estimated models from Table 4 and the basis observations from Figure 1b. For the last column, we target the cross-sectional average of the basis to compute the MAE. Empty cells correspond to periods where we do not have basis observations due to missing data; see also Table 3 for an overview. Model confidence set p-values are reported in squared brackets beneath the MAEs. MAEs that are part of the 95% model confidence set are marked by ⋆.

Table B.2: In-sample and out-of-sample model fit on Z-spread term-structure level for JPM

Model	(6M, 1Y]	(1Y, 3Y]	(3Y, 5Y]	(5Y, 7Y]	(7Y, 10Y]	(10Y, 20Y]	(20Y, 30Y]	Avg.
Panel A: In-sample MAE								
Base	51.558	42.553	23.494	19.639	18.702	40.996	62.313	32.883
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.003]	[0.000]	[0.000]
NN	15.147	13.174	8.695	7.649	7.934	14.578	20.903	11.848
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.015]	[0.000]	[0.000]
PB	14.921	13.481	10.955	8.932	6.18	20.737	16.532	11.87
	[0.003]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
PL	16.742	12.578	8.425	6.496	5.39	14.649	20.581	11.031
	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.019]	[0.000]	[0.000]
NS	16.364	10.921	7.252	5.192⋆	4.635	12.393⋆	17.295	9.65
	[0.000]	[0.000]	[0.000]	[0.104]	[0.012]	[0.063]	[0.000]	[0.000]
PL-ID	8.751⋆	6.748	5.584⋆	4.759⋆	4.75	7.511⋆	7.162⋆	6.22⋆
	[1.000]	[0.006]	[0.902]	[0.672]	[0.000]	[1.000]	[1.000]	[1.000]
NS-ID	10.837	6.964	6.217	5.052⋆	4.466⋆	8.996⋆	12.494	7.439
	[0.019]	[0.000]	[0.002]	[0.271]	[1.000]	[0.228]	[0.000]	[0.000]
PL-ID-VC	9.786	6.267⋆	5.576⋆	4.73⋆	4.779	7.937⋆	7.384⋆	6.246⋆
	[0.019]	[1.000]	[1.000]	[1.000]	[0.000]	[0.405]	[0.133]	[0.595]
NS-ID-VC	10.707	6.374⋆	5.944	5.154⋆	4.608	8.617⋆	10.94	7.143
	[0.019]	[0.546]	[0.018]	[0.147]	[0.012]	[0.063]	[0.000]	[0.000]
Panel B: Out-of-sample MAE								
Base	5.561⋆	6.93⋆	29.476	34.725		111.35	107.08	45.677
	[0.192]	[1.000]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]
NN	11.415	11.933	7.588	4.641		44.049	37.172	17.825
	[0.000]	[0.000]	[0.006]	[0.004]		[0.000]	[0.000]	[0.000]
PB	10.457	10.966	10.76	3.925⋆		47.525	33.519	18.016
	[0.000]	[0.000]	[0.000]	[0.205]		[0.000]	[0.000]	[0.000]
PL	10.383	11.069	8.809	4.708		44.554	37.922	17.958
	[0.000]	[0.000]	[0.000]	[0.004]		[0.000]	[0.000]	[0.000]
NS	12.405	9.265	6.964	3.193⋆		38.49	31.807	15.224
	[0.000]	[0.001]	[0.044]	[1.000]		[0.000]	[0.000]	[0.000]
PL-ID	5.22⋆	7.753⋆	6.861	3.504⋆		11.871	4.819⋆	6.761
	[0.254]	[0.122]	[0.006]	[0.205]		[0.000]	[1.000]	[0.003]
NS-ID	4.716⋆	7.404⋆	6.76⋆	3.873		21.579	13.096	9.179
	[1.000]	[0.5]	[0.078]	[0.005]		[0.000]	[0.000]	[0.000]
PL-ID-VC	5.648⋆	7.678⋆	6.237⋆	4.304		7.973⋆	5.741	6.356⋆
	[0.113]	[0.173]	[1.000]	[0.004]		[1.000]	[0.03]	[1.000]
NS-ID-VC	4.955⋆	7.633⋆	6.481⋆	6.653		12.063	5.661⋆	7.245
	[0.41]	[0.352]	[0.123]	[0.000]		[0.000]	[0.236]	[0.001]

Notes: This table contains the MAE (in bps) for the one-step ahead in-sample and out-of-sample CDS predictions for JPM across 7 bond maturity intervals, using the estimated models from Table 4 and the basis observations from Figure 1b. For the last column, we target the cross-sectional average of the basis to compute the MAE. Empty cells correspond to periods where we do not have basis observations due to missing data; see also Table 3 for an overview. Model confidence set p-values are reported in squared brackets beneath the MAEs. MAEs that are part of the 95% model confidence set are marked by ⋆.

B.3 Additional results for the NS-ID-VC model

Figure B.2: Filtered time-varying factors of the `NS-ID-VC` model for JPM

Notes: the top panel of this figure present the time-varying filtered factors of the `NS-ID-VC` model for JPM with 95% confidence level. The bottom panel presents the ML estimated volatility term-structures for CDS spreads and Z-spreads with 95% confidence levels.

Table B.3: Parameter estimates of the NS-ID-VC model for JPM

	f_c	f_b	f_1^c	f_2^c	$f_3^{\text{idio.}}$
Panel A: State variables					
c	0.000	-0.189***	-0.012	-0.676***	0.123**
	(0.412)	(0.061)	(0.299)	(0.206)	(0.058)
$\text{diag}(T)$	0.999***	0.990***	0.999***	0.995***	0.991***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$\text{diag}(Q)$	8.270***	2.678***	5.542***	16.614***	3.018***
	(0.154)	(0.096)	(0.313)	(1.594)	(0.205)

$\tau =$	6M	1Y	5Y	10Y	30Y
Panel B: Variance knots					
$\sigma^2_{\text{CDS}}(\tau)$	13.943***	40.898***	15.969***	16.146***	23.517***
	(0.454)	(1.318)	(0.814)	(0.620)	(0.894)
$\sigma^2_{\text{Z}}(\tau)$	395.112***	54.114***	28.423***	26.297***	209.882***
	(19.938)	(0.768)	(0.418)	(0.555)	(5.127)

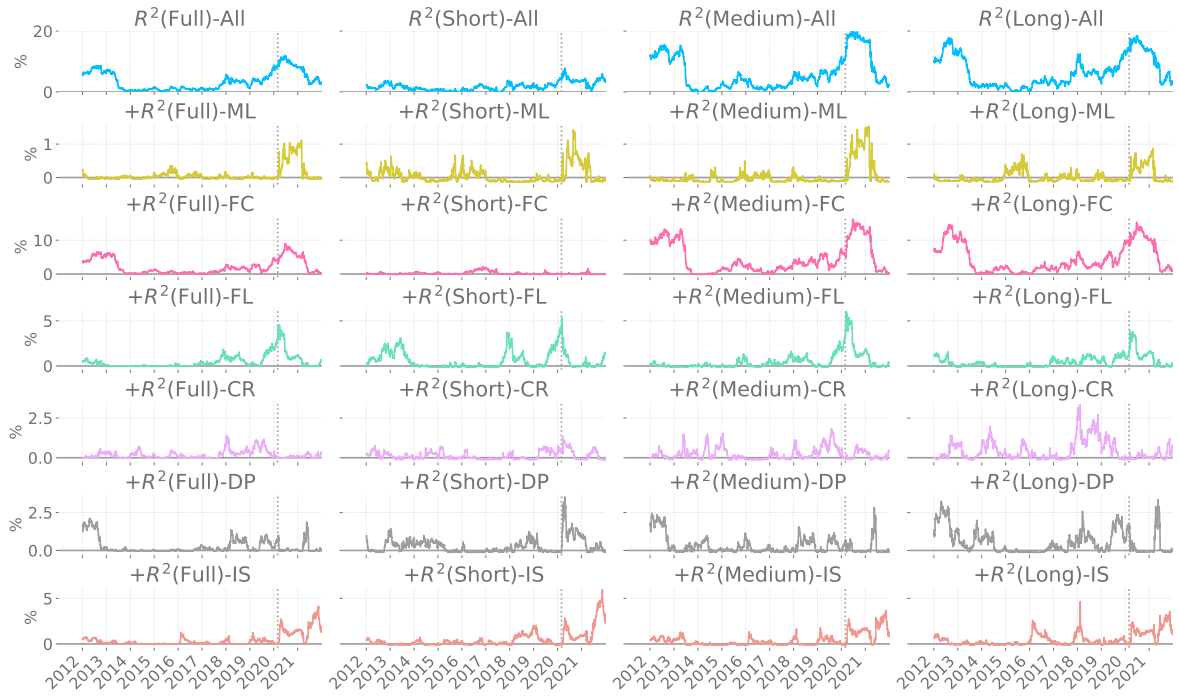
Panel C: Decay parameters	
λ^c	0.590***
	(0.003)
$\lambda^{\text{idio.}}$	0.225***
	(0.004)

Notes: This table reports the parameter estimates of the NS-ID-VC model for JPM. Model estimation was performed on credit spread data measured in basis points. Standard errors are presented in parentheses and the significance level of the coefficients at the 10%, 5% and 1% is indicated by '*', '**', '***', respectively.

C Additional basis explanatory results

C.1 Explanatory power on PL-ID-VC bases for JPM

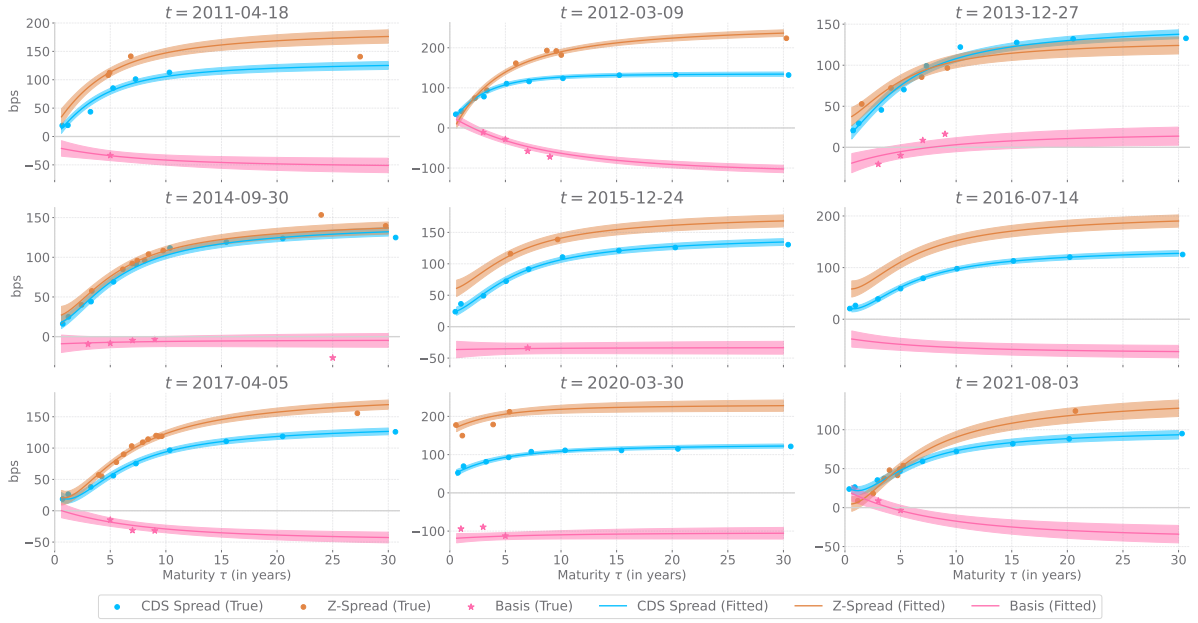
Figure C.1: Time-varying explanatory power per basis determinant



Notes: This figure plots the 261-day rolling-window adjusted R^2 estimates of diff-in-diff regression based on (12) for JPM with bases of `PL-ID-VC` model. Each column refers to a specific spectrum of basis maturities included in the dependant variable B. B(Full) stacks (includes) all maturities as dependant variable, whereas B(Short), B(Medium) and B(Long) only include the maturities [6M, 1Y, 3Y], [5M, 7Y, 10Y], and [15M, 20Y, 30Y], respectively. The explanatory variables (market liquidity, funding costs, funding liquidity, counterparty risk, default premium and idiosyncratic risk) are identical for all four regressions. The first row depicts the adjusted R^2 based on all regressors, whereas following rows present the individual contribution (increase/decrease) in adjusted R^2 via the leave-one-variable-out strategy.

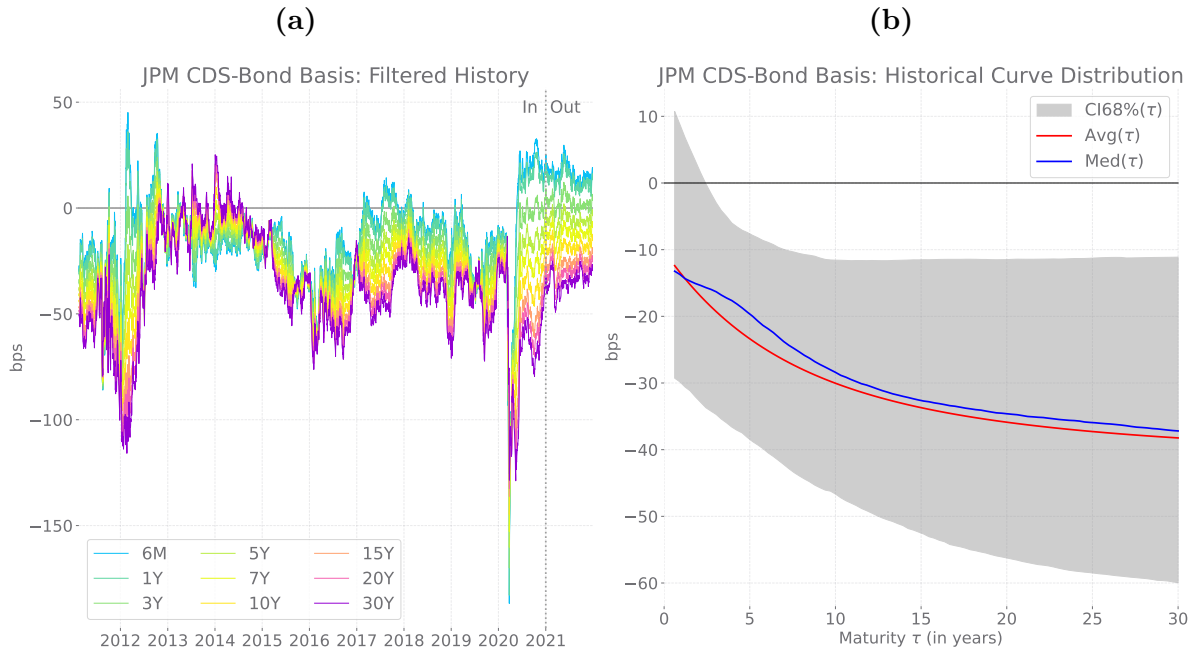
C.2 Robustness results for JPM based on NS-ID-VC model

Figure C.2: Fitted CDS-bond basis curve cross-sections of NS-ID-VC model



Notes: This figure displays fitted CDS, Z-spread and CDS-bond basis observations on 9 selected dates by the NS-ID-VC model for JPM. See caption of Figure B.1 for more details.

Figure C.3: Filtered constant maturity CDS-bond basis term-structure of JPM by NS-ID-VC model



Notes: Figure displays the C.3a constant maturity CDS-bond basis time-series (in bps) for JPM by the NS-ID-VC model. Figure C.3b displays curve distribution of the basis generated by the nowcasted time-series in Figure 4a.