HEC MONTRÉAL
École affliée à l'Université de Montréal

Three Essays on Credit Risk in the Bond Market

par
Delphine Boursicot

Thèse présentée en vue de l'obtention du grade de Ph. D.
en administration (Spécialisation Ingénierie financière)

May 2025

HEC MONTRÉAL

École affliée à l'Université de Montréal

Cette thèse intitulée:

Three Essays on Credit Risk in the Bond Market

Présentée par:

Delphine Boursicot

a été évaluée par un jury composé des personnes suivantes :

Michel Denault
HEC Montréal
Président-rapporteur

Geneviève Gauthier
HEC Montréal
Directrice de recherche

David Ardia
HEC Montréal
Membre du jury

Maher Kooli
Université du Québec à Montréal
Membre du jury

Cody Hyndman
Concordia University
Membre externe du comité conjoint

Nora Traum
HEC Montréal
Représentant du directeur de HEC Montréal

Résumé

Cette thèse se concentre sur des applications liées au risque de crédit sur le marché des titres à revenu fixe et explore la prévision de la structure par terme des écarts de crédit. Elle se divise en trois essais.

Dans le premier essai, les impacts de la dette Contingente Convertible (CoCo) sur les actionnaires sont étudiés. La dette contingente convertible est un instrument financier hybride émis par les banques, caractérisé par une conversion en actions en cas de détresse financière de l'établissement. Les CoCo comportent deux risques majeurs : le risque de défaut, inhérent à tout instrument de dette, et le risque spécifique de conversion obligatoire en actions. Dans cette étude, nous introduisons un modèle d'évaluation des instruments CoCo en fonction du ratio d'endettement de l'émetteur. Bien que les CoCo soient généralement plus coûteux que les instruments de dette traditionnels, leur présence dans la structure de capital de la banque réduit le coût de la dette ordinaire, diminuant ainsi le coût global du service de la dette. Pour les actionnaires initiaux, l'intégration des CoCo dans la structure de capital de la banque accroît la valeur totale des actions, ce qui a des implications bénéfiques pour la gestion du capital bancaire. Cet article a été publié dans *Risk, 2019, 7(2), 47*.

Le deuxième essai propose et évalue des méthodologies pour prévoir la structure par terme des écarts de crédit ainsi que les prix des obligations d'entreprise, en s'attaquant aux défis posés par les fréquences de négociation hétérogènes des obligations entre émetteurs. Pour les émetteurs à haute fréquence de négociation, nous implémentons un modèle en deux étapes de type Nelson–Siegel, couplé à un cadre autorégressif vectoriel. En raison d'une disponibilité de données plus limitée pour les émetteurs à faible fréquence, nous

développons une méthode de regroupement en deux étapes : d'abord, l'identification de groupes d'émetteurs à haute fréquence à l'aide d'un modèle de mélange gaussien ; ensuite, le rattachement de chaque émetteur à faible fréquence à ces groupes selon ses caractéristiques, puis le calcul d'une courbe d'écart de crédit moyennée d'après les probabilités d'appartenance aux clusters. L'analyse empirique, réalisée sur des obligations libellées en USD, montre des performances prédictives satisfaisantes, bien que la précision diminue pour les maturités longues et les émetteurs faiblement représentés.

Le dernier essai examine la modélisation prédictive de la structure par terme des écarts de crédit en intégrant des indicateurs d'incertitude macroéconomique et des méthodologies d'apprentissage automatique. Nous proposons une nouvelle fonction de perte pénalisée visant à améliorer l'interprétabilité économique et la stabilité temporelle des coefficients des écarts de crédit estimés selon le modèle Nelson-Siegel. Nous prédisons ensuite les coefficients de la structure par terme des écarts de crédit en étendant l'ensemble des variables traditionnellement utilisées dans la littérature par l'intégration de mesures d'incertitude macroéconomique et financière. Enfin, nous comparons des modèles de régression linéaire à des techniques d'apprentissage automatique basées sur des modèles d'ensemble. L'analyse empirique montre l'influence des variables spécifiques aux entreprises et renforce le lien entre l'incertitude économique et les spreads de crédit. Parmi les approches testées, le modèle des arbres renforcés (boosted trees) offre les meilleures performances pour prédire la structure par terme des écarts de crédit.

Mots clés: Dette contingente convertible; risque de crédit; écart de crédit; modélisation financière; gestion des risques; crise financière; analytique prédictive; apprentissage statistique.

Méthodes de recherche: Modèle d'ensemble; modèle de mélange gaussien; modélisation mathématique; optimisation numérique; régression; apprentissage statistique.

Abstract

This thesis concentrates on applications related to credit risk in the fixed-income market. It also explores the prediction of the credit-spread term structure. It is divided into three essays.

In the first essay, the impacts of Contingent Convertible (CoCo) debt on equity holders are investigated. Contingent Convertible debt is a hybrid financial instrument issued by banks, with a specific feature forcing its conversion to equity in the event of the bank's financial distress. CoCos incorporate two major risks: the default risk, inherent to all debt instruments, and the exclusive risk of mandatory conversion into equity. In this study, we introduce a valuation model for CoCo instruments as a function of the issuer's debt ratio. Although CoCos tend to be costlier than traditional debt instruments, their inclusion within the bank's capital structure reduces the cost associated with ordinary debt, thereby lowering the overall debt-servicing cost. For preliminary equity holders, the presence of CoCos in the banks' capital structure increases the shareholders' aggregate value, providing beneficial implications for the bank's capital-management strategy. This article has been published in *Risk, 2019, 7(2), 47*.

The second essay proposes and evaluates methodologies to forecast credit-spread term structures and corporate bond prices, by addressing challenges related to bond-trading frequencies across issuers. For high-frequency issuers, we implement a two-step Nelson–Siegel model coupled with a Vector-Autoregressive framework. Due to limited data availability for low-frequency issuers, we develop a two-step clustering approach: we first identify distinct clusters of high-frequency issuers using a Gaussian Mixture Model; we then link each low-frequency issuer to these clusters using issuer-specific information, and

we compute weighted-average credit-spread curves according to posterior cluster probabilities. Empirical analysis on USD-denominated corporate bond data demonstrates some predictive performance, although accuracy declines for longer maturities and sparsely populated issuer clusters.

The final essay investigates the predictive modelling of the credit-spread term structure by incorporating macroeconomic-uncertainty indicators and statistical-learning methodologies. We propose a new penalized-loss function designed to enhance the economic interpretability and temporal stability of the estimated Nelson-Siegel credit spread coefficients. We then predict these Nelson-Siegel credit spread coefficients by extending the set of features generally used in the literature to incorporate macroeconomic and financial-uncertainty measures. Finally, we compare linear regression models with ensemble based machine learning techniques. Empirical analyses show the influence of firm-level variables and reinforce the link between economic uncertainty and credit spreads. Boosted trees outperform other approaches to predict the term structure of credit spreads.

Keywords: Contingent convertible debt; credit risk, credit spread; financial modelling; risk management; financial crisis; predictive analytics; statistical learning.

Research methods: Ensemble model; Gaussian mixture model; mathematical modelling; numerical optimization; regression; statistical learning.

Contents

List of Figures

List of Tables

Abbreviations

ADS	**A**ruoba-**D**iebold-**S**cotti Business Conditions
AI	**A**ccured **I**nterest
AIC	**A**kaike **I**nformation **C**riterion
AR	**A**uto **R**regression
AT1	**A**dditional **T**ier **1**
BCBS	**B**asel **C**ommittee on **B**anking **S**upervision
BIC	**B**ayesian **I**nformation **C**riterion
CAR	**C**apital **A**dequacy **R**equirements
CET1	**C**ommon **E**quity **T**iers **1**
CPE	**C**onsumer **P**rice **I**ndex
CRD	**C**apital **R**equirements **D**irective
CRR	**C**apital **R**equirements **R**egulation
CoCo	**Co**ntingent **Co**nvertible
CRSP	**C**enter for **R**esearch in **S**ecurity **P**rices
EPU	**E**conomic **P**olicy **U**ncertainty
FINRA	**F**inancial **In**dustry **R**egulatory **A**uthority
FISD	**F**ixed **I**ncome **S**ecurities **D**atabase
FRED	**F**ederal **R**eserve **E**conomic **D**ata
FSL	**F**ederal and **S**tate/**L**ocal government purchases
GBM	**G**eometric **B**rownian **M**otion
GDP	**G**ross **D**omestic **P**roduct
GLS	**G**eneralized **L**east **S**quares

GMM	**G**aussian **M**ixture **M**odel
G-SIBs	**G**lobal- **S**ystemically **I**mportant **B**anks
IP	**I**ndustrial **P**roduction
IS	**I**n **S**ample
MAE	**M**ean **A**bsolute **E**rror
MSE	**M**ean **S**quared **E**rror
NBER	**N**ational **B**ureau of **E**conomic **R**esearch
NS	**N**elson-**S**iegel
OLS	**O**rdinary **L**east **S**quares
OS	**O**ut-of-**S**ample
OSFI	**O**ffice of the **S**uperintendent of **F**inancial **I**nstitutions
OTC	**O**ver-**T**he **C**ounter
RMSE	**R**oot **M**ean **S**quared **E**rror
RMSRE	**R**oot **M**ean **S**quared **R**elative **E**rror
ROE	**R**etrun-**o**n **E**quity
RWA	**R**isk-**W**eighted **A**sset
SHAP	**Sh**apley **A**dditive ex**P**lanations
SUR	**S**eemingly **U**nrelated **R**egression
TRACE	**T**rade **R**eporting **a**nd **C**ompliance **E**ngine
U.S.	**U**nited **S**tates
VAR	**V**ector **A**uto**r**egression
VIX	**V**olatility **I**ndex
XAI	**X**plainable **A**rtificial **I**ntelligence
YTM	**Y**ield-**T**o-**M**aturity
ZC	**Z**ero-**C**oupon

Acknowlegments

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Geneviève Gauthier, for her unwavering support, insightful guidance, thorough feedback, and remarkable patience throughout the entire research process. Her expertise and encouragement continually challenged me to push beyond my limits, and without her mentorship, kindness, and understanding, this thesis would not have reached its full potential.

I am also thankful to the members of my examining committee, who devoted their time to reading this work and provided insightful and suggestions that enhanced the quality of my work.

I also wish to thank Mohamed Jabir for his assistance and attentive listening throughout this journey. His understanding and support provided much-needed reassurance when the path felt uncertain.

My warmest thanks goes to my husband, Patrick. You stood by me, supported me through every stage of these doctoral years, and had faith in my abilities. You served as a vital source of strength and perseverance. I am deeply grateful for your patience and for always being there for me and our children when I was working. Your understanding and love have meant more to me than words can express.

I extend my deepest thanks to my schoolmates, whose camaraderie and stimulating discussions made the long hours of study far more rewarding, and helped shape this journey. A very special thank-you goes out to Ella, Saeed, Gabrielle, and Stephane my wonderful office mates. Your friendship has helped me overcome the many challenges of graduate life and I am grateful for it.

To my friends and family, you never stopped encouraging me, and for that I am deeply grateful for your support.

Finally, I wish to acknowledge my mother and grandmother. Their determination and unwavering strength they transmitted to me, empowered me the resilience to complete my doctorate, even when the odds were against me. With love.

This thesis is not simply my own achievement. It stands as a testament to the collective care, insight, and encouragement you have so generously provided. Thank you, from the bottom of my heart.

Chapter 1

Introduction

Banks face several distinct types of risks, including credit, market, liquidity, operational, and regulatory risks. This thesis explores the modelling and prediction of credit risk within the fixed-income market.

Risk management in the financial sector has evolved significantly over the decades, following major financial crises such as the 2008 global financial crisis, which triggered revisions in regulatory frameworks. The Basel Committee on Banking Supervision (BCBS) acts as a central coordinator for regulatory standards, notably through the Basel Accords. In response to the 2008 crisis, Basel III introduced stricter capital and liquidity requirements for banks and recommended the issuance of Contingent Convertible bonds (CoCos) as part of banks' Capital Adequacy Requirements (CAR). The first essay investigates the impact of Contingent Convertible bonds on equity holders.

The fixed-income market widely employs the term structure of credit spreads as a critical tool for quantifying credit risk. The term structure of credit spreads represents the yield spread between a risky bond and its risk-free benchmark across various maturities. The term structure of credit spreads emerges as the premium that investors require to compensate for credit risk. The last two essays of this thesis focus on developing predictive models of the term structure of credit spreads.

The first essay examines Contingent Convertible (CoCo) bonds and how t their inclusion in a bank's capital structure affects the value and strategic incentives of the equity

holders. CoCo bonds are hybrid financial instruments designed to convert into equity when a predefined trigger, linked to the issuer's financial health is breached. They absorb losses and act as automatic stabilizers by recapitalizing institutions during times of financial distress. CoCos carry two major risks: the risk of default, inherent to all debt instruments, and the unique risk of mandatory conversion. We propose a valuation model for CoCo instruments as a function of the issuer's debt ratio, and analyze how their introduction alters capital structure, debt composition, and dividend policy. While CoCos provide a buffer against default, conversion can inject new equity at potentially unfavourable terms. The presence of CoCos can alter the bank's risk profile from the perspective of equity holders. Anticipating the possibility of dilution, shareholders may adjust their investment and risk-taking behaviour. Although the CoCos are costlier instruments than traditional debt, their presence in the capital structure lower the cost of ordinary debt and reduce the total cost of debt. For preliminary equity holders, the presence of CoCos in the bank's capital structure increase the shareholder's aggregate value.

The second essay addresses the predictive challenges posed by issuer heterogeneity and disparities in bond-trading frequencies. We propose forecasting methodologies for credit spreads and bond prices that distinguish between frequently traded ("high-frequency") and infrequently traded ("low-frequency") issuers, based on a private database of USD-denominated corporate bond data. For high-frequency issuers, we implement a two-step Nelson–Siegel model combined with a Vector Autoregressive framework to predict credit spreads and bond prices. For low-frequency issuers, we develop a clustering approach based on Gaussian Mixture Models (GMM), which leverages issuer-specific attributes to infer credit-spread curves from comparable high-frequency issuers.

The third essay builds on the second and investigates the predictive modelling of the credit-spread term structure through the integration of macroeconomic-uncertainty indicators and statistical-learning approaches. Employing the Nelson-Siegel framework, we examine credit spreads using a comprehensive academic database of U.S. corporate bonds from July 2002 to June 2020, which encompasses periods of market stress and stability. To address known estimation challenges, we introduce a novel penalized-loss function that

enhances both the stability and economic interpretability of the estimated coefficients. We compare statistical-learning approaches and apply SHAP (Shapley Additive Explanations) values for interpretability. Our findings demonstrate the superior predictive performance of ensemble models relative to traditional linear models and provide valuable insights into the key variables influencing credit-spread dynamics.

Chapter 2

Contingent Convertible Debt: The Impact on Equity Holders

Abstract[1]

Contingent Convertible (CoCo) is a hybrid debt issued by banks with a specific feature forcing its conversion to equity in the event of the bank's financial distress. CoCo carries two major risks: the risk of default, which threatens any type of debt instrument, plus the exclusive risk of mandatory conversion. In this paper, we propose a model to value CoCo debt instruments as a function of the debt ratio. Although the CoCo is a more expensive instrument than traditional debt, its presence in the capital structure lowers the cost of ordinary debt and reduces the total cost of debt. For preliminary equity holders, the presence of CoCo in the bank's capital structure increases the shareholder's aggregate value.

[1] Joint work with Geneviève Gauthier and Farhad Pourkalbassi. Gauthier is affiliated with HEC Montréal.

2.1 Introduction

The 1988 Basel accord ties bank capitalization to portfolio risk by introducing the Capital Adequacy Requirements (CAR). Subsequently, Basel II obliges banks to maintain sufficient loss-absorbing capital on an annual basis. However, several studies on the 2008 financial crisis, such as Flannery, 2014 and Duffie, 2010 reveal that, in practice, regulators are unable to force banks to maintain adequate loss-absorbing capital. To alleviate banks undercapitalization, Flannery, 2005 introduced Contingent Convertibles, hereafter referred to as CoCo in accordance with most of the main-related studies. CoCo is a hybrid debt with a specific clause stipulating that the issuer would either convert it to equity or write down its face value if the level of loss-absorbing capital falls below a certain threshold. This is supposed to help with a firm's recapitalization under distress, while equity holders would be reluctant to raise capital voluntarily by issuing new shares. Basel III recommends large financial institutions to issue CoCo as a part of their Capital Adequacy Requirements.

CoCo carries two major risks: the risk of default, which threatens any type of debt instrument, plus the exclusive risk of mandatory conversion. CoCo differs greatly from a traditional convertible bond, which the conversion is optional and generally rewards the bondholder; hence, it is not a risk. Mandatory conversion of CoCo is a punishing mechanism which decreases the value of the bondholder in most scenarios; hence, it is a risk factor.

Most literature employs the structural approach to model CoCo dynamics. Studies generally define a trigger threshold as the barrier, then calculate the conditional probability of hitting the barrier. What makes this group of studies different is the choice of underlying instrument that triggers the conversion and the dynamics of the underlying trigger. The trigger is the book-equity-to-book-asset ratio in Glasserman and Nouri, 2012, where they model the book asset process using a Geometric Brownian motion (GBM). Conversion occurs if the book-equity-to-book-asset ratio exceeds a predetermined exogenous ratio. Chen et al., 2013 use the same underlying instrument while they model the book asset process using a jump-diffusion. Conversion is triggered if the asset value goes over a predetermined exogenous threshold. In Brigo et al., 2015, firm asset value is a GBM process and the conversion barrier is a linear function of the asset-to-equity ratio. De Spiegeleer et al., 2017,

in a more recent study, come closer to the Basel III accord and define the implied Common Equity Tier 1 (CET1) volatility as the conversion trigger while modelling the CET1 capital ratio as a GBM. Structural modelling based on a GBM makes the conversion time predictable, while being counterfactual. Uncertainty with respect to the conversion is due not only to the use of accounting capital ratio as the conversion trigger, but also to the stipulations in Basel III that allow regulators to choose the conversion time at their discretion. Studies such as Albul et al., 2010 and Hilscher and Raviv, 2014 model the CoCo as a contingent claim on the value of the bank's assets. The results are relatively tractable models in which the bank's incentive to issue CoCo voluntarily could be examined. Once again, the main challenge of these approaches is in triggering the conversion by the unobservable market value of assets.

In Cheridito and Xu, 2015, the CoCo price is modelled using a pure reduced form approach where the conversion and default events are modelled with a time-changed Poisson process. However, the reduced form approach is less intuitive by ignoring the link between the capital ratio and the trigger event (Brigo et al., 2015). Chung and Kwok, 2016 use a structural approach to model the conversion when the capital ratio falls below a certain threshold and also use a Poisson process to model the potential unexpected conversion imposed by the regulator.

Conversion price is also a matter of debate in the literature. As a basic design, Flannery, 2005 proposes that the number of shares received by CoCo holders at conversion is determined by the face value of the CoCo bonds divided by the market price of stock on the day of conversion. However, this basic conversion mechanism gives an opportunity to short sellers to bid down the share price, force conversion and dilute the market by increasing the number of outstanding shares. To avoid share price manipulation, Duffie, 2010 argues that the number of shares should be based on a multi-day average of closing prices. Other studies such as Flannery, 2016 and Pennacchi, 2010 propose converting the CoCo into a predetermined number of shares at a fixed price. However, there is a high risk that CoCo investors will suffer some value losses upon conversion due to a jump in the market price of shares.

Regulators insist that the CoCo conversion trigger should be based on the accounting

capital ratio. Indeed, this is determined by the Basel Committee on Banking Supervision (BCBS) at a global level in Basel III, the European Banking Authority through the Capital Requirements Directive IV/Capital Requirements Regulation (CRD IV/CRR) and the Office of the Superintendent of Financial Institutions (OSFI) in Canada through the CAR guidelines. The CET1 should not fall bellow a certain percent of Risk-Weighted Asset (RWA). Although accounting measures are not forward looking and can be manipulated by managers, they ensure that CoCo conversion occurs when a firm encounters serious financial problems. Glasserman and Nouri, 2012 and Chen et al., 2013 choose asset book value as the trigger through the book-equity-to-book-asset ratio, claiming that the book asset value truly approximates the market asset value. Many studies are against employing book value as a basis of conversion because it generates a delayed signal of financial distress. McDonald, 2013 and Bolton and Samama, 2012 assume that the market price of equity can measure a bank's loss-absorbing capacity. They propose a CoCo design in which the share price functions as the conversion trigger. Sundaresan and Wang, 2015 point out two shortcomings of employing the market price of shares as the conversion trigger within their framework, both of which are linked to the fact that CoCo conversion generates a value exchange between CoCo holders and initial equity holders. First, if the value is transferred from shareholders to CoCo investors at conversion, there can be more than one rational expected equilibrium price for both the stocks and CoCos. Second, if the value is transferred to shareholders, the model sometimes lacks an equilibrium share price. Sundaresan and Wang, 2015 conclude that a unique competitive equilibrium exists if the conversion does not induce a value transfer. Glasserman and Nouri, 2012 maintain that this multiple equilibrium problem is a feature of discrete-time models and can be alleviated in a continuous time framework. Closer to our paper, Chen et al., 2017 study the design and incentive effects of contingent convertible debt on optimal capital structure. The asset value is modelled as a jump diffusion, default and conversion occur whenever the firm value triggers some critical value. The default threshold is determined by the equity holder in such a way to maximize the equity value. They conclude that equity holders often have a positive incentive of issuing CoCos because the presence of CoCos reduces the bankruptcy costs when the conversion trigger is large enough.

In this paper, we also study the impact of having a contingent debt instrument in the debt structure in the perspective of equity holders. Based on a discrete time dynamic optimization approach, we reach similar conclusions as Chen et al., 2017. Indeed, our framework differs from the latter in various ways. First, the endogenous floating coupon rates of the standard and the CoCo debts account for the indebtedness of the firm. Second, the number of shares received by CoCo debt holders at conversion time is also designed differently. In our study, the optimal dividend stream is solved through a dynamic optimization approach: the equity holders have an incentive to control the firm debt ratio to maximize their share of the firm equity value, avoiding the dilution effect caused by conversion. It also benefits the bondholder as its mitigates the default risk.

A numerical simulation based on realistic data evaluates the benefits and the costs of having CoCos in the bank's debt structure. The parameters are estimated using three banks from three different regions (Europe, Canada, and the United States), for three different periods of time (pre-crisis, crisis, and post-crisis). The results help to understand how CoCos can help "Too big to fail" banks in different economic conditions. Although there are important differences among these three cases, common behaviours are observed: the presence of CoCos in the debt structure reduces the probability of default, the coupon of the standard debt, the cost of debt and capital. However, the CoCo is a more expensive instrument than the standard debt, mainly because the investor bears more risk. This study contributes to the literature by evaluating the effectiveness of adding CoCos to the financial firm's capital structure. We do not only evaluate the CoCo debt itself, but also examine its impact on the firm's management strategy by optimizing the per-share value of the cumulated dividend stream. Equity holders modify the optimal dividend policy to account for the conversion risk that affect them through the dilution effect (after conversion, there are more equity shares).

This paper is structured as follows. Section 2.2 presents the model to value CoCo and defines the conversion as well as the default intensity. Section 2.3 presents the dynamic programming model use to examine how CoCo impact the firm's management strategy. Section 2.4 is decomposed into three parts. First, the data used to have realistic scenarios are presented. Second, the results are presented for each bank and each year. Third, a

sensitivity analysis follows. Section 2.5 presents conclusions.

2.2 The Model

The asset value satisfies

$$A_t = E_t + D_t, \tag{2.1}$$

where E_t and D_t are the equity and debt values, respectively. The debt is decomposed into three main components: the deposit whose time t value is F_t, the coupon-paying bonds and other debt instruments for a value of B_t, and a convertible contingent instrument (CoCo) whose value is C_t:

$$D_t = F_t + B_t + C_t.$$

The debt ratio is defined as

$$X_t = \frac{D_t}{A_t}. \tag{2.2}$$

The presence of the CoCo debt alleviates the default risk since, in case of financial distress, the CoCo debt is converted into equity, leading to a smaller debt ratio. CoCo debt holders bear not only a default risk, but also a conversion risk; therefore, they require compensation in the form of a coupon payment that differs from that of a standard debt, which is subject only to credit risk. More precisely, once converted into equity, the CoCo debt has a zero recovery rate in case of default. However, if the firm survives after the conversion, then it is not clear whether the proportion of equity value held by the CoCo debt holders will be more or less profitable than for a standard debt.

The default time is denoted τ_D, and the conversion time satisfies $\tau_C = \min(\tau_\alpha, \tau_D)$, which means that the CoCo debt is always converted before the default event. Because other factors may influence the conversion decision, it does not necessarily occur as soon as the leverage ratio triggers α, but it is very likely to occur. Consequently, the conversion intensity driving τ_α,

$$G_t = g(X_t) = \left(\frac{X_t}{\theta_C}\right)^{\beta_C},$$

is a positive and convex function of X_t, taking large values whenever X_t is above $\theta_C \in (0, 1]$ and small values otherwise. θ_C should be close to α. The parameter β_C is usually a large

positive constant. Therefore,

$$\mathbb{P}_t\left(\tau_\alpha = t \,|\tau_\alpha > t-1, \tau_D > t-1\right) = \left[1-\exp\left(-G_t\right)\right]\mathbf{1}_{\tau_\alpha > t-1}\mathbf{1}_{\tau_D > t-1}.$$

Intuitively, the conversion should occur whenever the leverage ratio is close to the critical level α determined by the regulator. If the CoCo debt still exists at time t, then the superscript t^- denotes the pre-conversion values of the considered variables. The convertible contingent instrument is a coupon-paying bond with the floating coupon rate $c\left(X_t\right)$. At conversion, CoCo debt holders receive, in the form of equity shares, an amount equivalent to the debt's face value $C_{\tau_C^-}$ and a fraction ρ_C of the coupon. In other words, at conversion, the convertible debt holders receive

$$N_{\tau_C} = \frac{C_{\tau_C^-}\left(1+\rho_C c\left(X_{\tau_C - 1}\right)\right)}{S_{\tau_C}}$$

equity shares, where S_{τ_C} is the post-conversion price per share. Assuming that initially there were N outstanding equity shares, the post-conversion price per share becomes

$$S_{\tau_C} = \frac{E_{\tau_C}}{N + N_{\tau_C}} = \frac{E_{\tau_C^-} + C_{\tau_C^-}\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right)}{N + \frac{C_{\tau_C^-}\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right)}{S_{\tau_C}}}.$$

Multiplying both sides by $N + \frac{C_{\tau_C^-}\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right)}{S_{\tau_C^-}}$ and isolating S_{τ_C} leads to

$$S_{\tau_C} = \frac{E_{\tau_C^-}}{N} = S_{\tau_C^-},$$

which implies that the price per share is not affected by the conversion. Then, the equity value becomes

$$E_{\tau_C} = E_{\tau_C^-} + C_{\tau_C^-}\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right).$$

Letting

$$y_t = \frac{C_t}{D_t} \in [0,\gamma] \subseteq [0,1] \tag{2.3}$$

be the proportion of convertible debt in the total debt value. Because there are always deposits, then the convertible instrument represents less than 100% of the debt value, which implies that $\gamma < 1$. The number of additional shares issued at conversion satisfies

$$N_{\tau_C} = N\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right) y_{\tau_C^-} \frac{X_{\tau_C^-}}{1-X_{\tau_C^-}} \cong N\left(1+\rho_C c\left(\alpha\right)\right) y_{\tau_C^-}\left(\frac{\alpha}{1-\alpha}\right). \tag{2.4}$$

Proof.

$$\begin{aligned}
N_{\tau_C} &= \frac{C_{\tau_C^-}\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right)}{S_{\tau_C}} \\
&= N\frac{C_{\tau_C^-}\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right)}{E_{\tau_C^-}} \\
&= N\frac{C_{\tau_C^-}\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right)}{D_{\tau_C^-}}\frac{D_{\tau_C^-}}{A_{\tau_C^-}}\frac{A_{\tau_C^-}}{E_{\tau_C^-}} \\
&= N\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right)\frac{C_{\tau_C^-}}{D_{\tau_C^-}}\frac{D_{\tau_C^-}}{A_{\tau_C^-}}\frac{1}{1-\frac{D_{\tau_C^-}}{A_{\tau_C^-}}} \\
&= N\left(1+\rho_C c\left(X_{\tau_C-1}\right)\right)y_{\tau_C^-}\frac{X_{\tau_C^-}}{1-X_{\tau_C^-}}.
\end{aligned}$$

□

The standard debt pays the floating coupon rate of $b\left(X_t, y_t\right)$ unless default occurs. Interestingly, the debt coupon is affected by the presence of the CoCo instrument because the latter mitigates both the default and recovery risks. If

$$1-z_t=\frac{F_t}{F_t+B_t}=\frac{F_t}{D_t-C_t}=\frac{1}{1-y_t}\frac{F_t}{D_t}$$

denotes the proportion of the non-convertible contingent debt value that consists of deposits, then

$$F_t=\left(1-z_t\right)\left(1-y_t\right)D_t,\ B_t=z_t\left(1-y_t\right)D_t \text{ and } C_t=y_tD_t. \tag{2.5}$$

2.2.1 The Debt Ratio $\mathbb{P}-$Dynamics

Assume that the firm survives up to time $t:\tau_D>t$. At the beginning of the $t+1^{\text{th}}$ period, the invested capital yields returns:

$$R_{t+1}=m_{t+1}+\sigma_{t+1}\varepsilon_{t+1}^{\mathbb{P}},$$

where m and σ^2 are predictable processes and the sequence of ε_t is formed with independent standard normal random variables. The information structure is provided by the

filtration $\{\mathcal{F}_t\}_{t=1}^{\infty}$, where $\mathcal{F}_t = \sigma\left(\left\{\varepsilon_u^{\mathbb{P}}\right\}_{u=1}^{t}\right)$ is the $\sigma-$field modelling the information available at time t. Throughout the period, decisions about the convertible contingent debt conversion and the dividend payment affect the debt and debt ratio values. At time $t+1$, free cash flow (FCF) is

$$FCF_{t+1} = A_t R_{t+1}.$$

The financial flow (FF) is decomposed into the deposit interest payment $r_{t+1}F_t$, where r_{t+1} is the predictable risk-free rate, the interest payment on standard and convertible debts, $b\left(X_t, y_t\right) B_t$ and $c\left(X_t\right) C_t$ respectively, the dividend payment $\delta_{t+1}A_{t+1}$, and the debt structure variation,

$$D_{t+1} - D_t = \left(F_{t+1} - F_t\right) + \left(B_{t+1} - B_t\right) + \left(C_{t+1} - C_t\right)\mathbf{1}_{\tau_C > t+1} - C_t\mathbf{1}_{\tau_C = t+1}.$$

The floating interest rates reflect the risk embedded in both standard and CoCo debts. Both types of debts contain credit risk because the firm may default. See Appendix A.1.

The debt structure variation is expressed as a ratio:

$$\eta_{t+1} = \frac{D_{t+1} - D_t}{D_t}.$$

New debt issuance makes $\eta_{t+1} > 0$, whereas debt reaching maturity or CoCo debt conversion leads to $\eta_{t+1} < 0$.

Assumption 1. *Let $y_t = y_0\mathbf{1}_{\tau_C > t}$, where $y_0 = C_0/D_0$. In other words, whenever the contingent convertible debt exists, its proportion of the total debt value remains constant. Similarly, assuming that the proportion of deposits with respect to the non-convertible debt value remains constant over time, $(1 - z_t) = (1 - z_0) = F_0/(F_0 + B_0)$ and $z_0 = B_0/(F_0 + B_0)$.*

Assumption 2. *At conversion time, if there is no variation in the other types of debt, then $\eta_{\tau_C} = -y_{t_C - 1}$. In the numerical implementation, we assume that $\eta_{t+1} = -y_t\mathbf{1}_{\tau_C = t+1}$. In other words, the debt value at conversion is modified only by the conversion of the CoCo debt to equity.*

Thus, the weighted average interest rate is

$$\mu_{t+1} = \mu_{t+1}\left(X_t, z_t, y_t, \mathbf{1}_{\tau_C > t+1}\right)$$

$$= \left(r_{t+1}F_t + b\left(X_t, y_t\right)B_t + c\left(X_t\right)C_t\mathbf{1}_{\tau_C \geq t+1}\right)/D_t$$
$$= r_{t+1}\left(1-z_t\right)\left(1-y_t\right) + b\left(X_t, y_t\right)z_t\left(1-y_t\right) + c\left(X_t\right)y_t\mathbf{1}_{\tau_C > t+1}. \tag{2.6}$$

Since

$$D_{t+1} = D_t\left(1+\eta_{t+1}\right), \tag{2.7}$$

the financial flow satisfies

$$\begin{aligned} FF_{t+1} &= \mu_{t+1}D_t + D_t - D_{t+1} + \delta_{t+1}A_{t+1} \\ &= \left(\mu_{t+1} - \eta_{t+1}\right)D_t + \delta_{t+1}A_{t+1}. \end{aligned}$$

The bank's profit is the difference $FCF - FF$ between the free cash flow and the financial flow. Therefore,

$$\begin{aligned} A_{t+1} &= A_t + FCF_{t+1} - FF_{t+1} \\ &= \left(1+R_{t+1}\right)A_t - \left(\mu_{t+1} - \eta_{t+1}\right)D_t - \delta_{t+1}A_{t+1}, \end{aligned}$$

which is equivalent to

$$A_{t+1} = \frac{\left(1+R_{t+1}\right)A_t - \left(\mu_{t+1} - \eta_{t+1}\right)D_t}{1+\delta_{t+1}}. \tag{2.8}$$

Dividing both sides of Equation (2.8) by A_t,

$$\frac{A_{t+1}}{A_t} = \frac{\left(1+R_{t+1}\right) - \left(\mu_{t+1} - \eta_{t+1}\right)X_t}{1+\delta_{t+1}}.$$

Since

$$\frac{A_{t+1}}{A_t} = \frac{A_{t+1}}{D_{t+1}}\frac{D_{t+1}}{D_t}\frac{D_t}{A_t} = \frac{1}{X_{t+1}}\left(1+\eta_{t+1}\right)X_t,$$

comparing both equations and isolating X_{t+1} implies that the (post-dividend) debt ratio must satisfy

$$X_{t+1} = \frac{\left(1+\delta_{t+1}\right)\left(1+\eta_{t+1}\right)X_t}{\left(1+R_{t+1}\right) - \left(\mu_{t+1} - \eta_{t+1}\right)X_t}. \tag{2.9}$$

Note that

$$X_{t+1} = \left(1+\delta_{t+1}\right)X^0_{t+1}, \tag{2.10}$$

where X^0_{t+1} denotes the pre-dividend debt ratio

$$X^0_{t+1} = X_{t+1}|_{\delta_{t+1}=0} = \frac{\left(1+\eta_{t+1}\right)X_t}{\left(1+R_{t+1}\right) - \left(\mu_{t+1} - \eta_{t+1}\right)X_t}.$$

2.2.2 Conversion

The conversion decision is taken under the assumption that $\delta_{t+1} = 0$, $\eta_{t+1} = 0$ (no debt issuing or refunding), and that the full interest payment includes the CoCo debt coupon:

$$\mu_{t+1}^{C} = r_{t+1}\left(1 - z_t\right)\left(1 - y_t\right) + b\left(X_t, y_t\right) z_t \left(1 - y_t\right) + c\left(X_t\right) y_t. \tag{2.11}$$

The conversion intensity

$$G_{t+1} = g\left(X_{t+1}^{C}\right) = \left(\frac{X_{t+1}^{C}}{\theta_C}\right)^{\beta_C} \tag{2.12}$$

is a positive, increasing, and convex function of the debt ratio

$$X_{t+1}^{C} = \frac{X_t}{\left(1 + R_{t+1}\right) - \mu_{t+1}^{C} X_t}, \tag{2.13}$$

which is a particular case of Equation (2.9). The parameter θ_C is the critical level from which the conversion probability grows fast beyond this threshold. Because of the convex relation between the conversion intensity and the conversion probability, θ_C is not exactly equal to the regulator critical level α, but it is in its neighborhood. The parameter β_C captures the growth speed. In the numerical implementation, both parameters are obtained through a calibration approach based on the regulator critical level α. For this reason, τ_α denotes the conversion time characterized by the conversion intensity α. The conditional conversion probability at time $t+1$ triggered by the critical level α (letting $\beta_C \to \infty$, we recover the case where conversion occurs as soon as $X_{t^-} > \alpha$) is

$$P_{t+1}\left(\tau_\alpha = t + 1 \,|\tau_C > t\right) = \left[1 - \exp\left(-G_{t+1}\right)\right] \mathbf{1}_{\tau_C > t}.$$

Because immediate conversion may also arise as a consequence of an unpredicted default, the survival conversion probability is

$$\begin{aligned}
\mathbb{P}_{t+1}\left(\tau_C > t + 1 \,|\tau_C > t\right) &= \mathbb{P}_{t+1}\left(\tau_\alpha > t + 1, \tau_D > t + 1 \,|\tau_C > t\right) \\
&= \mathbb{P}_{t+1}\left(\tau_\alpha > t + 1 \,|\tau_C > t\right) \mathbb{P}_{t+1}\left(\tau_D > t + 1 \,|\tau_C > t\right) \\
&= \exp\left(-H_{t+1}\mathbf{1}_{\tau_C > t} - G_{t+1}\right).
\end{aligned}$$

Then, the conditional conversion probability arising from both the critical level and the potential default is

$$\mathbb{P}_{t+1}\left(\tau_C = t+1 \,|\tau_C > t\right) = \left[1 - \exp\left(-H_{t+1}\mathbf{1}_{\tau_C>t} - G_{t+1}\right)\right]\mathbf{1}_{\tau_C>t}, \tag{2.14}$$

where the default intensity H_{t+1} is defined at Equation (2.17).

2.2.3 Default

Since default occurs after conversion, the interest payment does not include the CoCo debt coupon. Indeed, a fraction of the CoCo coupon is paid back to CoCo debt holders in the form of equity shares. This means that

$$\mu^D_{t+1} = r_{t+1}\left(1 - z_t\right)\left(1 - y_t\right) + b\left(X_t, y_t\right) z_t \left(1 - y_t\right). \tag{2.15}$$

The default intensity is based on the pre-dividend debt ratio. More precisely, assuming that there are no dividends, $\delta_{t+1} = 0$, and that the only debt variation considered is the one arising from an immediate conversion, $\eta_{t+1} = -y_t\mathbf{1}_{\tau_C=t+1}$, the pre-dividend debt ratio is

$$X^D_{t+1} = \frac{\left(1 - y_t\right)X_t}{\left(1 + R_{t+1}\right) - \left(\mu^D_{t+1} + y_t\right)X_t}\mathbf{1}_{\tau_\alpha>t} + \frac{X_t}{\left(1 + R_{t+1}\right) - \mu^D_{t+1}X_t}\mathbf{1}_{\tau_\alpha\leq t}. \tag{2.16}$$

Since a debt ratio augmentation has more impact on the default probability when the debt ratio is already high, the default intensity H_{t+1} is a positive, increasing, and convex function of the debt ratio. More precisely,

$$H_{t+1} = h\left(X^D_{t+1}, y_t\right) = \lambda_D + \left(\frac{X^D_{t+1}}{\theta_{D,t}}\right)^{\beta_D}, \tag{2.17}$$

where $\theta_{D,t} = \theta_D + y_t$, $\lambda_D \geq 0, \theta_D > \alpha, \beta_D > 1$. Indeed, $\theta_{D,t}$ represents the critical debt ratio from which the increasing behaviour of the default probability (seen as a function of the debt ratio) accelerates. Because the CoCo instrument provides the standard debt holders with an additional protection against default risk, the critical debt ratio is $\theta_{D,t}$ slightly higher whenever the CoCo debt is present in the debt structure. Consequently, the conditional default probabilities are

$$\mathbb{P}_{t+1}\left(\tau_D = t+1 \,|\tau_C > t, \tau_D > t\right) = \left[1 - \exp\left(-H_{t+1}\mathbf{1}_{\tau_C>t}\right)\right]\mathbf{1}_{\tau_D>t}, \tag{2.18}$$

and

$$\mathbb{P}_{t+1}\left(\tau_D = t+1 \,|\tau_C \leq t, \tau_D > t\right) = \left[1 - \exp\left(-H_{t+1}\mathbf{1}_{\tau_C \leq t}\right)\right]\mathbf{1}_{\tau_D > t}. \tag{2.19}$$

2.3 Stochastic Optimum Control Problem

The period $]t, t+s]$ discount factor is

$$DF_{t,t+s} = \prod_{u=t}^{t+s-1} \frac{1}{1 + w\left(X_u\right)},$$

where the cost of capital is a weighted average of the cost of equity, r^E, and the cost of debt, μ:

$$w\left(X_u\right) = \left(1 - X_u\right) r_u^E + X_u \mu_u. \tag{2.20}$$

The current equity holders want to maximize their share of dividends. More precisely, given a stream of dividend rates $\delta_{1:\infty} = \{\delta_i\}_{i=1}^{\infty}$, the expected value of the discounted dividends at time t is

$$\begin{aligned} V\left(t, X_t^0, \delta_{t:\infty}\right)\mathbf{1}_{\tau_D > t} &= \sum_{u=t}^{\infty} \mathbb{E}_t^{\mathbb{P}}\left[DF_{t,u}\delta_u A_u\left(\mathbf{1}_{\tau_C > u} + \frac{N}{N + N_{\tau_C}}\mathbf{1}_{\tau_C \leq u}\right)\mathbf{1}_{\tau_D > u}\right]\mathbf{1}_{\tau_D > t} \\ &\cong \sum_{u=t}^{\infty} \mathbb{E}_t^{\mathbb{P}}\left[DF_{t,u}\delta_u A_u\left(1 - \frac{\left(1 + \rho_C c\left(\alpha\right)\right) y_0 \alpha}{1 - \alpha + \left(1 + \rho_C c\left(\alpha\right)\right) y_0 \alpha}\mathbf{1}_{\tau_C \leq u}\right)\mathbf{1}_{\tau_D > u}\right]\mathbf{1}_{\tau_D > t}, \end{aligned}$$

since Assumption 1 and Equation (2.4) imply that

$$\begin{aligned} \mathbf{1}_{\tau_C > u} + \frac{N}{N + N_{\tau_C}}\mathbf{1}_{\tau_C \leq u} &= 1 - \frac{N_{\tau_C}}{N + N_{\tau_C}}\mathbf{1}_{\tau_C \leq u} \\ &\cong 1 - \frac{\left(1 + \rho_C c\left(\alpha\right)\right) y_0 \alpha}{1 - \alpha + \left(1 + \rho_C c\left(\alpha\right)\right) y_0 \alpha}\mathbf{1}_{\tau_C \leq u}. \end{aligned}$$

Since V allows for the decomposition

$$\begin{aligned} V\left(t, X_t^0, \delta_{t:\infty}\right)\mathbf{1}_{\tau_D > t} \cong \delta_t A_t &\left(1 - \frac{\left(1 + \rho_C c\left(\alpha\right)\right) y_0 \alpha}{1 - \alpha + \left(1 + \rho_C c\left(\alpha\right)\right) y_0 \alpha}\mathbf{1}_{\tau_C \leq t}\right)\mathbf{1}_{\tau_D > t} \\ &+ \mathbb{E}_t^{\mathbb{P}}\left[DF_{t,t+1} V\left(t+1, X_{t+1}^0, \delta_{t+1:\infty}\right)\mathbf{1}_{\tau_D > t+1}\right]\mathbf{1}_{\tau_D > t}, \end{aligned} \tag{2.21}$$

the optimal dividend rate sequence $\delta_{1:\infty}^{\text{opt}}$ can be constructed recursively using backward recursion over time:

$$\begin{aligned}\delta_t^{\text{opt}} \;=\; \arg\max_{\delta_t\in[0,\delta_t^{\max}]} \Bigg\{ & A_t\delta_t\left(1-\frac{\left(1+\rho_C c\left(\alpha\right)\right)y_0\alpha}{1-\alpha+\left(1+\rho_C c\left(\alpha\right)\right)y_0\alpha}\mathbf{1}_{\tau_C\leq t}\right) \\ & +\mathbb{E}_t^{\mathbb{P}}\left[\frac{V\left(t+1,X_{t+1}^0,\delta_{t+1:\infty}^{\text{opt}}\right)}{1+w\left(X_t\right)}\mathbf{1}_{\tau_D>t+1}\right]\Bigg\}\mathbf{1}_{\tau_D>t}. \end{aligned} \tag{2.22}$$

See Appendix A.4. Because $\frac{A_{t+1}}{A_t}=\frac{D_t}{A_t}\frac{D_{t+1}}{D_t}\frac{A_{t+1}}{D_{t+1}}=\frac{X_t}{X_{t+1}}\left(1+\eta_t\right)$,

$$\begin{aligned} v\left(t,X_t^0,\delta_{t:\infty}\right)\mathbf{1}_{\tau_D>t} &= \frac{V\left(t,X_t^0,\delta_{t:\infty}\right)}{A_t}\mathbf{1}_{\tau_D>t} \\ &\cong \delta_t\left(1-\frac{\left(1+\rho_C c\left(\alpha\right)\right)y_0\alpha}{1-\alpha+\left(1+\rho_C c\left(\alpha\right)\right)y_0\alpha}\mathbf{1}_{\tau_C\leq t}\right)\mathbf{1}_{\tau_D>t} \\ &\quad+\mathbb{E}_t^{\mathbb{P}}\left[DF_{t,t+1}\frac{X_t}{X_{t+1}}\left(1+\eta_t\right)v\left(t+1,X_{t+1}^0,\delta_{t+1:\infty}\right)\mathbf{1}_{\tau_D>t+1}\right]\mathbf{1}_{\tau_D>t}, \end{aligned} \tag{2.23}$$

and

$$\begin{aligned}\delta_t^{\text{opt}} = \arg\max_{\delta_t\in[0,\delta_t^{\max}]} \Bigg\{ & \delta_t\left(1-\frac{\left(1+\rho_C c\left(\alpha\right)\right)y_0\alpha}{1-\alpha+\left(1+\rho_C c\left(\alpha\right)\right)y_0\alpha}\mathbf{1}_{\tau_C\leq t}\right) \\ & +\mathbb{E}_t^{\mathbb{P}}\left[\frac{X_t}{X_{t+1}}\left(1+\eta_t\right)\frac{v\left(t+1,X_{t+1}^0,\delta_{t+1:\infty}^{\text{opt}}\right)}{1+w\left(X_t\right)}\mathbf{1}_{\tau_D>t+1}\right]\Bigg\}\mathbf{1}_{\tau_D>t}. \end{aligned}$$

Therefore, we work with a standardized version of the primary equity holders' share of cumulated discounted dividends. Indeed, under this form, the dynamic optimization does not require the modelling of the dynamics of A.

The dividend rate is bounded above. Indeed, if the dividend payment is too large, the equity value will drop below its current level. More precisely, noting that $X_t=\left(1+\delta_t\right)X_t^0$, the expected variation of the pre-dividend equity is (see Appendix A.2.3)

$$\mathbb{E}_t^{\mathbb{P}}\left[E_{t+1}|_{\delta_{t+1}=0}-E_t\right]\leq\left(m_{t+1}-\left(1+\delta_t\right)X_t^0\mu_{t+1}\left(\left(1+\delta_t\right)X_t^0\right)\right)A_t.$$

We restrict δ_t, making the right-hand side bound positive. Now, let x_0 be the solution of $m_{t+1}-x_0\mu_{t+1}\left(x_0\right)=0$. Indeed, since $x\mu\left(x\right)$ is an increasing function of x starting at

0, a unique solution exists. Since equity holders do not reduce the expected equity value deliberately, it follows that

$$\begin{aligned} m_{t+1} - (1+\delta_t) X_t^0 \mu_{t+1} \left((1+\delta_t) X_t^0\right) \geq 0 &\iff (1+\delta_t) X_t^0 \mu_{t+1} \left((1+\delta_t) X_t^0\right) \leq m_{t+1} \\ &\iff (1+\delta_t) X_t^0 \leq x_0 \\ &\iff \delta_t \leq \frac{x_0}{X_t^0} - 1. \end{aligned}$$

In addition, since $\frac{x_0}{X_t^0} - 1 \to \infty$ as $X_t^0 \to 0$, this upper dividend rate bound is not active for small debt ratio values. However, the dividend rate is generally lower than the expected asset returns, which leads to $\delta_t \in [0, \delta_t^{\max}]$, where

$$\delta_t^{\max} = \min\left(\max\left(\frac{x_0 - X_t^0}{X_t^0}, 0\right), m_{t+1}\right). \tag{2.24}$$

Dynamic programming optimization allows using a recursive method starting from T with backward induction. At each time period, the value function is the sum of the immediate dividend and the expected future dividend gain. To initialize the recursion, a long time horizon T is chosen for which some simplifications are made. Because the model is Markovian, after some iterations, the terminal conditions vanish and, for that reason, the following assumption has no impact on our numerical results. The algorithm stops when the variations in the optimal dividend become very small. Then, for time T, the following simplifications are to be assumed.

Assumption 3. *For any $t > T$,*

1. *The asset returns are no longer uncertain, that is, $R_t = m_T$, $\sigma_t^2 = 0$;*
2. *There is no more possibility of conversion, that is, the CoCo debt becomes a standard debt;*
3. *The dividends are the remaining part of the returns once the interest rate payment on the debts is deducted:*

$$\delta_t A_t = \max\left(m_T A_t - \mu_t D_{t-1}, 0\right); \tag{2.25}$$

4. *If the dividend payment $m_T A_t - \mu_t D_{t-1}$ is positive, then there is no variation of the debt value, that is, $D_t = D_{t-1}$ or, equivalently, $\eta_t = 0$;*

5. *The risk-free rate r_t is constant and equal to r.*

Since a potential conversion is no longer possible (Assumption 3-2), for all $t > T$, the coupon on the CoCo debt is the same as the one on the ordinary debt, that is,

$$\begin{aligned}\mu_t &= r\left(1-z_{t-1}\right)\left(1-y_T\right) + b\left(X_{t-1},0\right) z_{t-1}\left(1-y_T\right) + b\left(X_{t-1},0\right) y_T \\ &= r\left(1-z_{t-1}\right)\left(1-y_T\right) + b\left(X_{t-1},0\right)\left(1-\left(1-z_{t-1}\right)\left(1-y_T\right)\right).\end{aligned}$$

Assume for a moment that $m_T A_t - \mu_t D_{t-1} \geq 0$. As a consequence of Assumption 3-3, $\forall t > T$,

$$FCF_t - FF_t = m_T A_{t-1} - \left(\mu_t - \eta_t\right) D_{t-1} - \delta_t A_t = 0,$$

which implies that $A_t = A_{t-1}$. Therefore,

$$\begin{aligned}\delta_t &= \frac{1}{A_t}\max\left(m_T A_t - \mu_t D_{t-1}, 0\right) \\ &= \max\left(m_T - \mu_t \frac{A_{t-1}}{A_t}\frac{D_{t-1}}{A_{t-1}}, 0\right) \\ &= \max\left(m_T - \mu_t X_{t-1}, 0\right).\end{aligned}$$

Moreover, from Assumption 3-4, $D_t = D_{t-1}$, it follows that $X_t = X_{t-1}$ and $\mu_{t+1} = \mu_t$. In addition, a recursive argument leads to the following conclusion: if the dividend rate is positive, that is

$$\begin{aligned}0 \leq \delta_{T+1} &= \frac{m_T A_{T+1} - \mu_{T+1} D_T}{A_{T+1}} \\ &= m_T - \mu_{T+1}\frac{A_T}{A_{T+1}} X_T \\ &= m_T - \mu_{T+1} X_T,\end{aligned}$$

then for all $t > T$, $X_t = X_T$, $A_t = A_T$, $\mu_t = \mu_{T+1}$ and $\delta_t = \delta_{T+1} = m_T - \mu_{T+1} X_T$. The discount factor then becomes

$$DF_{t,t+s} = \prod_{u=t}^{t+s-1}\left(1 + w(X_T)\right)^{-1} = \left(1 + w\left(X_T\right)\right)^{-s}.$$

If $m_T A_t - \mu_t D_{t-1} < 0$, then not only is there no dividend payment, but also the firm needs to issue more debt to cover the interest expenses: $D_t = D_{t-1} + \mu_t D_{t-1} - m_T A_t$. In that case,

$$\begin{aligned} FF_t &= \mu_t D_{t-1} + D_{t-1} - D_t \\ &= \mu_t D_{t-1} + D_{t-1} - D_{t-1} - \mu_t D_{t-1} + m_T A_t \\ &= m_T A_t \\ &= FCF_t, \end{aligned}$$

and $A_t = A_{t-1} + FCF_t - FF_t = A_{t-1}$. Therefore, the debt is growing and the asset value is stable, which implies that the debt ratio will increase until default.

Lemma 1. *Under Assumption 3, the expected value of the discounted dividends at time T satisfies*

$$\begin{aligned} & V\left(T, X_T^0, \delta_T\right) \mathbf{1}_{\tau_D > T} \qquad (2.26) \\ \cong\ & A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T}\right) \left\{\delta_T + \delta_{T+1} \left(\frac{\exp\left(-h\left(X_T\right)\right)}{1 + w\left(X_T\right) - \exp\left(-h\left(X_T\right)\right)}\right)\right\} \mathbf{1}_{\tau_D > T}, \end{aligned}$$

where $\delta_{T+1} = m_T - \mu_{T+1} X_T$, $h\left(X_T\right) = \lambda_D + \left(\frac{1}{\theta_D} \max\left(X_T; 0\right)\right)^{\beta_D}$ *and* $X_T = \left(1 + \delta_T\right) X_T^0$.

See proof in Appendix A.2.4.

2.4 Numerical Results

To generate realistic scenarios, the parameters correspond to the financial ratio of three banks in three different countries (Europe, Canada, and the United States (U.S.)). The aim is to analyze how "too big to fail" banks can react in the case where there is or not CoCo in the debt structure. We focus the analysis on three different periods corresponding to the pre-crisis (2006), the crisis (2008), and the post-crisis (2015).

2.4.1 Data

The sample is composed of three banks listed as Global-Systemically Important Banks (G-SIBs): Société Générale for Europe, Royal Bank of Canada for Canada, and Bank of America in the United States. For each bank and for each year, the debt ratio is calculated from Equation (2.2). The total amount of asset (A_t), deposit (F_t), long and short-term debt (B_t) and CoCo (C_t) are obtained using Bloomberg (Bloomberg Financial Analysis and Bloomberg Contingent Convertibles Search).

Figure 2.1 shows the evolution of the debt ratio over time, according to the banks. The French bank has the highest debt ratio during the financial crisis of 2008. Nowadays, the French and the Canadian banks appear to have the same level of debt in their financial structure. The decomposition of the debt structure, as presented in Table 2.1, bring out that the French bank finance these activities with fewer deposit and more bonds than the Canadian and U.S. banks.

CoCo was first launched in Europe at the end of 2009, after the financial crisis, to fulfill a need in terms of risk management. CoCos appear as a potential solution to absorb losses when the capital of banks fall below a certain level. CoCos issuance has started to rapidly increase since 2013/2014: under Basel III, certain specific CoCos are categorized as Additional Tier 1 capital (AT1). Nowadays, CoCos throughout the world are mainly issued by European and Asian companies. Canadian banks started to issue CoCos in 2014 while U.S. banks do not.

Figure 2.1. Evolution of the Debt Ratio Over Time

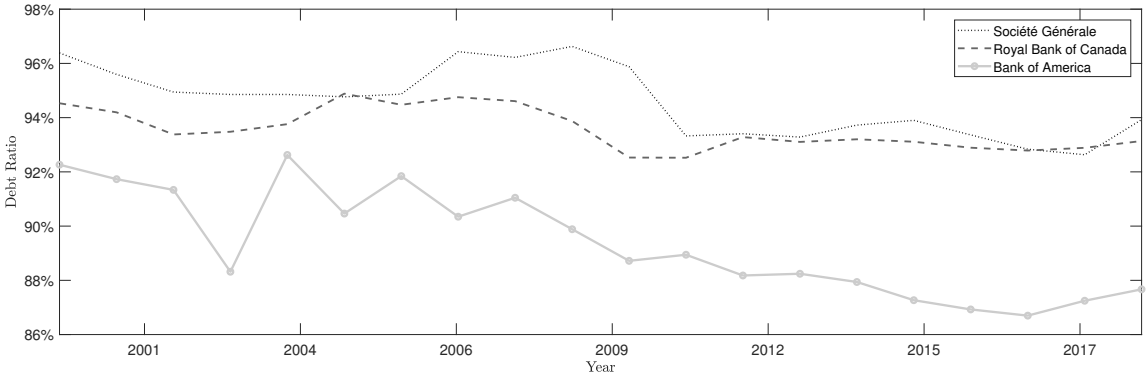


Table 2.1. Source of Assets

	Société Générale	**Royal Bank of Canada**	**Bank of America**
Equity/Asset	5.41%	6.43%	10.62%
Deposit/Asset	40.60%	77.37%	58.86%
Bonds (including CoCo)/Asset	53.99%	16.20%	32.52%

The table presents the average of three financial ratios over 19 years, from 1999 to 2018. Asset is computed using Equation (2.1). Deposit, Equity and Bonds (including CoCo) are based on the financial statement available on Bloomberg.

To initialize the dynamic optimization, the percentage of CoCos in the debt structure is assumed to be 1%, since it is quite representative of the European and Canadian banks. Indeed, based on the debt value and the amount of CoCos issued by banks, obtained from Bloomberg, the percentage of CoCos in the debt structure is around 1.15% for European banks and around 0.64% for Canadian banks. For the U.S. banks, assuming 1% of CoCos could shed light on how it could help the bank in case of financial distress.

Table 2.2 presents the parameter values obtained from the financial ratio of three banks, for three different periods (pre-crisis (2006), crisis (2008), and post-crisis (2015)). The critical debt ratio parameters (θ_C and θ_D) and the convexity parameters (β_C and β_D) characterizing the conversion and default intensities were obtained by calibrating the conversion and default probabilities (c.f., Appendix A.3). The results obtained for the critical debt level of the conversion intensity θ_C is near 95%. Indeed, under Basel III, CoCos that have a minimum trigger level of 5.125% in terms of CET1 by RWA and that have perpetual coupons are qualified as AT1. Hence, there had been a trend towards issuing CoCos with such characteristics. This corresponds to a trigger level imposed by the regulator (α) in terms of debt ratio close to 1–5.125% of RWA. The critical debt ratio of the default intensity obtained, θ_D, is slightly greater than 100%. When a firm has more liabilities than assets, there is a high risk of insolvent and excess leverage. The convexity conversion (default) parameter, β_C (β_D), is designed to increase the conversion probability (default probability) rapidly around the critical debt ratio θ_C (θ_D). This leads to high convexity parameters.

Table 2.2. Parameter Values of the Dynamic Optimization per Bank per Year.

	Société Générale			**Royal Bank of Canada**			**Bank of America**		
	2006	**2008**	**2015**	**2006**	**2008**	**2015**	**2006**	**2008**	**2015**
					Returns				
r_t	3.76%	1.99%	−0.2%	4.07%	0.89%	0.51%	4.94%	0.28%	0.73%
r_t^E	20.04%	7.02%	6.23%	23.21%	17.64%	18.42%	18.07%	1.81%	6.27%
m	3.07%	3.43%	1.26%	3.5%	4.35%	2.25%	3.84%	4.66%	0.36%
σ^2		2%			2%			2%	
					Initial debt structure				
F_0	31.41%	29.71%	44.35%	79.58%	86.39%	83.45%	54.78%	56.13%	70.29%
y_0		1%			1%			1%	
					Conversion risk				
α	98.47%	98.43%	98.63%	97.86%	98.03%	98.03%	96.30%	96.28%	96.65%
β_C		74.48			66.55			39.85	
θ_C		97.42%			96.82%			94.69%	
ρ_C		90%			90%			90%	
					Default risk				
θ_D		107.7%			104.11%			107.29%	
β_D		47.2			69.9			40.46	
λ_D		0			0			0	
ρ_D		0.4			0.4			0.4	
					Numerical scheme				
T		30			30			30	
Δ_x		0.002			0.002			0.002	
Δ_t		1			1			1	

r_t stands for risk-free rate and corresponds to the one-year zero coupon curve computed by each central bank of each region. r_t^E stands for the Return on Equity. m stands for the average of the expected return on capital and σ^2 its variance. F_0 means the percentage of deposit at time 0. $y_0 = C_0/D_0$ is the initial proportion of CoCo debt. The conversion and default intensities functions are respectively Equations (2.12) and (2.17). The debt ratio mesh parameter Δ_x represents the distance between two consecutive debt ratios. The time discretization is $\Delta_t = 1$ year.

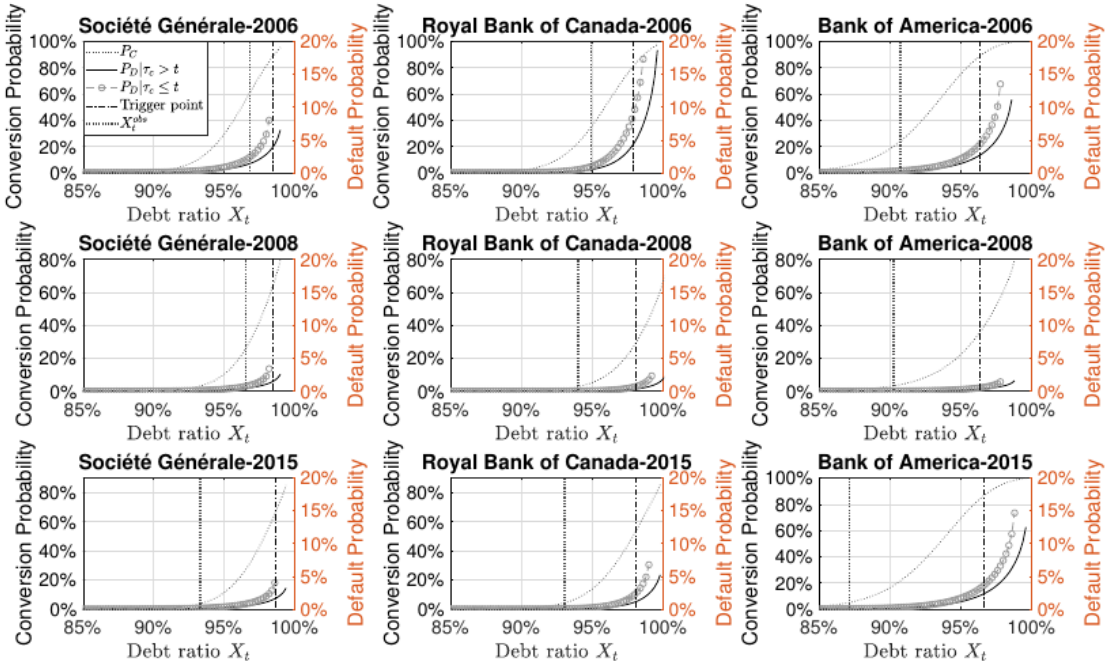
The credit risk management literature suggests that the recovery rate is around 40%, based mainly on observed data (Altman and Kishore, 1996; Duffie and Lando, 2001). The risk-free rate corresponds to the one-year zero coupon curve computed by the European Central Banking, Bank of Canada, and the Federal Reserve Bank of St. Louis. r_t^E corresponds to the Return on Equity (ROE). m is the expected capital return expected by investors and is therefore not directly observable. We have chosen to proxy m_t by the average ROE over the last five years and the average risk-free rate, such that $m_t = (1-x)\bar{r}_t^E + x\bar{r}_t$.

2.4.2 Empirical Results

Using the parameters presented in Table 2.2, Equations (2.14), (2.18) and (2.19) are used to compute the one-year conversion probability and the one-year default probabilities as a function of the actual debt ratio. Figure 2.2 shows that, for each bank and for each year, the conversion probability increases with the debt ratio. Around the debt ratio critical level, the conversion probability tends to climb to 100% reflecting the mandatory conversion. As expected, the presence of CoCo debt instruments reduces the default probabilities. When there is no more CoCo in the debt structure, the default probability increases more rapidly. It confirms that having CoCos in the debt structure acts like a safety buffer. The conversion and default probabilities are not only driven by the intensity parameters, but also by the actual market conditions, that is, the risk-free rate, the expected return on the capital, and the debt structure. Figure 2.2 shows that the probabilities of default were greater in 2006, with more indebted banks.

Equations (A.4)–(A.6) give the coupon rate on the standard debt (with and without CoCos) and the CoCo debt. CoCo debt holders face two risks: the risk of default (like the standard debt holders), plus the risk of mandatory conversion. However, since at conversion, CoCo debt holders receive in terms of equity the value of their investment, there is no conversion recovery risk. Figure 2.3 shows that when the contingent debt is present in the debt structure, the CoCo debt is more expensive than the standard debt. Moreover, as the contingent debt mitigates the default risk, the coupon rate of the ordinary debt is notably lower in the presence of CoCos. Comparing the result for each bank and each year, Figure 2.3 shows that the coupons are lower over the years due to lower interest rate. The risk-free

Figure 2.2. One-Year Conversion and Default Probabilities

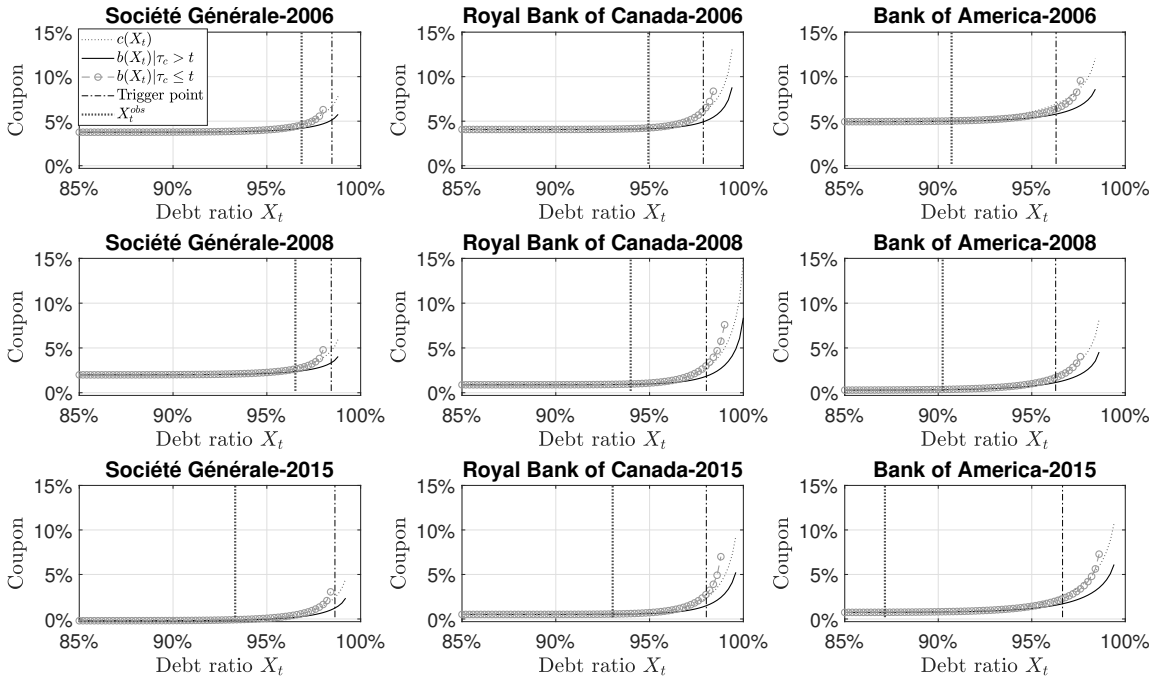


All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2.2. Each column corresponds to a specific bank and each line corresponds to a specific year. The dark grey dotted line represents the conversion probability. The black line represents the one-year default probability in the presence of CoCos in the debt structure. The light grey circle dash line represents the one-year default probability without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1-5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

rates were higher before the financial crisis and stayed at low levels after the crisis.

The cost of debt (Equation (2.6)) is a weighted average of each component (deposit, standard debt, and CoCo debt). Figure 2.4 shows that the cost of debt increases for large debt ratio which is a direct consequence of the coupon curve behaviour, as observed in Figure 2.3. Even if the CoCo coupon rate is large relatively to the standard debt coupon rate, for high-debt ratio, the presence of CoCo in the debt structure reduces the cost of debt. Over the years, the cost of debt decreases, mainly due to the lower risk-free rate. Because

Figure 2.3. CoCos and Standard Debts Coupons

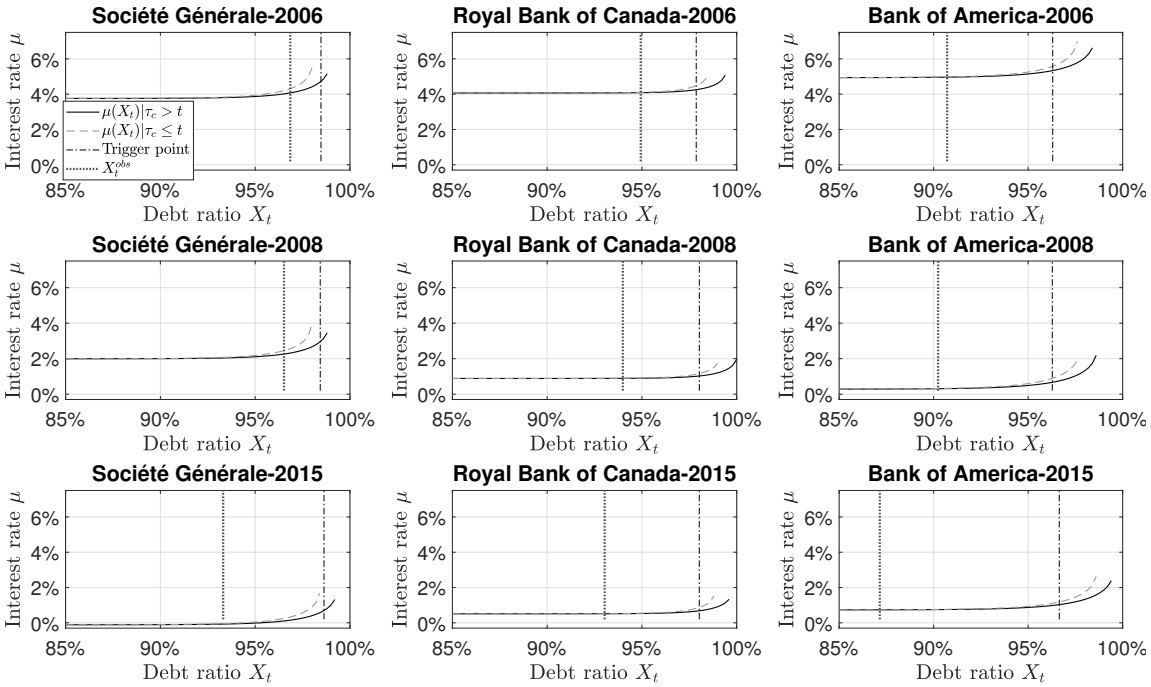


All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2.2. Each column corresponds to a specific bank and each line corresponds to a specific year. The dark grey dotted line represents the coupon on the CoCo debt. The black line represents the coupon on the standard debt in the presence of CoCos in the debt structure. The light grey circle dash line represents the coupon on the standard debt without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1-5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

Royal Bank of Canada uses mainly deposits over bonds, the cost of debt is not so much affected by the increase in the coupon value for high-debt ratio.

The cost of debt directly affects the cost of capital (Equation (2.20)). Indeed, the lower the cost of capital, the more likely the bank is creating value. The cost of capital indicates the minimum rate of return before generating profit. When the debt ratio is close to 0%, the bank is financed by equity: the cost of capital corresponds to the cost of equity. Between the two extremes, the cost of capital tends to decrease: the cost of debt is generally a cheaper source of financing than equity, except when the firm is in financial distress. As

Figure 2.4. Cost of the Debt

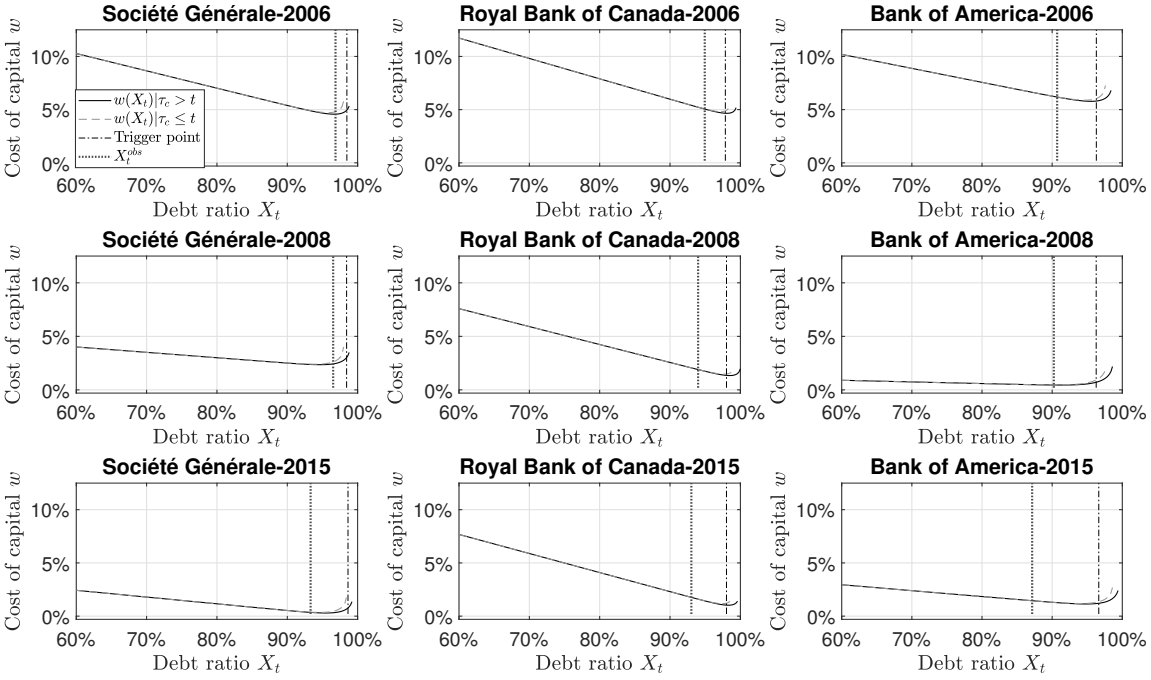


All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2.2. Each column corresponds to a specific bank and each line corresponds to a specific year. The black line represents the cost of debt when there is CoCos in the debt structure. The light grey dashed line represents the cost of debt without CoCos in the debt structure. The vertical black dashed-dotted line corresponds to the trigger level (α) of 1-5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

shown in Figure 2.5, the equity return declines substantially during the 2008 financial crisis, especially for the U.S. and European banks and stays at a low level after the crisis. The low cost of capital for Société Générale and Bank of America in 2008 and 2010 is in line with the low ROE and risk-free rate observed for these periods. The effects of the 2008 financial crisis were less important for the Canadian banks. In Europe, the 2011 debt crisis is also a reason for these low interest rates and ROE. In all studied cases, the cost of capital is at its lowest just before the debt ratio conversion trigger, around 95%. In Figure 2.5, we see that, for all studied cases, the debt ratio is to the left of the debt ratio level that minimizes the cost of capital. In a multi-period setting, banks use a precautionary cushion to stay away

from the conversion and default thresholds.

Figure 2.5. Cost of the Capital

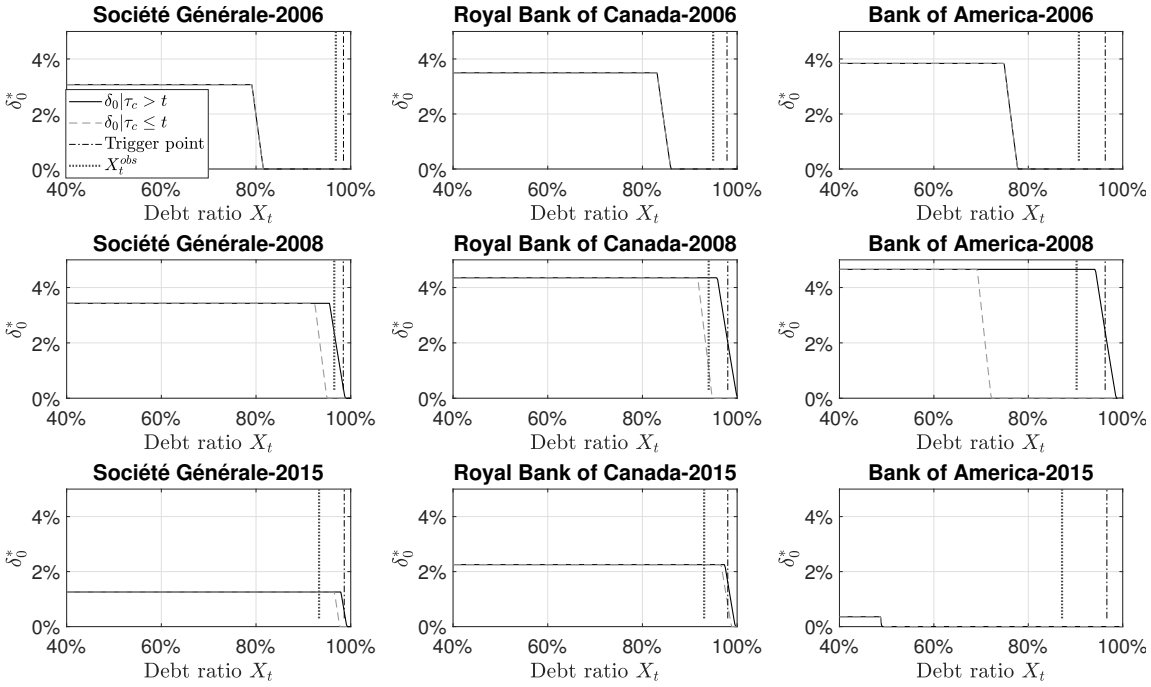


All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2.2. Each column corresponds to a specific bank and each line corresponds to a specific year. The black line represents the cost of capital when there is CoCos in the debt structure. The light grey dashed line represents the cost of capital without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1-5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

To mitigate the effect of Assumption 3, the dynamic optimization program is applied recursively using a backward recursion until there are no significant changes in the dividend policy and the value of the discounted aggregated dividend per share (Equation 2.21) and Equation 2.22). Figure 2.6 shows the optimal dividend rate in function of the debt ratio at time $t = T - 30$. The optimal dividend rate is more permissive when there is CoCo in the debt structure in 2008 and 2015. The difference between the optimal dividend rate with and without CoCos is the largest for Bank of America during the financial crisis. The U.S. banks were the most affected by the subprime crisis. The presence of CoCos in the debt

structure has a significant effect on the optimal dividend rate. The optimal dividend rate in 2015 for Bank of America is very small (0.36%) due to the expected return on capital: we use a weighted average over the mean ROE and risk-free rate over the last five years, and these parameters are very small due to the low interest rate policy and the effect of the crisis.

Figure 2.6. Optimal Dividend at Time $T = 0$



All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2.2. Each column corresponds to a specific bank and each line corresponds to a specific year. The black line represents the optimal dividend rate when there is CoCos in the debt structure. The light grey dashed line represents the optimal dividend rate without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1-5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

Figure 2.7 shows the (normalized) discounted cumulated dividend value of the primary equity holders at time $t = T - 30$. The primary equity holders have larger dividend gains when the CoCo debt is part of the debt structure. This is due to the dilution effect created by the CoCo conversion: CoCo debt holders become equity holders, then, mechanically,

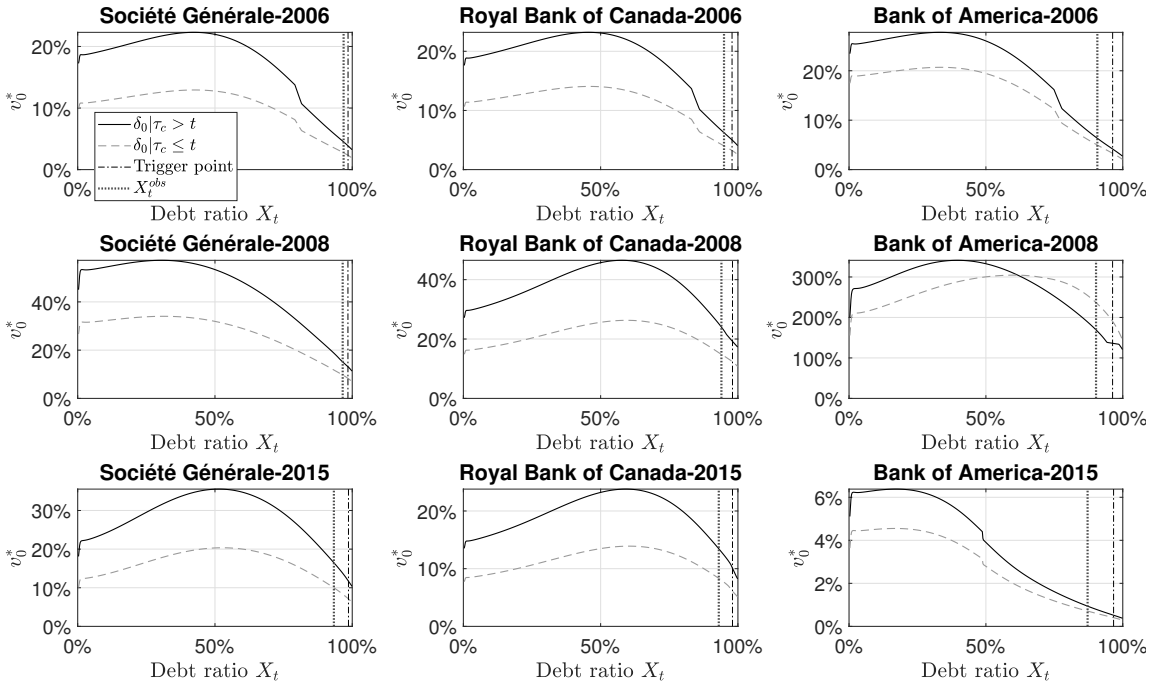
decrease the dividend-per-share. The (normalized) discounted cumulated dividend value is at its maximum for a debt ratio around 50%, except for Bank of America in 2015. For all studied cases, the observed debt ratio does not maximize the future dividends. The level of the discounted cumulated dividend for Bank of America in 2008 reflects the sudden financial crisis and our parametrization of the expected return on capital. The ROE was in average equal to 18.82% from 2003 to 2006, 10.77% in 2007 and 1.81% in 2008. The low ROE in 2008 affected the level of the cost of capital and so the discount factor used in the dynamic program. However, the expected return on capital is difficult to estimate, and so we use past observations on the ROE and risk free rate instead of prospective data. This leads to a high expected return on capital, which affects the level of the optimal dividend rate. A sensitivity analysis follows.

2.4.3 Sensitivity Analysis

Sensitivity analysis is performed to examine how changes in the parameters affect our conclusions and in particular the discounted cumulated dividend. First, what is the impact of letting the proportion of CoCos in the debt structure increases from 1% to 10%, all other parameters being the same? Obviously, Figure 2.8 shows that, after conversion, discounted cumulated dividends are much lower because the dilution effect is more important: there is more debt converted to equity. Before the conversion, there is not much impact or a slight decrease due to the modest increase of the conversion risk. For a large debt ratio, the discounted cumulated dividends are slightly smaller.

Second, we analyze how good the situation is for the firms, that is, how firms with high ROE affect the dividends. We fix the ROE for each bank equal to the maximum ROE observed from 1999 to 2017, all other parameters being the same as in the base case. Changing the ROE also impacts the expected return on capital (m). These two parameters impact the cost of capital, the discount factor, and thus the discounted cumulated dividends, as well as the optimal dividend rates. Indeed, an increase of the expected return on capital leads to higher optimal dividend rates. For example, the optimal dividend rate for Bank of America in 2015 increases at 2.97% and create higher discounted cumulated dividends. In 2008, for Bank of America, increasing the ROE allows to have reasonable values of

Figure 2.7. (Standardized) Value of Discounted Cumulated Dividends at Time $T = 0$

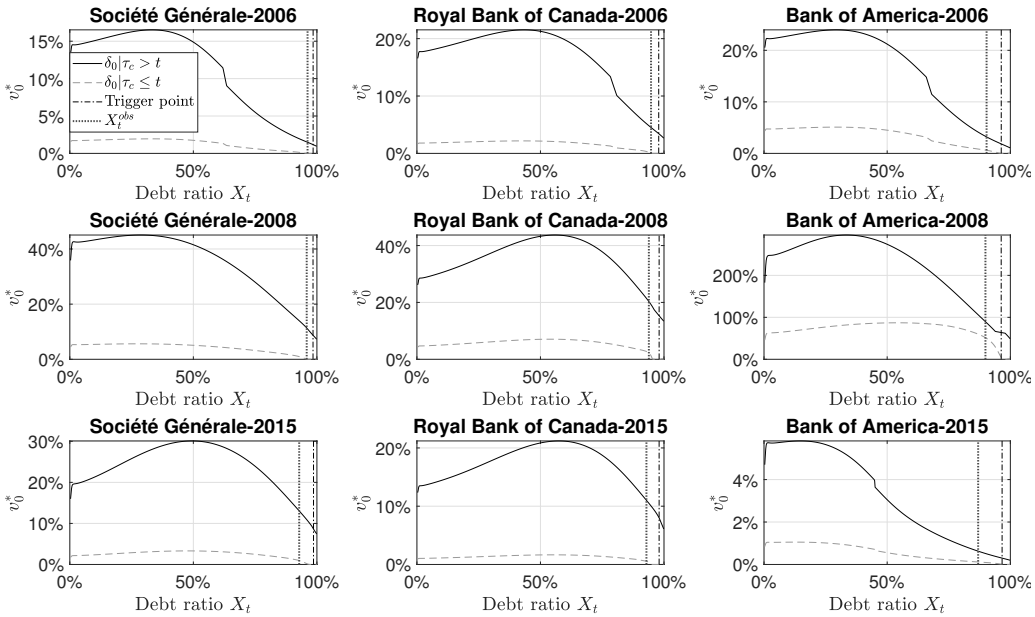


All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2.2. Each column corresponds to a specific bank and each line corresponds to a specific year. The black line represents the case when there is CoCos in the debt structure. The light grey dashed line represents the case without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1-5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

discounted cumulated dividends, as shown in Figure 2.9. For Société Générale in 2008 and 2015, increasing the ROE slightly decrease the discounted cumulated dividends. There is a trade-off between having high dividend today and a high dividend in the future, which can create impatience for the equity holder. Indeed, increasing the ROE leads to an increase of the expected return on capital due to our parametrization. This also leads to an increase in the cost of capital that affects the discount factor of the future dividend. It is not clear what is the expectation of the discounted cumulated dividends in that context.

Finally, we change the default intensity to have higher default probabilities, reflecting bad conditions for banks. We assume a one-year default probability of 8% around the

Figure 2.8. Discounted Cumulated Dividends at Time $T = 0$, with $y = 10\%$



critical debt ratio imposed by the regulator. An increase of the default probability decreases the value of the discounted cumulated dividend faster when the bank is largely indebted, as shown in Figure 2.10. As expected, when the bank is not so much indebted, there is no or a slight decrease in the level of the discounted cumulated dividend, especially, in the absence of CoCo in the debt structure, as observed for Société Générale in 2006 and 2008.

Figure 2.9. Optimal Dividend and Discounted Cumulated Dividends at Time $T = 0$, With High Return on Equity.

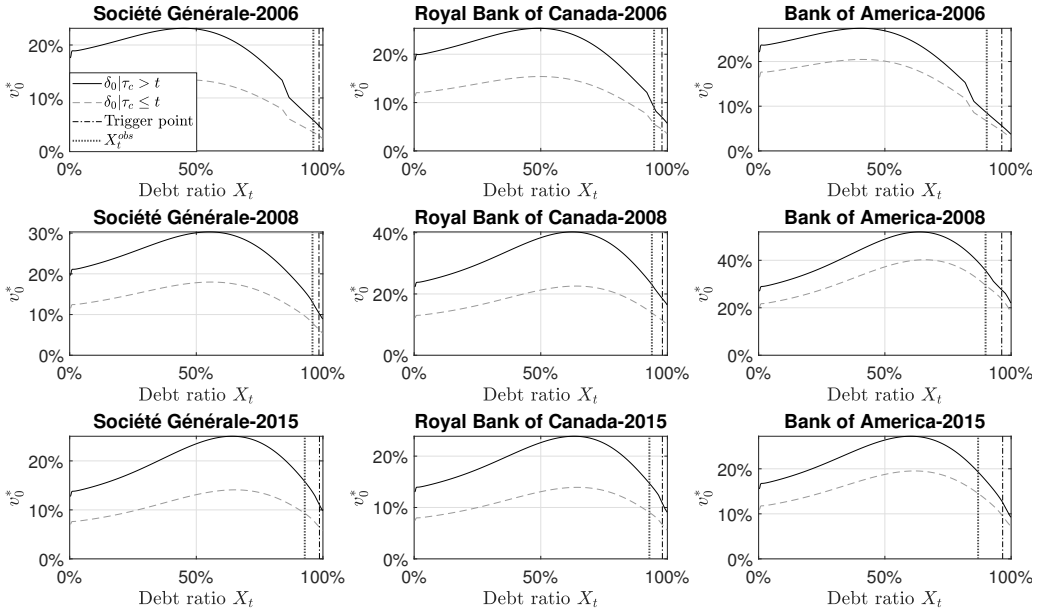
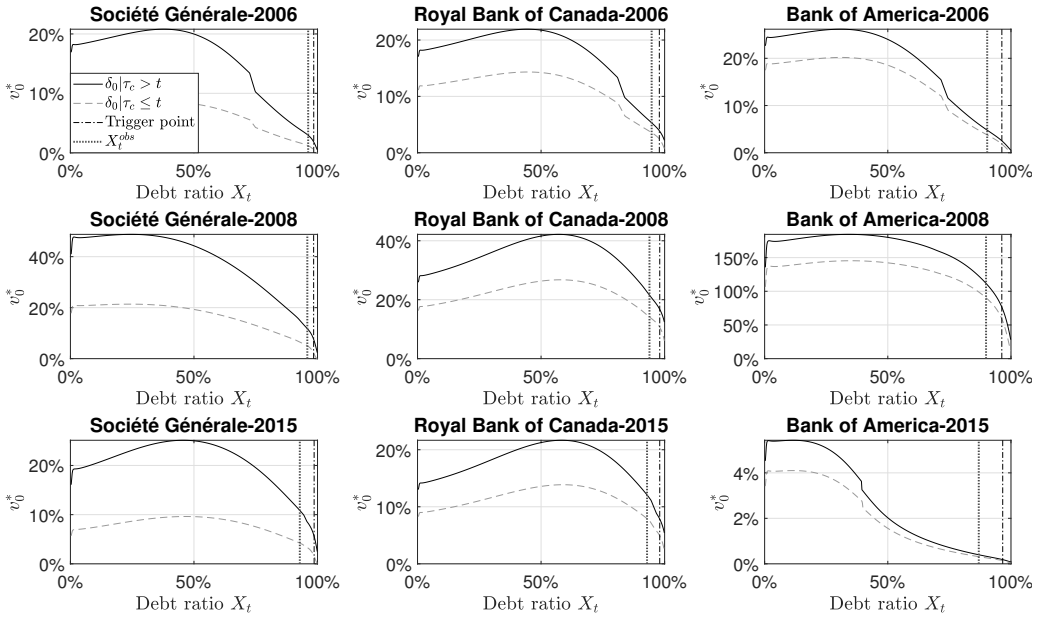


Figure 2.10. Optimal Dividend and Discounted Cumulated Dividends at Time $T = 0$, With Higher Default Probabilities



2.5 Conclusion

We introduced a reduced form approach that also includes information about the financial health of the firm to model the impact of CoCos in the debt structure by taking into account the uncertainty and the time-varying default and conversion risks. To analyze the impact of CoCos in the bank's capital structure, we set up a hypothetical capital structure consisting of equity, ordinary debt, and CoCo debt in order to understand the benefits/costs of convertible contingent debt. Real parameters are used, corresponding to three banks considered as G-SIBs (too big to fail) in three different markets (Europe, Canada, and U.S.) for three periods (pre-crisis, crisis, and post-crisis) to generate scenarios. Results reveal that CoCos reduce the cost and the risk of standard debt when they are being added to the capital structure. In fact, CoCos acts as a precautionary buffer to prolong the default time and hence to reduce the risk of default for ordinary debt. Meanwhile, CoCo itself is a more expensive instrument compared to the ordinary debt, remembering that it has zero chance of recovery on default. Furthermore, the results show that the presence of CoCos in the debt structure reduce the total cost of debt, knowing that the cost of debt is a weighted average of each debt component. From the point of view of primary equity holders, the presence of CoCos in the bank's capital structure increases the aggregated value for shareholders. The optimal dividend policy derived from the dynamic optimization suggests paying more dividends when the CoCo is active and leads to higher aggregated dividends, especially in the context of crisis and low return.

References

Albul, B., Jaffee, D. M., & Tchistyi, A. (2010). Contingent convertible bonds and capital structure decisions.

Altman, E. I., & Kishore, V. M. (1996). Almost everything you wanted to know about recoveries on defaulted bonds. *Financial Analysts Journal*, *52*(6), 57–64.

Bolton, P., & Samama, F. (2012). Capital access bonds: Contingent capital with an option to convert. *Economic Policy*, *27*(70), 275–317.

Brigo, D., Garcia, J., & Pede, N. (2015). Coco bonds pricing with credit and equity calibrated first-passage firm value models. *International Journal of Theoretical and Applied Finance*, *18*(03), 1550015.

Chen, N., Glasserman, P., Nouri, B., & Pelger, M. (2013). Cocos, bail-in, and tail risk. *Working Paper, Office of Financial Research, U.S. Treasury Department,Washington, DC*.

Chen, N., Glasserman, P., Nouri, B., & Pelger, M. (2017). Contingent capital, tail risk, and debt-induced collapse. *The Review of Financial Studies*, *30*(11), 3921–3969.

Cheridito, P., & Xu, Z. (2015). A reduced-form contingent convertible bond model with deterministic conversion intensity. *Journal of Risk*, *17*(3).

Chung, T.-K., & Kwok, Y.-K. (2016). Enhanced equity-credit modelling for contingent convertibles. *Quantitative Finance*, *16*(10), 1511–1527.

De Spiegeleer, J., Höcht, S., Marquet, I., & Schoutens, W. (2017). Coco bonds and implied cet1 volatility. *Quantitative Finance*, *17*(6), 813–824.

Duffie, D. (2010). *A contractual approach to restructuring financial institutions* (K. E. Scott, G. P. Shultz, & J. B. Taylor, Eds.). Hoover Institution, Stanford University.

Duffie, D., & Lando, D. (2001). Term structures of credit spreads with incomplete accounting information. *Econometrica*, *69*(3), 633–664.

Flannery, M. J. (2005). *No pain, no gain? effecting market discipline via reverse convertible debentures* (H. S. Scoot, Ed.). Oxford University Press Oxford, UK.

Flannery, M. J. (2014). Maintaining adequate bank capital. *Journal of Money, Credit and Banking*, *46*(s1), 157–180.

Flannery, M. J. (2016). Stabilizing large financial institutions with contingent capital certificates. *Quarterly Journal of Finance*, *6*(02), 1650006.

Glasserman, P., & Nouri, B. (2012). Contingent capital with a capital-ratio trigger. *Management Science*, *58*(10), 1816–1833.

Hilscher, J., & Raviv, A. (2014). Bank stability and market discipline: The effect of contingent capital on risk taking and default probability. *Journal of Corporate Finance*, *29*, 542–560.

McDonald, R. L. (2013). Contingent capital with a dual price trigger. *Journal of Financial Stability*, *9*(2), 230–241.

Pennacchi, G. (2010). A structural model of contingent bank capital. *FRB of Cleveland Working Paper No. 10-04*.

Sundaresan, S., & Wang, Z. (2015). On the design of contingent capital with a market trigger. *The Journal of Finance*, *70*(2), 881–920.

Chapter 3

Predicting the Term Structure of Credit Spreads for High- and Low-Frequency Issuers: A Combined Nelson-Siegel and Clustering Approach

Abstract

We propose and evaluate methodologies to forecast credit spread term structures and corporate bond prices by addressing challenges related to bond-trading frequencies across issuers. For high-frequency issuers, we implement a two-step Nelson–Siegel model coupled with a Vector Autoregressive (VAR) framework. Due to limited data availability for low-frequency issuers, we develop a two-step clustering approach: we first identify distinct clusters of high-frequency issuers using a Gaussian Mixture Model; we then link each low-frequency issuer to these clusters using issuer-specific information, and we compute weighted-average credit-spread curves according to posterior cluster probabilities. Empirical analysis on USD-denominated corporate bond data demonstrates predictive performance, although accuracy declines for longer maturities and in sparsely populated issuer clusters, emphasizing the critical importance of data quality.

3.1 Introduction

This study examines methodologies for predicting the term structure of corporate credit spreads and bond prices, developed in collaboration with a partner company, a Toronto-based financial analytics firm specializing in fixed-income markets, and supported by a Mitacs scholarship. In exchange for access to a promising, unique data lake, the partner company requested forecasts of bond prices at different time horizons. We chose to predict the corporate credit spread term structure to address the bond-price prediction challenge. Corporate credit spreads are defined as the differences between corporate zero-coupon yield curves and a risk-free benchmark, typically U.S. Treasury yields for USD-denominated corporate issuers. Corporate issuers differ in their frequency and volume of trading activity. We classify issuers into two categories: those with extensive historical bond-trading data ("high-frequency issuers") and those with limited or infrequent trading activity ("low-frequency issuers"). This paper addresses three research questions: How can the zero-coupon spread curves and associated bond prices be predicted from one- to five-day-ahead time horizons for high-frequency issuers? Can issuance prices be forecast for newly issued bonds by high-frequency issuers? For low-frequency issuers, or issuers without sufficient historical bond data, how can we reliably estimate zero-coupon corporate spread curves and subsequently predict issuance prices for their bonds?

We approach these questions by first estimating issuer-specific zero-coupon spread curves through the Nelson-Siegel framework, originally proposed by Nelson and Siegel, 1987. The Nelson-Siegel model is well-known for its flexibility and parsimony, which can accommodate diverse yield curve shapes using only four coefficients: level, slope, curvature, and hump position of the term structure (Diebold & Li, 2006; Sundaram & Das, 2011). Despite these advantages, the Nelson-Siegel model is known for parameter instability, posing challenges for reliable forecasting. To enhance parameter stability, we adopt the two-stage estimation approach proposed by Gauthier and Simonato, 2012. This approach initially utilizes a linearized version of the Nelson-Siegel model informed by prior information, providing starting values for subsequent optimization. Moreover, we introduce a penalization component into the optimization criterion whenever parameter estimates deviate beyond one standard error from previous periods, thus further stabilizing the parameter

estimation.

We then forecast the zero-coupon spread curves from one- to five-day ahead time horizons. While Diebold and Li, 2006 applied an autoregressive approach, our analysis identifies substantial correlations among estimated coefficients, which motivate us to employ a Vector AutoRegressive (VAR) model to better capture these dependencies. This methodology allows us to predict new bonds issued by high-frequency issuers.

For low-frequency issuers, the Nelson-Siegel methodology is infeasible due to its minimum requirement of four actively traded bonds per issuer per day. Furthermore, extrapolating credit spreads beyond observed maturities can introduce significant pricing biases. To address these constraints, we propose a two-step clustering methodology. First, we apply a Gaussian Mixture Model to group high-frequency issuers into distinct clusters based on their credit-risk characteristics. We then build a confusion matrix summarizing the probability of membership in each cluster given sector, seniority, and credit rating. Second, for low-frequency issuers, we compute a weighted-average credit-spread curve using cluster probabilities derived from the confusion matrix. This allows us to predict the credit spread and thus the bond prices for those low-frequency issuers.

Although the partner company initially committed to providing an extensive, high-quality fixed-income securities database essential for robust predictive modelling, practical implementation encountered significant obstacles due to limited data accessibility and data quality issues. These limitations influenced both the methodological development and the reliability of the resulting predictions.

The remainder of the paper is structured as follows. Section 3.2 describes the dataset and discusses our data-cleaning and validation protocols. Section 3.3 details the Nelson–Siegel estimation and VAR forecasting for high-frequency issuers. Section 3.4 presents the clustering methodology for low-frequency issuers. Section 3.5 concludes the paper.

3.2 The Data

Significant challenges emerged related to data accessibility and quality. We encountered inconsistencies such as unflagged simulated data, which complicated the analysis. For

the estimation of the zero-coupon (ZC) yield curve for U.S. government securities and the ZC spread curves for corporate issuers, we require precise inputs: dirty price, bond maturity (tenor), coupon rate, and coupon frequency. Market platforms typically provide bond prices as clean prices, which exclude accrued interest. Consequently, we convert these clean prices into dirty prices by incorporating accrued interest. Given data-quality issues, our analysis began with rigorous cleaning procedures. We focused exclusively on plain-vanilla bonds, excluding covered bonds, Islamic bonds, structured products (e.g., asset-backed or mortgage-backed securities), and private placements. Only bonds with at least USD 250 million outstanding were retained, per the partner company's requirement. We further refined the dataset by excluding securities with anomalous pricing. Observations were removed if bid or ask yields were negative or if the bid yield was inexplicably lower than the ask yield. Because the yields in the dataset are averages across recent trades rather than closing yields, bid yields can occasionally appear below ask yields when trades occur at wide spreads. Although infrequent, this inconsistency affected up to 25 percent of observations for certain issuers, underscoring a major data-quality concern. We also removed records where the issue date occurred after the trade date. Descriptive statistics of the average number of bonds per year and per issuer can be found in Table 3.1 and Table 3.3.

Given the absence of prices with accrued interest[1], we derived mid prices based on the provided mid yields as the sum of discounted cash flows with Accrued Interest (AI):

$$B_i = 100 \times \frac{c_i}{f_i} \frac{1 - \left(1 + \frac{ytm_i}{f_i}\right)^{-T \times f_i}}{\frac{ytm_i}{f_i}} + \frac{100}{\left(1 + \frac{ytm_i}{f_i}\right)^{T \times f_i}} + AI_i,$$

$$AI_i = 100 \times \frac{c_i}{f_i} \times \left(\frac{\text{days since last coupon}}{\text{days in coupon period}}\right),$$

where 100 is the face value, c_i is the coupon rate of bond i, f_i represents the coupon frequency, T is the time to maturity in years, and ytm_i is the yield-to-maturity. To mitigate the impact of outliers, we eliminated bond prices falling outside the 0.1% and 99.9% percentiles. Credit spread curves for each issuer were constructed using bonds sharing uniform

[1] dirty price

seniority. For 16.81% of USD-denominated bonds, we are missing coupon frequency information; hence, we assumed that coupons are paid semi-annually, aligning with North American market standards. This assumption is supported by our data: when not missing, semi-annual coupon frequency accounts for approximately 57.95% of USD-denominated bonds in our dataset. Due to incomplete data on day-count conventions, we estimated the remaining number of coupon payments by rounding up the product of each bond's tenor and coupon frequency. These extensive data-cleaning and adjustment procedures were critical to ensuring the integrity of our empirical analysis and subsequent predictive-modelling efforts.

3.3 Prediction for High-Frequency Issuers: Nelson-Siegel Model Using the Two-Step Approach

3.3.1 Nelson-Siegel Two-Step Approach

The model is based on the construction of the zero-coupon curve using coupon-bearing bond prices. The price of a coupon-bearing bond can be represented as a function of zero-coupon bond prices, as follows:

$$\begin{aligned} B_i &= \sum_{j=1}^{m_i} c_{i,j}, P\big(t_{i,j}\big) \\ &= \sum_{j=1}^{m_i} c_{i,j} e^{-y^{ZC}(t_{i,j})t_{i,j}}, \end{aligned} \tag{3.1}$$

where $P\left(t_{i,j}\right)$ is the zero-coupon bond price serving as the discount factor; m_i is the number of remaining coupon payments of bond i; $t_{i,j}$ is the time-to-maturity to the jth coupon payment of bond i; and y^{ZC} is the zero-coupon yield curve. The jth coupon payment of bond i, $c_{i,j}$, is defined as

$$c_{i,j} = 100\Big(\tfrac{c_i}{f_i}\mathbb{I}_{j<m_i} + \big(1+\tfrac{c_i}{f_i}\big)\mathbb{I}_{j=m_i}\Big),$$

where 100 is the face value, c_i is the coupon rate, and f_i is the coupon frequency of bond i.

The zero-coupon (ZC) yield y^{ZC} is modelled using the Nelson-Siegel framework (Nelson & Siegel, 1987) as

$$y^{ZC}(\theta, t) = \beta_0 + \beta_1\phi_1(t, \tau) + \beta_2\phi_2(t, \tau), \tag{3.2}$$

$$\phi_1(t, \tau) = \frac{1 - e^{-t/\tau}}{t/\tau}, \tag{3.3}$$

$$\phi_2(t, \tau) = \phi_1(t, \tau) - e^{-t/\tau}, \tag{3.4}$$

where $\theta = \beta_0, \beta_1, \beta_2, \tau$ is the vector of unknown parameters. Here, β_0 denotes the long-term yield; β_1 controls the slope (so that $\beta_0 + \beta_1$ corresponds to the short-term yield); β_2 captures the curvature; and τ determines the placement of the hump.

3.3.1.1 Estimation of the Zero-Coupon Yield Curve for Government Issuers

Our initial step is to estimate the zero-coupon yield curve for government bonds using Equation (3.1). We determine the Nelson-Siegel coefficients by numerically minimizing the root-mean-squared relative error (RMSRE), defined as:

$$\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{\widehat{B}_i - B_i}{B_i}\right)^2}, \tag{3.5}$$

where N_t is the number of bonds available at time t, $\widehat{B}_i$ is the model-estimated price, and B_i is the observed bond price.

The Nelson-Siegel approach is known to encounter optimization challenges: different coefficient sets can produce almost identical yield curves, and there are several local optima (Gauthier & Simonato, 2012; Sundaram & Das, 2011). To mitigate this issue, we adopt the two-step linearized Nelson-Siegel approach proposed by Gauthier and Simonato, 2012 to establish initial coefficient values prior to nonlinear optimization. The intuition underlying this linearization is as follows: for a given fixed τ, the three other coefficients (β_0, β_1, β_2) can be estimated by ordinary least squares (OLS); then, over a grid of τ values, we select the one that minimizes the RMSRE in Equation (3.5). This procedure is based on a Bayesian framework, as prior information on the coefficients is incorporated into the estimation process.

For the estimation of U.S. Treasury yields, our priors are derived from the long-term yield, slope, and curvature of the Treasury constant-maturity yields, available through the Federal Reserve Economic Data (FRED) database.[2] The long-term yield prior corresponds to the 30-year constant-maturity yield; the slope prior corresponds to the spread between the 3-month and 30-year constant-maturity yields; and the curvature prior corresponds to twice the 5-year yield minus the sum of the 3-month and 30-year yields. The details of the linearization step are presented in Appendix B.1.

We verify coefficient stability by ensuring each estimated coefficient remains within one standard error of its previously estimated value.[3] If any coefficient deviates beyond this threshold, we re-estimate the coefficient set by introducing a penalty term into the RMSRE objective function to enforce coefficient stability over time. The penalized-loss function is defined as:

$$\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{\widehat{B}_i - B_i}{B_i}\right)^2} + \sum_{j=0}^{2}\left(\widehat{\beta}_{j,t} - \beta_{j,t-1}\right)^2 + \left(\frac{\widehat{\tau}_t - \tau_{t-1}}{\tau_{t-1}}\right)^2, \tag{3.6}$$

where $\widehat{B}_i$ is the estimated price from Equation (3.1) using the estimated Nelson-Siegel coefficients $\widehat{\beta}_{0,t}, \widehat{\beta}_{1,t}, \widehat{\beta}_{2,t}, \widehat{\tau}_t$. This penalty term ensures stability and interpretability of the coefficients over time, which is essential for subsequent forecasting.

3.3.1.2 In-Sample Results for United States Treasury Bonds

Table 3.1 provides descriptive statistics for the dataset of U.S. Treasury bonds, covering the period from January 2013 through December 2019. The daily number of bonds available increases over this period, from an average of 59 bonds per trading day in 2013 to 178 bonds per day in 2019, reflecting improved database coverage in recent years. The average coupon rate decreases over the period, from 4.12% in 2013 to 2.25% in 2019, consistent with the general downward trend in U.S. interest rates during this period. The average bond maturity decreases from 15.76 years in 2013 to 5.80 years in 2019; the median indicates

[2] https://fred.stlouisfed.org/categories/115

[3] The standard errors are calculated using a rolling window of three months. During the first three months of the dataset, standard errors are computed based on the linearized model's coefficient estimates.

Table 3.1. Descriptive Statistics of U.S. Treasury Bonds per Year

	Bonds per day			Coupon	Tenor			
Year	Min	Avg	Max	Avg	Min	Avg	Med	Max
2013	7	58.79	76	4.12%	3.87	15.76	13.14	30
2014	4	39.96	80	3.12%	2.87	13.50	8.43	30
2015	4	57,38	114	2.65%	1.88	11.87	6.63	30
2016	5	81.61	143	2.33%	0.88	10.19	5.74	30
2017	8	114.04	185	2.16%	0.25	8.31	4.7	30
2018	4	151,77	227	2.3%	0.83	7.23	3.90	30
2019	29	177.98	460	2.25%	0.025	5.8 4	2.88	30

The data cover the period from January 2, 2013, to December 31, 2019. Dates with fewer than four traded securities (8 out of 1,752 days) were excluded.

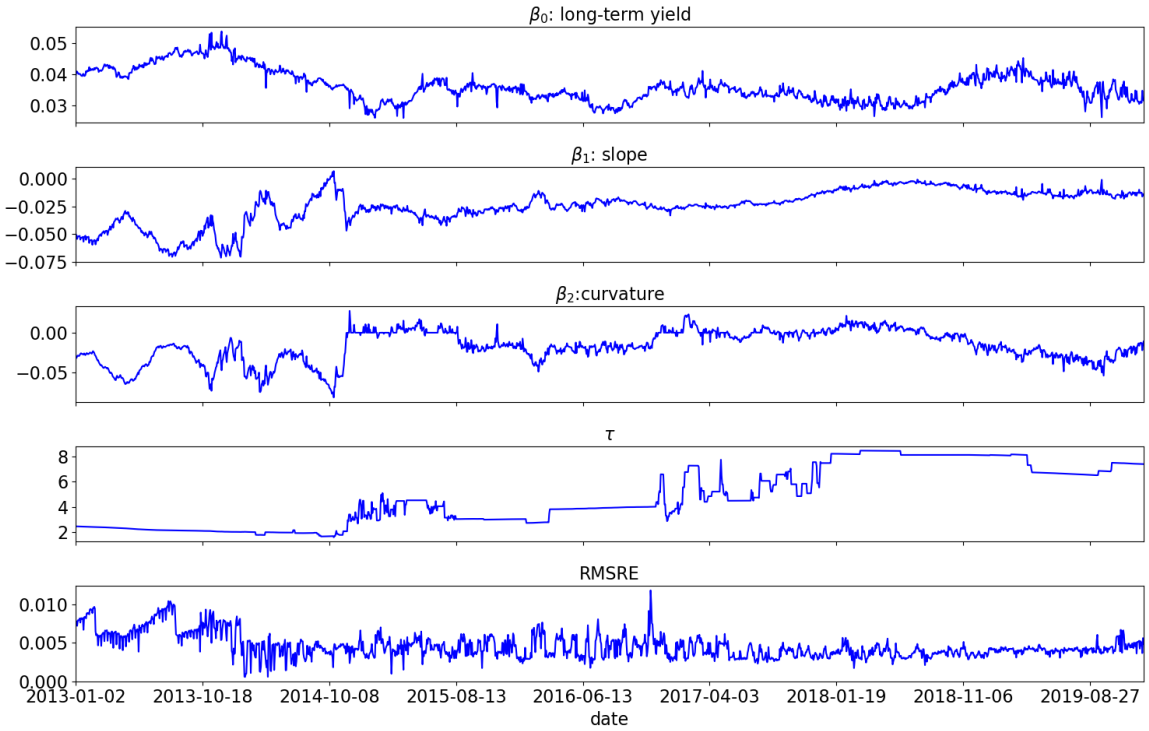
a concentration of bonds at shorter maturities. Despite these observations, bonds span the full range of the zero-coupon curve up to 30 years.

The time series of the estimated U.S. Treasury Nelson-Siegel coefficients are presented in Figure 3.1. The estimated coefficients are stable across the sample period, with the long-term yield coefficient (β_0) decreasing over time. The slope coefficient (β_1) approaches zero beginning in 2018, reflecting a flattening yield curve. The τ coefficient increases during 2018 and 2019, coinciding with periods when the U.S. Treasury zero-coupon yield curve flattened or became inverted.

Figure 3.2 presents the estimated zero-coupon yield surface for U.S. Treasury bonds over the 2013–2019 period. The surface reveals negative short-term rates in 2013, reflecting both historically low short-term yields and the known limitations of the Nelson-Siegel model in capturing yields at very short maturities, exacerbated by the scarcity of very short-term bonds in the data. Starting in 2018, the yield curve flattens and transitions into an inversion observed in 2019. These dynamics mirror the evolution of the Nelson-Siegel coefficients shown in Figure 3.1.

To assess model precision, we compute three accuracy measures: Mean Absolute Error

Figure 3.1. Time Series of the Estimated Nelson-Siegel Coefficients of the Zero-Coupon Yield Curve for U.S. Treasuries



The panels visually document estimated coefficients from January 2, 2013, to December 31, 2019. The coefficients are estimated using the two-step linearized NS approach on Equations (3.1) and (3.2), minimizing either Equation (3.5) or Equation (3.11) when the coefficients deviate beyond one standard error of previous estimates. From top to bottom: long-term yield (β_0), slope (β_1), curvature (β_2), hump position (τ), and RMSRE as defined in Equation (3.5).

(MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), defined as:

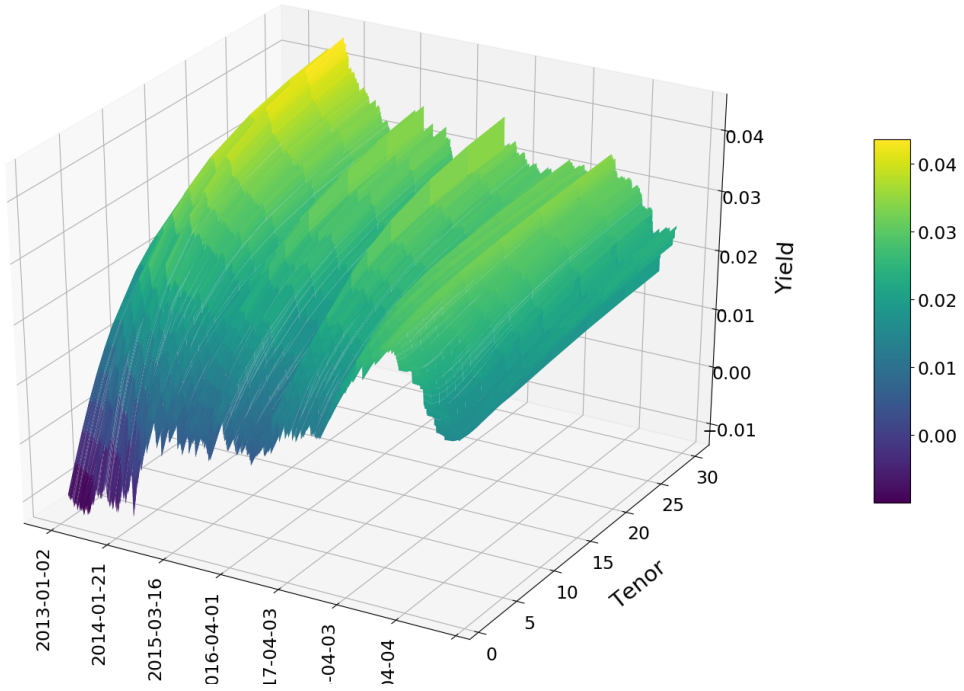
$$MAE = \frac{1}{N_t}\sum_{i=1}^{N_t}|\widehat{B}_i - B_i|, \tag{3.7}$$

$$MSE = \frac{1}{N_t}\sum_{i=1}^{N_t}\left(\widehat{B}_i - B_i\right)^2, \tag{3.8}$$

$$RMSE = \sqrt{MSE}. \tag{3.9}$$

Table 3.2 presents these results by maturity bucket. Accuracy improves as maturity

Figure 3.2. Surface of the Estimated U.S. Treasury Yield Curve from 2013 to 2019



The surface is computed using the estimated coefficients shown in Figure 3.1, covering January 2, 2013, to December 31, 2019.

shortens: the overall average absolute difference between estimated and observed prices is approximately \$17 cents, declining to less than \$0.11 for maturities up to ten years. Figure 3.3 depicts the distribution of pricing errors, which is leptokurtic with many small deviations and occasional large outliers.

3.3.1.3 Estimation of the Zero-Coupon Spread Curve for Corporate Issuers

To capture issuer-specific risks in corporate bond prices, we estimate the zero-coupon spread curve rather than the zero-coupon yield curve. This approach requires first estimating the zero-coupon yield curve for the corresponding government issuer. Indeed, Equation (3.1) can be decomposed to account for both government (risk-free) yields and

Table 3.2. Precision Errors Between the Estimated Price and the Observed Price by Bucket of Maturity for U.S. Treasury Bonds

	MAE	MSE	RMSE
$T \leq 1$	0.04	0.00	0.06
$1 < T \leq 2$	0.04	0.00	0.07
$2 < T \leq 5$	0.07	0.01	0.12
$5 < T \leq 10$	0.11	0.03	0.17
$10 < T \leq 15$	0.12	0.04	0.20
$15 < T \leq 20$	0.15	0.08	0.28
$20 < T \leq 25$	0.17	0.10	0.32
$25 < T \leq 31$	0.18	0.11	0.33
Overall	0.17	0.10	0.32

The results cover the period from January 2, 2013, to December 31, 2019. The estimated prices are computed using the estimated coefficients presented in Figure 3.1. Accuracy metrics are defined in Equations (3.7)–(3.9).

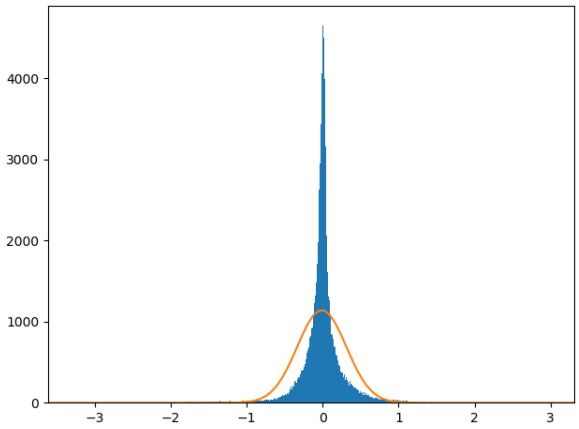
issuer-specific spreads as follows:

$$B_i = \sum_{j=1}^{m_i} c_{i,j} e^{-y^{ZC}(\theta^{(Gov)}, t_{i,j}) t_{i,j}} e^{-y^{ZC}(\theta^{(issuer)}, t_{i,j}) t_{i,j}}, \tag{3.10}$$

where $\theta^{(Gov)}$ is the coefficient set for the government's zero-coupon yield curve, and $\theta^{(issuer)}$ is the coefficient set for the issuer-specific zero-coupon spread curve.

The estimation procedure closely follows that for government issuers. We numerically optimize the issuer-specific spread coefficients using initial values obtained from a linearized Nelson-Siegel approach adapted explicitly for the zero-coupon spread (see Appendix B.2). After initial optimization, coefficient stability is assessed by verifying that each coefficient remains within one standard error of its value computed over the previous three-month rolling window. If any coefficient deviates beyond this threshold, we perform a penalized re-estimation by modifying the loss function to incorporate penalties based on deviations from the previous day's coefficient values and on prior information about the

Figure 3.3. Distribution of the Pricing Errors Between Estimated and Observed Prices from 2013 to 2019 for US (U.S. Treasury Bonds)



The histogram shows the distribution of errors between estimated and observed bond prices from January 2, 2013, to December 31, 2019. The orange line corresponds to a fitted Gaussian density.

plausible range of coefficients. The modified loss function is defined as:

$$\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{\widehat{B}_i - B_i}{B_i}\right)^2} + \sum_{j=0}^{2}\left(\widehat{\beta}_{j,t} - \beta_{j,t-1}\right)^2 + \left(\frac{\widehat{\tau}_t - \tau_{t-1}}{\tau_{t-1}}\right)^2 + \sum_{j=0}^{2}\left(\widehat{\beta}_{j,t} - \beta_{j,t-1}^{prior}\right)^2 \tag{3.11}$$

Since our primary objective is prediction, coefficient stability is crucial. Although many coefficient sets can generate the spread, only a few produce interpretable weight structures. Incorporating prior information on long-term spread levels, slope spreads, and curvature significantly enhances coefficient interpretability and stability. A complete description of the priors' construction is provided in Appendix C.1 of Chapter 4.

3.3.1.4 In-Sample Results for U.S. Corporate Issuers

We evaluate the performance of our approach using data on six U.S. corporate issuers: Goldman Sachs (ID=580), RBC[4] (ID=270), Johnson & Johnson (ID=939), The Walt Disney Company (ID=973), Amgen (ID=979), and Verizon (ID=995). Table 3.3 summarizes

[4]USD-denominated bonds.

the average daily number of traded bonds per issuer and year. Our analysis covers bond prices from 2015 through 2019, although data availability for RBC and Verizon begins in 2016 and mid-2017, respectively. RBC has a limited number of traded bonds per day compared to the other five firms, which have between 10 and 51 traded bonds per day.

Table 3.3. Average Number of Bonds per Issuer and Year

	RBC (ID=270)	Goldman Sachs (ID=580)	Johnson (ID=939)	Walt Disney (ID=973)	Amgen (ID=979)	Verizon (ID=995)
2015	-	12.27	9.81	10.22	17.44	-
2016	4.32	16.92	13.77	14.90	21.68	-
2017	6.38	20.75	18.30	18.53	28.15	40.78
2018	7.85	22.46	23.44	19.55	29.31	41.67
2019	9.67	19.26	23.67	51.05	29.88	42.98

The data cover the period from January 2, 2015, to December 31, 2019. Data for RBC and Verizon begin in 2016 and mid-2017, respectively. Dates with fewer than four traded securities are excluded.

Figure 3.4 shows the time series of Mean Absolute Error and Mean Squared Error over the sample period for each issuer. RBC has fewer traded bonds and exhibits lower MAE and MSE values, indicating higher estimation accuracy. Conversely, Verizon has numerous traded bonds per day and shows higher MAE and MSE, especially in 2019. Verizon's performance divergence reflects substantial variability in yields across maturities within the dataset. For instance, on August 30, 2019, among the 13 Verizon bonds with maturities exceeding 20 years, yields ranged from 2.4% to 4.02%, with mean and median values around 3.4%. Variations in yield-to-maturity and, accordingly, in prices, are observed across different maturity buckets and trading dates, which complicates accurate price estimation and shows significant data-quality limitations within the dataset.

Appendix B.3 visually documents the estimated Nelson-Siegel spread coefficients and the corresponding root-mean-squared relative errors (RMSRE) for each issuer over the sample period. For the six issuers considered, the long-term spread coefficients range from 1% and 3.5%, while the slope coefficients tend to remain negative. The τ coefficient is relatively stable over time; notably, RBC, Walt Disney, and Verizon have τ between 1.5 and 3 years, while Goldman Sachs, Johnson & Johnson, and Amgen exhibit a higher and

broader range, from 4 to 8 years.

Figure 3.5 presents the surface of the Nelson-Siegel spread curve for each issuer. All issuers generally exhibit positively sloped (normal) spread curves. RBC and Johnson & Johnson display consistently lower yield spreads relative to the other issuers, although RBC's spreads rise during 2019. Given RBC's limited bond maturities, the estimation accuracy of its long-term spread coefficient might be compromised. We observe two periods where the spread increases: early 2016 and throughout 2019. Between these two periods, the spread reduces and the curve flattens, except for Verizon. These corporate credit spread dynamics differ markedly from the U.S. Treasury curve, which began flattening in 2018 and became inverted in 2019.

Table 3.4 summarizes the in-sample pricing errors (MAE, MSE, and RMSE) for each issuer. Similar conclusions to those observed in Figure 3.5 are reached. RBC achieves the highest estimation accuracy, benefiting from fewer traded bonds and reduced variability. In contrast, Verizon demonstrates substantially lower accuracy, reflecting extensive yield fluctuations across maturities, particularly evident in the 2019 data.

Table 3.4. In-Sample Precision Errors Between the Estimated Price and the Observed Price for U.S. Corporate Issuers

	MAE	MSE	RMSE
RBC (ID=270)	0.10	0.07	0.27
Goldman Sachs (ID=580)	0.50	0.69	0.83
Johnson (ID=939)	0.75	1.26	1.12
Walt Disney (ID=973)	1.02	3.04	1.74
Amgen (ID=979)	1.31	3.58	1.89
Verizon (ID=995)	2.29	20.07	4.48

The data cover the period from January 2, 2015, to December 31, 2019. Data for RBC and Verizon begin in 2016 and mid-2017, respectively. The statistics are computed using Eqs. (3.7)–(3.9).

3.3.2 Prediction of the Yield and Spread Curve

To forecast yield curves and bond prices, we first predict the dynamics of the Nelson-Siegel coefficients (i.e., the long-term yield, β_0; the slope, β_1; and the curvature, β_2) by modelling their joint evolution using a Vector Autoregressive (VAR) model:

$$\beta_{0,t} = \gamma_{0,0} + \sum_{i=1}^{AR} \gamma_{0,i}\beta_{0,t-i} + \varepsilon_{0,t}, \tag{3.12}$$

$$\beta_{1,t} = \gamma_{1,0} + \sum_{i=1}^{AR} \gamma_{1,i}\beta_{1,t-i} + \varepsilon_{1,t}, \tag{3.13}$$

$$\beta_{2,t} = \gamma_{2,0} + \sum_{i=1}^{AR} \gamma_{2,i}\beta_{2,t-i} + \varepsilon_{2,t}, \tag{3.14}$$

where t is the time period, AR is the selected autoregressive lag, and the residual terms $\varepsilon_{0,t}$, $\varepsilon_{1,t}$, and $\varepsilon_{2,t}$ are correlated. We use the estimated VAR parameters ($\hat{\gamma}_{j,i}$, with $j = 0, 1, 2$ and $i = 0, 1, ..., AR$) to obtain one- to h-step-ahead forecasts of each the Nelson-Siegel coefficients via its conditional expectations given the information available at time $t-1$:

$$\begin{aligned}
\mathbb{E}_{t-1}\left[\beta_{j,t}\right] &= \mathbb{E}_{t-1}\left[\hat{\gamma}_{j,0} + \sum_{i=1}^{AR} \hat{\gamma}_{j,i}\beta_{j,t-i} + \varepsilon_{j,t}\right] \\
&= \hat{\gamma}_{j,0} + \sum_{i=1}^{AR} \hat{\gamma}_{j,i}\mathbb{E}_{t-1}\left[\beta_{j,t-i}\right] + \mathbb{E}_{t-1}\left[\varepsilon_{j,t}\right] \\
&= \hat{\gamma}_{j,0} + \sum_{i=1}^{AR} \hat{\gamma}_{j,i}\beta_{j,t-i},
\end{aligned}$$

for $j = \{0, 1, 2\}$. Finally, these forecasted coefficients are then used to reconstruct the predicted yield curves and associated bond prices.

The first three Nelson-Siegel coefficients are linear; however, the fourth coefficient, τ, introduces nonlinear complexity. Due to Jensen's inequality, we cannot directly use the conditional expectation of τ to forecast the yield curve without introducing bias. Indeed, the Nelson-Siegel basis functions, $\phi_1(T, \tau)$ and $\phi_2(T, \tau)$ in Equations (3.3) and (3.4), exhibit nonlinear (concave or convex) relationships in maturity T and coefficient τ. The

function $\phi_1(T,\tau)$ is strictly concave, starting at 1 for $T = 0$ and decreasing monotonically to 0 as T increases; thus, smaller τ accelerates this decay. The function $\phi_2(T,\tau)$ is concave for smaller values of T and convex otherwise. Jensen's inequality implies that the expectation of a nonlinear function does not equal the function evaluated at the expectation, introducing a bias if the expectation of τ is used directly to forecast the yield curve:

$$\begin{aligned}
&\mathbb{E}_{t-1}\left[y_t^{ZC}(\theta,T)\right] \\
&= \mathbb{E}_{t-1}\left[\beta_{0,t} + \beta_{1,t}\phi_1(T,\tau) + \beta_{2,t}\phi_2(T,\tau)\right] \\
&= \mathbb{E}_{t-1}\left[\beta_{0,t}\right] + \mathbb{E}_{t-1}\left[\beta_{1,t}\phi_1(T,\tau)\right] + \mathbb{E}_{t-1}\left[\beta_{2,t}\phi_2(T,\tau)\right] \\
&\neq \mathbb{E}_{t-1}\left[\beta_{0,t}\right] + \mathbb{E}_{t-1}\left[\beta_{1,t}\right]\mathbb{E}_{t-1}\left[\phi_1(T,\tau)\right] + \mathbb{E}_{t-1}\left[\beta_{2,t}\right]\mathbb{E}_{t-1}\left[\phi_2(T,\tau)\right].
\end{aligned}$$

To mitigate this issue and avoid bias, we fix the value of τ at its most recently observed value and forecast only the linear coefficients (β_0, β_1, and β_2). Furthermore, given the strong correlations among the NS coefficients both within and across issuers, we explicitly model these correlations among each series of β_i ($i = 0, 1, 2$) within the VAR framework. Forecasting is performed using a Seemingly Unrelated Regression (SUR) approach. Further details on the SUR methodology and its implementation are provided in Appendix B.4.

3.3.2.1 Out-of-Sample Prediction Results for U.S. Treasury Bonds

To evaluate the predictive performance of our model, we conduct an out-of-sample analysis by dividing the dataset into two distinct subsets: an in-sample training set with observations from January 2015 to June 2018, and an out-of-sample test set with observations from July 2018 to December 2019. This split corresponds to a 70% in-sample and 30% out-of-sample division. Using the in-sample data, we estimate the VAR model parameters of the NS coefficients' dynamics, which are then employed to forecast these coefficients on the out-of-sample data.

To measure predictive accuracy, we compute the MAE between the predicted and observed bond prices for forecast horizons of one to five days ahead. Table 3.5 presents the out-of-sample MAE results for U.S. Treasury bonds across various maturity buckets.

Consistent with prior findings, prediction accuracy improves with shorter maturities and remains stable across all forecast horizons from one to five days ahead. Compared to the in-sample analysis, out-of-sample MAEs are generally stable, except for bonds with maturities between 25 and 31 years, where the MAE increases by approximately five cents, indicating slightly lower accuracy at the long end of the yield curve.

Table 3.5. Mean Absolute Error for X-day Ahead Out-of-Sample Price Predictions for U.S. Treasury Bonds

	1-day	2-day	3-day	4-day	5-day
$T \leq 1$	0.0400	0.0397	0.0398	0.0399	0.0395
$1 < T \leq 2$	0.0539	0.0540	0.0545	0.0552	0.0552
$2 < T \leq 5$	0.0879	0.0903	0.0917	0.0934	0.0944
$5 < T \leq 10$	0.1207	0.1243	0.1261	0.1293	0.1304
$10 < T \leq 15$	0.1224	0.1262	0.1283	0.1316	0.1328
$15 < T \leq 20$	0.1424	0.1464	0.1497	0.1542	0.1557
$20 < T \leq 25$	0.1735	0.1779	0.1822	0.1882	0.1922
$25 < T \leq 31$	0.2251	0.2286	0.2346	0.2428	0.2467
Overall	0.2068	0.2099	0.2153	0.2227	0.2261

Mean absolute error by maturity bucket between forecasted and observed bond prices for U.S. Treasury bonds. Predictions are generated one to five days ahead using VAR parameters estimated from the in-sample dataset (January 2015-June 2018) and evaluated over the out-of-sample period (July 2018-December 2019).

3.3.2.2 Out-of-Sample Precision Errors for One- to Five-Days-Ahead Predicted Psrices for Corporates

s We apply the same out-of-sample forecasting framework to our six U.S. corporate issuers. The dataset from January 2015 through June 2018 serves as the in-sample estimation window, while July 2018 through December 2019 constitutes the out-of-sample test period (adjusted slightly for Verizon and RBC to account for their later data start dates). We generate one- to five-day-ahead forecasts of bond prices by propagating the estimated VAR–SUR

system of Nelson-Siegel coefficients and then reconstructing prices via the coupon-bond pricing formula. Predictive accuracy is measured by the Mean Absolute Error (MAE) between forecasted and realized prices. Tables 3.6 through 3.10 report the detailed MAE results for each issuer and forecast horizon.

We observe a decline in predictive accuracy as maturity and forecast horizon increase. Comparing the out-of-sample MAE with the in-sample MAE (cf. Table 3.4), Goldman Sachs's predictive performance surpasses its in-sample accuracy, while RBC, Johnson & Johnson, and Amgen exhibit out-of-sample accuracy comparable to their respective in-sample levels; however, Walt Disney and Verizon experience decreased accuracy. These conclusions hold consistently across one- to five-day forecast horizons.

Figure 3.6 presents the distribution of one-day-ahead forecast errors between forecasted and observed prices for all six U.S. issuers. Each distribution exhibits leptokurtosis, characterized by many small errors concentrated around the mean and a limited number of extreme deviations. Verizon displays a significant number of outliers, reinforcing the potential data-quality issues noted earlier. Given these findings, further verification of data-integration processes within the database is recommended to enhance predictive reliability.

Table 3.6. One-Day Ahead Precision Errors (MAE) Between Forecasted and Observed Prices by Maturity Bucket for U.S. High-Frequency Issuers

	RBC (ID=270)	Goldman Sachs (ID=580)	Johnson (ID=939)	Walt Disney (ID=973)	Amgen (ID=979)	Verizon (ID=995)
$T \leq 1$	0.185	0.099	0.116	0.070	0.216	-
$1 < T \leq 2$	0.105	0.142	0.131	0.139	0.185	0.210
$2 < T \leq 5$	0.158	0.207	0.247	0.328	0.310	0.485
$5 < T \leq 10$	0.180	0.284	0.515	1.016	1.848	1.848
$10 < T \leq 15$	-	1.345	1.256	1.179	-	2.863
$15 < T \leq 20$	-	-	1.235	2.724	3.610	3.268
$20 < T \leq 25$	-	1.807	1.726	1.584	1.876	2.845
$25 < T \leq 31$	-	0.889	1.029	2.830	1.765	4.340
Overall	0.144	0.348	0.744	1.319	1.334	2.688

MAE is defined in Eq.(3.7). The in-sample data set covers the period of January 2015 to June 2018. The out-of-sample data set covers the period of July 2018 to December 2019. The out-of-sample set is used to test the precision of the forecasted prices versus the observed prices.

Table 3.7. Two-Day Ahead Precision Errors (MAE) Between Forecasted and Observed Prices by Maturity Bucket for U.S. High-Frequency Issuers

	RBC (ID=270)	Goldman Sachs (ID=580)	Johnson (ID=939)	Walt Disney (ID=973)	Amgen (ID=979)	Verizon (ID=995)
$T \leq 1$	0.185	0.099	0.115	0.069	0.216	-
$1 < T \leq 2$	0.105	0.142	0.130	0.139	0.184	0.209
$2 < T \leq 5$	0.164	0.211	0.249	0.330	0.313	0.487
$5 < T \leq 10$	0.175	0.301	0.522	1.020	1.852	1.856
$10 < T \leq 15$	-	1.333	1.276	1.192	-	2.860
$15 < T \leq 20$	-	-	1.240	2.710	3.602	3.272
$20 < T \leq 25$	-	1.821	1.732	1.591	1.877	2.851
$25 < T \leq 31$	-	0.896	1.048	2.836	1.776	4.362
Overall	0.147	0.354	0.750	1.321	1.337	2.694

MAE is defined in Eq.(3.7). The in-sample data set covers the period of January 2015 to June 2018. The out-of-sample data set covers the period of July 2018 to December 2019. The out-of-sample set is used to test the precision of the forecasted prices versus the observed prices.

Table 3.8. Three-Days Ahead Precision Errors (MAE) Between Forecasted and Observed Prices by Maturity Bucket for U.S. High-Frequency Issuers

	RBC (ID=270)	Goldman Sachs (ID=580)	Johnson (ID=939)	Walt Disney (ID=973)	Amgen (ID=979)	Verizon (ID=995)
$T \leq 1$	0.185	0.099	0.115	0.069	0.215	-
$1 < T \leq 2$	0.107	0.143	0.130	0.140	0.183	0.209
$2 < T \leq 5$	0.168	0.221	0.252	0.332	0.316	0.492
$5 < T \leq 10$	0.170	0.328	0.529	1.028	1.858	1.864
$10 < T \leq 15$	-	1.360	1.291	1.192	-	2.858
$15 < T \leq 20$	-	-	1.251	2.692	3.589	3.277
$20 < T \leq 25$	-	1.832	1.732	1.606	1.883	2.857
$25 < T \leq 31$	-	0.937	1.070	2.851	1.796	4.390
Overall	0.149	0.367	0.757	1.325	1.344	2.702

MAE is defined in Eq.(3.7). The in-sample data set covers the period of January 2015 to June 2018. The out-of-sample data set covers the period of July 2018 to December 2019. The out-of-sample set is used to test the precision of the forecasted prices versus the observed prices.

Table 3.9. Four-days Ahead Precision Errors (MAE) Between Forecasted and Observed Prices by Maturity Bucket for U.S. High-Frequency Issuers

	RBC (ID=270)	Goldman Sachs (ID=580)	Johnson (ID=939)	Walt Disney (ID=973)	Amgen (ID=979)	Verizon (ID=995)
$T \leq 1$	0.184	0.099	0.115	0.068	0.214	-
$1 < T \leq 2$	0.107	0.144	0.129	0.140	0.183	0.208
$2 < T \leq 5$	0.172	0.227	0.253	0.334	0.318	0.493
$5 < T \leq 10$	0.173	0.345	0.535	1.032	1.861	1.866
$10 < T \leq 15$	-	1.376	1.309	1.213	-	2.861
$15 < T \leq 20$	-	-	1.265	2.690	3.585	3.287
$20 < T \leq 25$	-	1.850	1.750	1.627	1.897	2.868
$25 < T \leq 31$	-	0.985	1.089	2.868	1.802	4.406
Overall	0.151	0.378	0.766	1.333	1.349	2.709

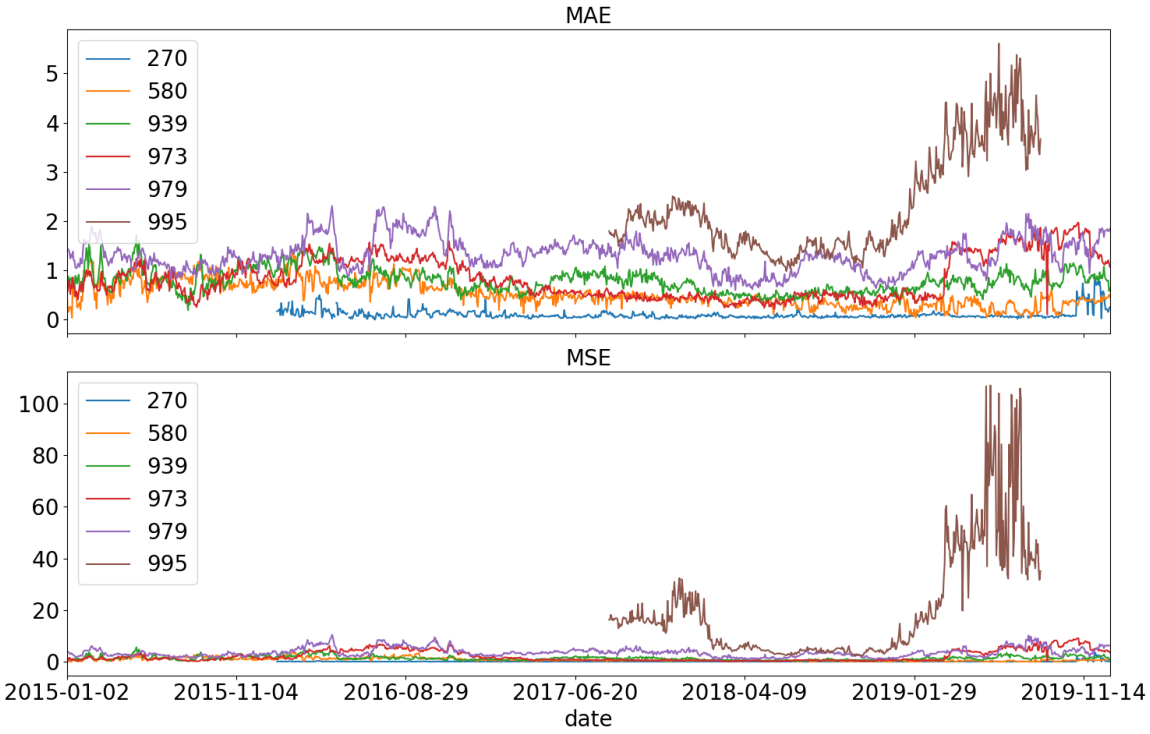
MAE is defined in Eq.(3.7). The in-sample data set covers the period of January 2015 to June 2018. The out-of-sample data set covers the period of July 2018 to December 2019. The out-of-sample set is used to test the precision of the forecasted prices versus the observed prices.

Table 3.10. Five-days Ahead Precision Errors (MAE) Between Forecasted and Observed Prices by Maturity Bucket for U.S. High-Frequency Issuers

	RBC (ID=270)	Goldman Sachs (ID=580)	Johnson (ID=939)	Walt Disney (ID=973)	Amgen (ID=979)	Verizon (ID=995)
$T \leq 1$	0.185	0.099	0.114	0.068	0.214	-
$1 < T \leq 2$	0.108	0.146	0.129	0.140	0.183	0.207
$2 < T \leq 5$	0.175	0.232	0.255	0.335	0.318	0.493
$5 < T \leq 10$	0.173	0.364	0.538	1.034	1.864	1.866
$10 < T \leq 15$	-	1.361	1.321	1.219	-	2.863
$15 < T \leq 20$	-	-	1.275	2.683	3.582	3.299
$20 < T \leq 25$	-	1.875	1.762	1.639	1.906	2.874
$25 < T \leq 31$	-	0.987	1.094	2.873	1.804	4.414
Overall	0.153	0.385	0.771	1.336	1.352	2.715

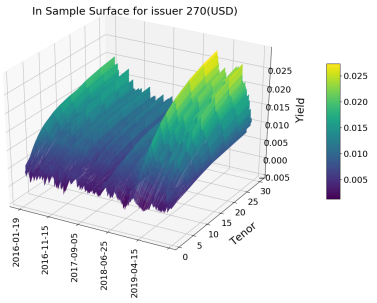
MAE is defined in Eq.(3.7). The in-sample data set covers the period of January 2015 to June 2018. The out-of-sample data set covers the period of July 2018 to December 2019. The out-of-sample set is used to test the precision of the forecasted prices versus the observed prices.

Figure 3.4. MAE and MSE for U.S. Corporate Issuers from 2015 to 2019

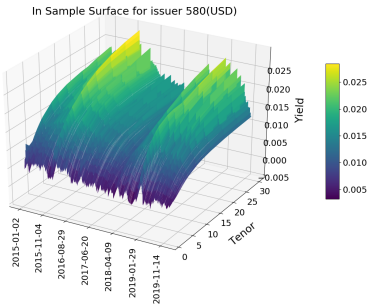


MAE and MSE are defined in Eq. (3.7) and Eq. (3.8), respectively. The issuers are RBC (ID=270), Goldman Sachs (ID=580), Johnson & Johnson (ID=939), Walt Disney (ID=973), Amgen (ID=979), and Verizon (ID=995). The data cover January 2, 2015, to December 31, 2019; data for RBC and Verizon begin in 2016 and mid-2017, respectively.

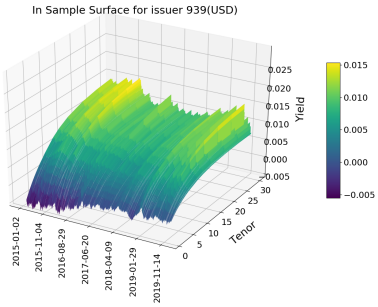
Figure 3.5. Surface of Credit Spreads for U.S. Corporate Issuers for 2015 to 2019



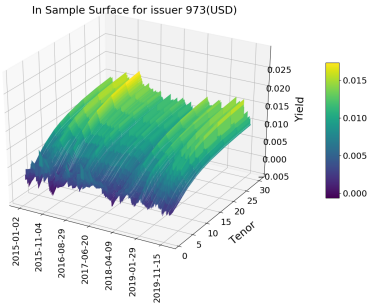
RBC (ID=270)



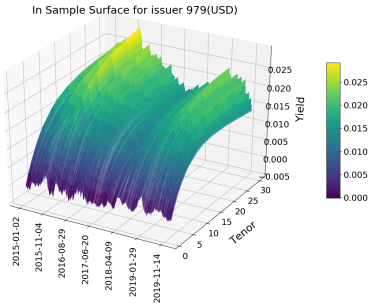
Goldman Sachs (ID=580)



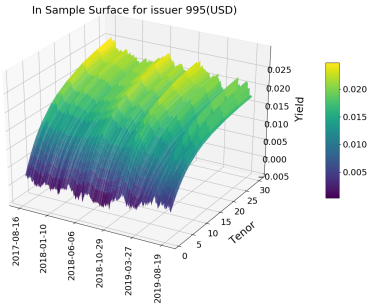
Johnson & Johnson (ID=939)



The Walt Disney Company (ID=973)



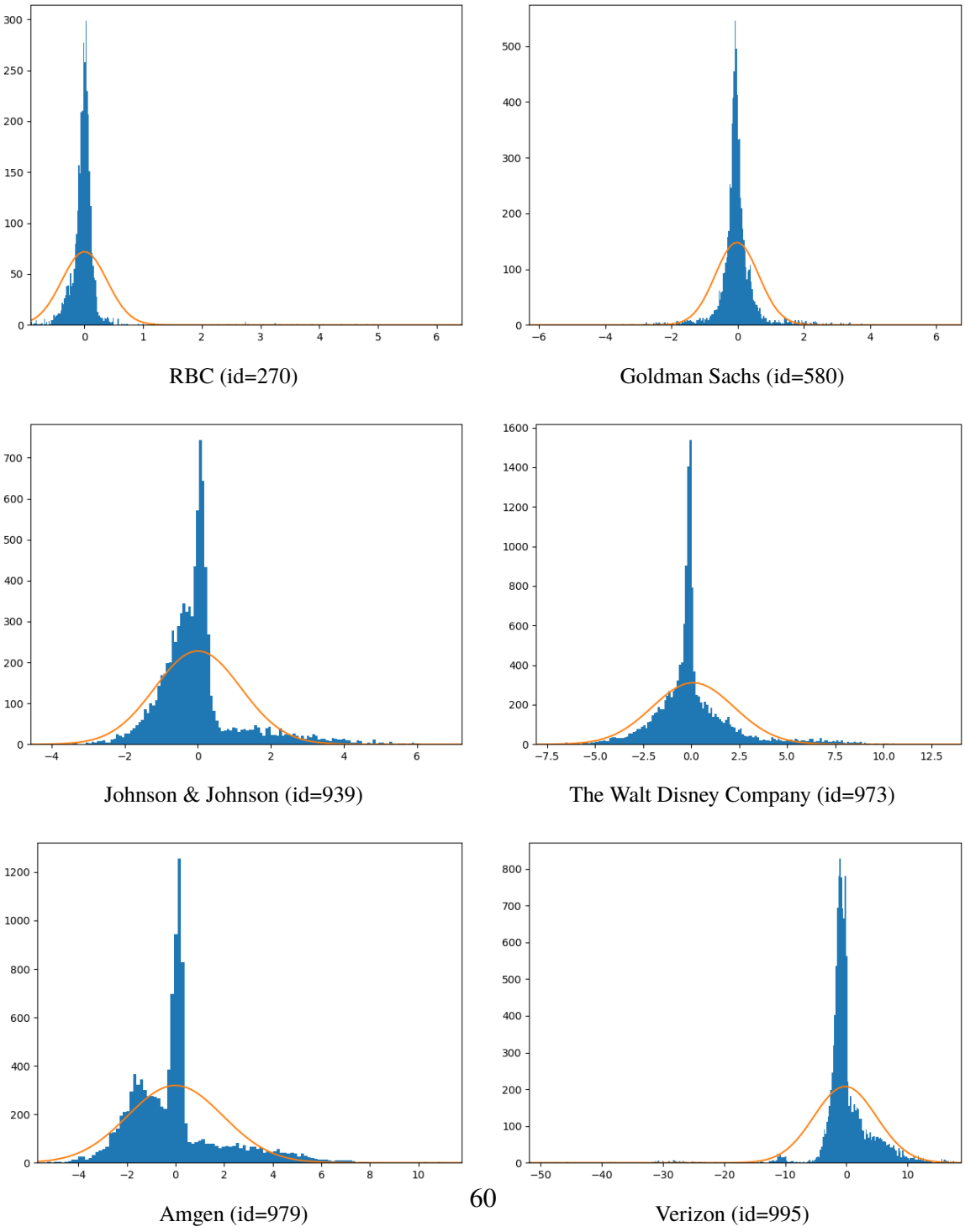
Amgen (ID=979)



Verizon (ID=995)

Zero-coupon spread curves are computed using the estimated coefficients presented in Appendix B.3, which are obtained via the Nelson-Siegel two-step approach. The data cover January 2, 2015, to December 31, 2019; data for RBC and Verizon begin in 2016 and mid-2017, respectively.

Figure 3.6. Distributions of the Errors Between Forecasted and Observed Prices for U.S. High-Frequency Issuers



Out-of-sample errors between the forecasted prices and the observed prices from July 2018 to December 2019. The orange line corresponds to a Gaussian distribution with the parameters (mean and variance) of the error distribution.

3.4 Prediction for Low-Frequency Issuers or New Issuers: Clustering Approach

In this section, we present the two-step methodology used to forecast credit spread curves for low-frequency issuers. Our approach involves, firstly, clustering high-frequency issuers exhibiting similar credit-risk profiles through a Gaussian Mixture Model. Secondly, we match low-frequency issuers to the appropriate high-frequency issuer clusters based on issuer characteristics, allowing us to use the cluster-specific credit spread curves as proxies. These proxy curves are then leveraged to predict credit spread curves for low-frequency issuers. Empirical results follow.

3.4.1 Gaussian Mixture Model

We employ an unsupervised clustering methodology to forecast credit spread curves where limited bond data prevents the direct estimation of credit spread curves. We start by estimating the Nelson–Siegel zero-coupon credit spread curves on subsets of issuers having at least five traded bonds per day,[5] covering selected periods: April 2017, August 2017, December 2017, and June 2018. This approach ensures representative spread curves across distinct market conditions.

In addition to the credit spread term structure values, we incorporate issuer-specific characteristics into the database. Firstly, we classify bonds by seniority into four categories: unsecured, secured, other unsecured,[6] and other secured.[7] Secondly, we standardize credit ratings across the three primary rating agencies (S&P, Fitch, and Moody's) into comparable categories as described in Table 3.11. Thirdly, we enrich our dataset by including one-year

[5]Given the four-parameter Nelson–Siegel model, a minimum of four bonds per issuer-day is required to ensure identifiability; we impose a slightly stricter threshold of five bonds to ensure sufficient data points for robust estimation.

[6]The category "other unsecured" aggregates bond seniorities labeled as subordinated unsecured, senior subordinated unsecured, junior unsecured, junior subordinated unsecured, senior non-preferred unsecured, or similar subordinated rankings.

[7]The category "other secured" aggregates bonds labeled senior secured mortgage, senior secured first mortgage, senior secured first lien, and other secured classes.

and five-year default probabilities obtained from Bloomberg. Default probabilities offer a continuous, nonlinear metric of credit risk, providing more granular insights than discrete credit ratings alone. We end up with a database of $43,507$ rows.

Table 3.11. Cluster of Credit Rating Note in Function of the Rating Agency

Cluster of Credit Rating	S&P	Fitch 500	Moody's
Prime		AAA	Aaa
High Grade		AA+/AA/AA-	Aa1/Aa2/Aa3
Upper Medium Grade		A+/A/A-	A1/A2/A3
Lower Medium Grade		BBB+/BBB/BBB-	Baa1/Baa2/Baa3
Non Investment Grade Speculative		BB+/BB/BB-	Ba1/Ba2/Ba3
Highly Speculative		B+/B/B-	B1/B2/B3
Substantial risks		CCC+	Caa1
Extremely Speculative		CCC	Caa2
Default imminent		CC+/CC/CC-	Caa3 Ca
In Default		D	C
NR		missing or unrated	

Comparison of the credit rating note by rating agency.

Since issuer clusters are unobservable, we apply three unsupervised learning methods to cluster corporate issuers based on similarities in their estimated spread curves: K-Means, K-Prototypes, and Gaussian Mixture Models (GMM). These unsupervised algorithms identify patterns in the data by grouping issuers with similar profiles into clusters while maximizing differentiation between distinct clusters. Due to superior predictive results, we subsequently focus exclusively on the Gaussian Mixture Model.

The Gaussian Mixture Model aims to identify a mixture of multidimensional Gaussian probability distributions that best fit the dataset, treating cluster membership as a latent variable. Given a set of N observations $S = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$, we aim to partition the data into K distinct clusters $C = \{C_1, C_2, ..., C_K\}$. For each observation x_n, the conditional probability distribution given membership in cluster k is assumed to follow a multivariate

Gaussian distribution:

$$f(\mathbf{x_n} \mid C_n = k; \theta) = \frac{1}{\sqrt{2\pi\sigma_k}} \exp - \left(\frac{\mathbf{x}_n - \mu_k}{\sigma_k} \right)^2,$$

where $\theta = \{\mu_k, \sigma_k\}$ denotes the parameter set, μ_k and σ_k represent the mean vector and covariance matrix of cluster k, for $k = 1, 2, ..., K$, and C_n is the cluster of observation n. The probability of an observation belonging to a cluster k is modeled by:

$$p\left(C_n = k\right) = \pi_k,$$

for $k = 1, 2, ..., K$. Using Bayes' theorem, the joint distribution of each observation and its latent cluster membership is thus given by:

$$f(\mathbf{x}_n, C_n; \theta) = p\left(C_n = k\right) f(\mathbf{x}_n \mid C_n = k; \theta).$$

The marginal distribution of each observation is obtained by summing the joint distribution over all clusters, which gives

$$p\left(\mathbf{x}_n\right) = \sum_{k=1}^{K} p\left(C_n = k\right) f\left(\mathbf{x}_n \mid C_n = k; \theta\right).$$

Thus, the Gaussian mixture distribution is expressed as a linear combination of Gaussian distributions. The corresponding log-likelihood function is:

$$\mathcal{L}\left(\theta \mid \mathbf{x}\right) = \log p\left(\mathbf{x}; \theta\right) = \sum_{n=1}^{N} \log \left\{ \sum_{k=1}^{K} p\left(C_n = k; \theta\right) f\left(\mathbf{x}_n \mid C_n = k; \theta\right) \right\}.$$

We estimate the parameters of the GMM model via the Expectation-Maximization (EM) algorithm proposed by Dempster et al., 1977, which iteratively maximizes the likelihood by alternating between expectation and maximization steps. The first step of the EM algorithm, the expectation step, evaluates the conditional expectation of the log-likelihood, i.e., the likelihood that the data belongs to cluster k given a set of parameters (or initial values), that is

$$\mathcal{Q}\left(\theta, \theta^{old}\right) = \mathbb{E}[\log\{f\left(\mathbf{x}, C = k; \theta\right) \mid \mathbf{x}; \theta^{old}\}]$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} p\left(C_n = k \mid \mathbf{x}_n; \theta^{old}\right) \log\{f\left(\mathbf{x}_n, C_n = k; \theta\right)\}$$

where θ^{old} represents the current parameter estimates. The second step, the maximization step, finds the new estimates of the parameters that maximize this expectation:

$$\theta^{new} = \arg\max_{\theta} \mathcal{Q}\left(\theta, \theta^{old}\right)$$

These steps are repeated until convergence. A detailed description of the EM algorithm implementation can be found in Bishop, 2006.

To construct clusters of high-frequency issuers based on credit risk, we utilize corporate credit spreads at selected tenors (i.e., 1, 2, 3, 4, 5, 7, 10, 12, 15, 17, 20, 25, and 30 years), both with and without incorporating one-year and five-year default probabilities. The inclusion of default probabilities, sourced from Bloomberg, is motivated by their ability to capture nonlinearities in issuer-specific default risk. Issuer characteristics (i.e., bond seniority, sector, and aggregated credit ratings from S&P, Fitch, and Moody's) are included to ensure economically meaningful cluster identification.

To select the optimal number of clusters K, we use the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC), defined respectively as:

$$\text{BIC} = q \ln(n) - 2\mathcal{L}\left(\theta \mid \mathbf{x}\right), \tag{3.15}$$

$$\text{AIC} = 2q - 2\mathcal{L}\left(\theta \mid \mathbf{x}\right), \tag{3.16}$$

where q is the total number of parameters estimated, n is the number of observations, and $\mathcal{L}\left(\theta \mid \mathbf{x}\right)$ is the maximized log-likelihood. The optimal number of clusters K is the one minimizing these information criteria.

Table 3.12 summarizes the optimal number of clusters suggested by each criterion, both with and without incorporating default probabilities. Consistent with expectations, the BIC tends to suggest fewer clusters (18) compared to the AIC (47).

The decision regarding the optimal number of clusters balances statistical fit against economic interpretability and practical forecasting considerations. While models with more clusters can capture finer nuances in issuer-specific risk profiles, a large number of

Table 3.12. Optimal Number of Clusters Based on Information Criteria for the Gaussian Mixture Model

Data	AIC	BIC
Credit spread	40	9
Credit spread + default probabilities	47	18

AIC and BIC are based on the log-likelihood. AIC is defined in Equation (3.16), and BIC is defined in Equation (3.15).

clusters may result in sparsity, with some clusters potentially having few observations or even no available data during forecast periods. In our data, employing 18 clusters provides better predictive performance than 47 clusters. A disproportionate representation among clusters has been identified: five clusters represent approximately 80% of the data. The weight of the other clusters is less than 3.5%. A pie chart illustrating the proportion of each cluster is provided in Appendix B.5.1. Some clusters contain a very limited number of bonds, presenting challenges for subsequent out-of-sample forecasting.

3.4.2 Prediction Based on the Clustering

The GMM provides posterior probabilities of belonging to each cluster for each traded bond in the database. We aggregate these posterior probabilities into a confusion matrix summarizing the likelihood of being in cluster k based on sector, seniority, and credit rating. For each bond to be priced, we match its triplet (sector, seniority, and credit rating) against the confusion matrix, identifying the cluster that best characterizes its risk profile. The confusion matrix is available in Appendix B.5.2. The advantage of this approach is its ability to produce price predictions even when no bonds with identical characteristics (sector, seniority, or credit rating) have recently traded, as each cluster typically aggregates multiple issuer profiles sharing common risk characteristics.

For each tenor and each cluster, we compute the average credit spread curve and its predictive confidence intervals. Formally, for $x \in C_k$ and for $t = \{1, 2, ..., 30\}$, we define

the mean credit spread $\bar{x}_{k,t}$ at tenor t for cluster k as follows:

$$y_{k,t} = \bar{x}_{k,t}, \tag{3.17}$$

and construct the predictive interval for the spread as

$$CI_{k,t} = y_{k,t} \pm t_{\alpha/2,n_i-1}\sigma_{k,t}, \tag{3.18}$$

where $t_{\alpha/2,n_k-1}$ is the critical value of the Student's t-distribution at significance level α, and $\sigma_{k,t}$ is the standard deviation of spreads within cluster k a tenor t. We assume the absence of correlation between the clusters.

Bond prices are then predicted using the government zero-coupon yield curve combined with the weighted-average spread curve of the assigned clusters. This leads to:

$$\hat{B}_i = \sum_{j=1}^{m_i} c_{i,j} e^{-y^{ZC}(\theta^{(Gov)},t_{i,j}),t_{i,j}} e^{-\sum_{k=1}^{K}\left(w_k y_{k,t} t_{i,j}\right)}, \tag{3.19}$$

where $\theta^{(Gov)}$ is the coefficients set of the government ZC yield curve, and w_k is the posterior probability of belonging to cluster k.

3.4.3 Low-Frequency Issuer Predicted Bond Price Results

We have considered a set of 114 newly issued bonds in March and April 2019, and we were able to predict 101 of them. For the remaining 13, issuer-specific information was missing. To evaluate the predictive accuracy of the proposed clustering approach, we assess how frequently the observed bond issuance prices fall within the calculated predictive intervals (i.e., Eq. (3.18)): 97 bonds fall within the predictive interval. We also examine absolute pricing errors by evaluating whether the differences between predicted and actual issuance prices are within 50 and 1 dollar thresholds. These thresholds were chosen to be consistent with the MAE observed during the estimation of credit spreads and the prediction of credit spread curves for high-frequency issuers using the VAR model. 19 bonds have differences of less than 50 cents from the observed price, while 27 bonds have differences of less than 1 dollar from the observed price.

Table 3.13 summarizes the standard errors of the predictive intervals across maturity buckets. The table shows that predictive intervals widen as the maturity of the bonds increases. Longer-maturity bonds entail greater uncertainty due to less frequent trading and heightened sensitivity to underlying economic and credit-risk assumptions, as reflected by larger standard errors.

Table 3.13. Descriptive Statistics of Standard Errors for Predictive Intervals by Tenor

Tenor	count	mean	std	min	50%	max
$2 < T \leq 5$	15	1.29	0.44	0.63	1.10	2.15
$5 < T \leq 10$	34	3.56	2.21	1.54	2.61	9.67
$10 < T \leq 20$	26	6.27	4.35	3.11	4.79	24.03
$20 < T \leq 30$	26	8.71	3.35	0.60	9.02	13.00

Observations correspond to bonds issued in March and April 2019. The prediction intervals are derived from estimated GMM cluster spreads using Equation (3.17) and Equation (3.18). Specifically, using the average credit spread and predictive interval credit spread, we compute the predicted price and the corresponding bounds. Finally, knowing $t_{\alpha/2,n_i-1}$ we obtain:

$$SE_i = \frac{\hat{B}_i - B}{t_{\alpha/2,n_i-1}}, \tag{3.20}$$

where $\hat{B}_i$ is computed using Equation (3.19), and B is the observed issuance price.

Table 3.14 reports the mean absolute prediction errors by cluster and maturity bucket. For each low-frequency issuer, the cluster presented is the one with the highest weight in computing the predicted credit spread. Consistent with theoretical expectations, predictive accuracy generally deteriorates as bond maturity increases. Short- to intermediate-term maturities (2–5 and 5–10 years) exhibit relatively lower mean absolute errors, whereas maturities beyond 20 years display substantially higher errors. For example, bonds in Cluster 11 exhibit large deviations (average absolute error exceeding 28 dollars) at maturities beyond 20 years. Cluster 10 maintains prediction accuracy consistently below 1 dollar, as does one tenor bucket for clusters 13 and 0. These three clusters collectively account for 63% of the data from high-frequency issuer clusters. Thus, clusters with substantial representation in the dataset tend to exhibit higher forecasting accuracy, while clusters with

Table 3.14. Mean Absolute Prediction Errors by Cluster and by Bucket of Tenor for Low-Issuers

Cluster	$2 < T \leq 5$	$5 < T \leq 10$	$10 < T \leq 20$	$20 < T \leq 30$
0	1.72	5.92	0.56	5.11
1	-	5.49	5.07	-
3	-	1.30	-	-
4	-	4.46	-	-
5	-	1.08	3.47	2.96
10	1.04	0.87	0.26	0.71
11	-	-	1.43	28.13
12	-	3.01	7.14	-
13	0.94	1.54	4.99	4.14
16	-	1.64	2.34	-

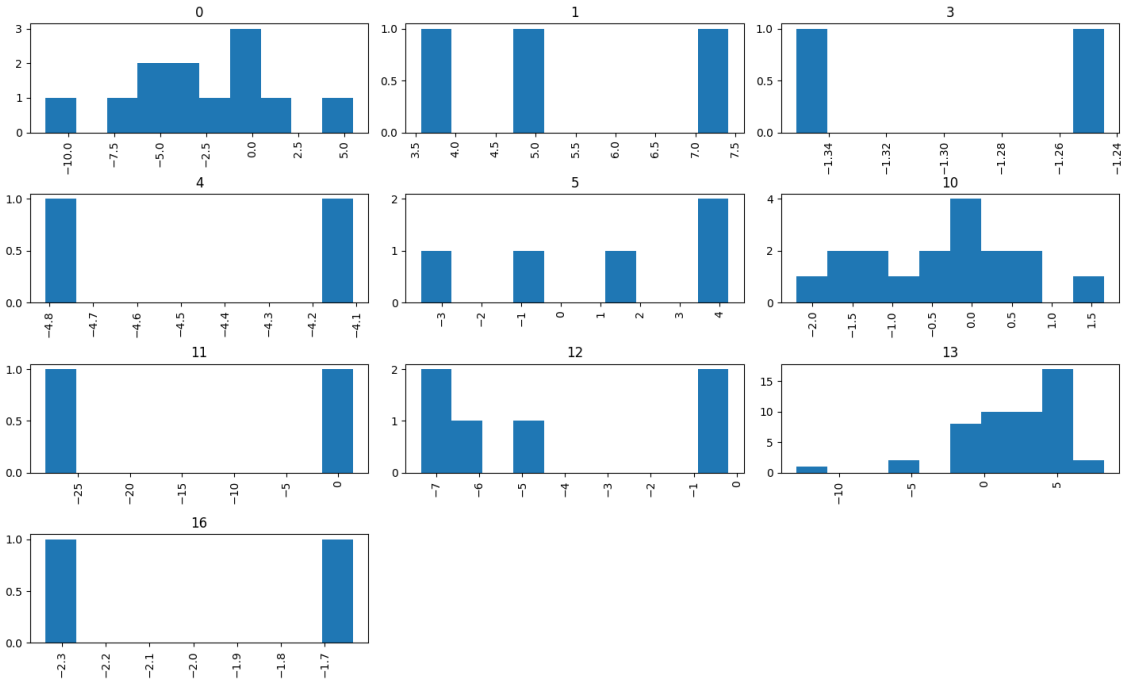
Mean absolute errors are calculated as absolute deviations between observed and predicted bond issuance prices. The predicted price is computed using Equation (3.19). Missing entries indicate no observations within the respective cluster-tenor combination. For each low-frequency issuer, the cluster presented is the one with the highest weight in computing the predicted credit spread. Predicted prices are computed using Equation (3.19). Data cover bond issuances from March and April 2019.

fewer observations, particularly at longer maturities, display larger prediction errors. This highlights potential challenges in accurately forecasting credit spreads for sparse clusters, which is further confirmed by the histogram of bond price errors shown in Figure (3.7).

The three clusters most represented among high-frequency issuers are also the most prevalent among low-frequency issuers. These clusters exhibit relatively symmetric distributions, though with varying dispersion levels. The remaining clusters, accounting for 21.8% of low-frequency issuers in March and April 2019, contain fewer observations, limiting interpretability and complicating robust conclusions. Cluster 11 notably displays extreme negative errors, with magnitudes significantly larger than those observed in other clusters. This suggests substantial prediction difficulties likely attributable to limited or inconsistent data within the cluster.

Overall, these results emphasize variability in predictive performance across clusters. They suggest that while the clustering approach effectively captures issuer heterogeneity for well-represented groups, predictive accuracy can deteriorate significantly in clusters

Figure 3.7. Histogram of the Prediction Errors by Cluster



Prediction error is the difference between the predicted price and the observed price for each cluster. For each low-frequency issuer, the cluster presented is the one that has the highest weight in the computation of the predicted credit spread.

with limited or heterogeneous observations. This indicates the potential need for further refining the clustering strategy or improving data quality.

3.5 Conclusion

This paper proposes and evaluates methodologies for forecasting corporate bond prices and credit spread term structures, addressing challenges related to varying bond trading frequencies across issuers. Utilizing a proprietary dataset, we apply distinct methodological approaches depending on issuer trading frequency. For high-frequency issuers, we employ a two-step Nelson-Siegel framework combined with a Vector Autoregressive (VAR)

model to predict issuer-specific zero-coupon credit spread curves and bond prices for forecast horizons ranging from one to five days ahead. The two-step approach leverages a linearized version of the Nelson-Siegel model with informative priors to generate stable initial estimates. To ensure coefficient stability over time, a penalty term is included in the optimization loss function whenever model coefficients deviate by more than one standard error from prior estimates. Subsequently, a VAR model is used to predict the Nelson-Siegel coefficients by capturing interdependencies among them. Empirical evidence based on USD-denominated corporate bonds indicates good predictive accuracy for short- and medium-term maturities. Precision notably declines for longer maturities, highlighting uncertainties associated with modeling longer-dated fixed-income securities.

For low-frequency issuers, we introduce an innovative clustering approach based on Gaussian Mixture Model (GMM). This methodology addresses data scarcity by aggregating issuer characteristics and market information derived from high-frequency issuers exhibiting similar risk profiles. The clustering model leverages posterior probabilities of cluster membership, conditioned on issuer characteristics such as sector, seniority, and credit ratings, to construct a weighted-average credit spread curve for each low-frequency issuer. Including issuer-specific default probabilities alongside credit spread data improves cluster identification and forecasting accuracy. Our analysis reveals heterogeneity in predictive performance across clusters: clusters with more observations yield more accurate price predictions, whereas clusters with fewer observations exhibit higher prediction errors. This emphasizes the necessity for cautious interpretation and further model refinement.

Throughout our analysis, we identified significant data-quality issues, including missing bond characteristics, the unflagged inclusion of simulated data, and substantial yield variability within clusters. These deficiencies impact predictive accuracy, highlighting the need for rigorous data validation protocols and increased transparency in data sourcing practices.

In summary, this research contributes methodologically to the literature on fixed-income analytics by proposing predictive frameworks tailored specifically to high-frequency and low-frequency corporate bond issuers. Future research could benefit from employing higher-quality datasets, thereby reducing data-driven noise and enhancing forecast reliability, and

extending our approach by incorporating additional issuer-specific attributes and macroeconomic indicators.

References

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

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

Gauthier, G., & Simonato, J.-G. (2012). Linearized nelson–siegel and svensson models for the estimation of spot interest rates. *European Journal of Operational Research*, *219*(2), 442–451.

Nelson, C., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, *60*(4), 473–89.

Sundaram, R. K., & Das, S. R. (2011). *Derivatives: Principles and practice*. McGraw-Hill Irwin New York, NY.

Chapter 4

Enhancing Credit Spread Forecasts through Macroeconomic Uncertainty Variables and Statistical Learning Approaches

Abstract[1]

This essay investigates the predictive modelling of the credit spread term structure by incorporating macroeconomic uncertainty indicators and statistical learning methodologies. We propose a new penalty loss function designed to enhance the economic interpretability and temporal stability of the estimated Nelson-Siegel credit spread coefficients. We then predict these coefficients by extending the set of features generally used in the literature to include macroeconomic and financial uncertainty measures. Finally, we compare linear regression models with ensemble-based machine learning techniques. Empirical analysis using SHapley Additive exPlanation (SHAP) values show the influence of firm-level variables and reinforce the link between economic uncertainty and credit spreads. Boosted trees outperform other approaches to predict the term structure of credit spread.

[1]Joint work with Geneviève Gauthier. Gauthier is affiliated with HEC Montréal.

4.1 Introduction

In the fixed-income market, credit risk reflects the possibility that a bond issuer may fail to meet its contractual obligations. The credit spread emerges as the premium that investors require to compensate for this risk. The term structure of credit spreads represents the yield spread between a risky bond and its risk-free benchmark across various maturities. Consequently, modelling and predicting this term structure is essential from both a risk-management perspective and a valuation perspective.

Credit spreads are not fixed but rather vary across maturities (Litterman & Iben, 1991), seniority levels, and credit-rating classes (Jarrow et al., 1997). The literature recognizes that these factors are not the sole determinants of credit-spread changes. For instance, Longstaff and Schwartz, 1995, Duffee, 1998, and Collin-Dufresne et al., 2001 document how changes in the spot rate and the slope of the yield curve drive fluctuations in credit spreads. Litterman, 1991 find that level and slope factors explain a substantial portion of yield-curve dynamics. Collin-Dufresne et al., 2001 underscore that aggregate variables, such as interest rates, business-cycle indicators, and market volatility, directly influence credit spreads through their impact on default expectations and recovery processes.

Firm-specific characteristics also play a role in determining credit-spread behaviour. Traditional asset-pricing models (e.g., the three-factor model of Fama and French 1993, 1996) reveal that size and book-to-market ratios capture risk components not explained by market betas. While initially developed in the equity context, these firm-level factors are also relevant to credit risk, given the close link between equity valuations and default risk (Jarrow, 2001; Kwan, 1996).

Credit spreads reflect default probability and loss given default, with the latter typically inversely related to recovery rates. Empirical evidence suggests that recovery rates are driven by idiosyncratic factors, such as the seniority of the debt or the economic sector in which the firm operates (Altman & Kalotay, 2014; Boudreault et al., 2013). Macroeconomic variables also exert an impact on recovery outcomes. For instance, Altman et al., 2005 shows that historical default rates negatively impact recovery levels. Recent contributions highlight economic uncertainty as a central systematic driver of both default and recovery (Gambetti et al., 2019), in line with structural and reduced-form credit risk mod-

els that link macroeconomic variables to default probabilities (Duffie & Singleton, 1999; Merton, 1974).

Macroeconomic factors in the bond literature have shown promising results. Ang and Piazzesi, 2003 found that incorporating observable macroeconomic variables, such as inflation and output growth, significantly improves the forecasting of Treasury bond yields compared to models relying solely on latent factors, especially at short and medium maturities. Recent macro-finance models of the term structure have shown that Treasury bond risk premia are driven by macroeconomic variables. Ludvigson and Ng, 2009 and Cooper and Priestley, 2009 document that various macroeconomic indicators contain information about future excess bond returns. Joslin et al., 2014 also show that variation in economic activity and inflation in the United States affects the term structure of U.S. Treasury yields. Nevertheless, most existing research focuses on Treasury or sovereign yield curves, with comparatively fewer studies investigating credit spreads. Amato and Luisi, 2006 show that sector- and rating-specific credit-spread dynamics are influenced by macroeconomic conditions. Despite these insights, the prediction of credit spreads via macroeconomic indicators and macroeconomic uncertainty variables remains limited. This study asks whether incorporating macroeconomic indicators and uncertainty measures can improve the predictive term structure of credit spreads.

Despite the growing recognition of uncertainty's importance, there is no universal consensus on how to define or measure it (Bloom, 2014; Knight, 1921; Zhang, 2006). Proxies range from financial-market volatilities (e.g., the VIX) to factor-based uncertainty measures derived from large datasets (Jurado et al., 2015; Ludvigson et al., 2021), each capturing different facets of the underlying economic environment.

Recent advances in computer science have made machine learning the new trend in finance. Numerous studies have explored its potential for asset pricing and risk-premium measurement (Bianchi et al., 2021; Gu et al., 2020), return predictability (Feng et al., 2018), and portfolio hedging (Carbonneau, 2021). Comparative analyses often conclude that tree-based ensemble models (e.g., random forests, gradient-boosted trees) and neural networks outperform traditional parametric regressions in a variety of forecasting tasks—such as recovery-rate prediction (Bellotti et al., 2021; Gambetti et al., 2022), loss-given-default

estimation (Hartmann-Wendels et al., 2014; Loterman et al., 2012; Qi & Zhao, 2011), excess Treasury bond-return prediction across different maturities (Bianchi et al., 2021), and credit scoring (Liu, Fan, & Xia, 2022; Liu, Fan, Xia, & Xia, 2022). Despite their predictive advantages, machine-learning models have faced criticism in financial economics for their "black-box" nature, which complicates parameter inference and interpretability (Bianchi et al., 2021; Mullainathan & Spiess, 2017). However, recent developments in eXplainable Artificial Intelligence (XAI) now provide methodologies to isolate the contributions of each variable to the model's output, thereby addressing key interpretability challenges.

In this study, we examine whether machine-learning approaches can outperform more traditional models by capturing the interdependencies among the coefficients of the credit-spread term structure.

Based on the Nelson-Siegel framework (Nelson & Siegel, 1987), we estimate the term structure of credit spreads for U.S. firms from July 2002 through June 2020. This long sample enables us to evaluate model performance across periods of market stress and volatility as well as periods of stability. To ensure both stability and economic interpretability in the estimated coefficients, we introduce a penalized loss function that mitigates known challenges in estimating Nelson-Siegel credit spread coefficients.

We then predict these Nelson-Siegel credit spread coefficients by extending the feature set commonly used in the literature to include macroeconomic and financial uncertainty measures highlighted by Gambetti et al., 2019, 2022. Following Bellotti et al., 2021; Gambetti et al., 2022, we compare the performance of linear regression models with a range of machine learning algorithms (bagging, random forests, and boosted trees) to assess both in-sample explanatory power and out-of-sample predictive accuracy. We also employ SHAP values (Lundberg & Lee, 2017) to interpret the machine learning models, thereby quantifying each feature's contribution to the predictions. Finally, we conduct robustness tests to validate our results.

Our study makes three main contributions to the literature. First, we propose a new penalized loss function designed to enhance the economic interpretability and temporal stability of the estimated Nelson-Siegel credit spread coefficients. This penalty loss function addresses known problems when estimating Nelson-Siegel credit spread coefficients,

thereby extending the literature. It also improves the prediction of these coefficients. Second, our empirical analysis illustrates the influence of firm-level variables (firm age, equity price, and book-to-market ratio) on credit spreads, aligning with prior findings. Moreover, we reinforce the link between economic uncertainty and credit spreads by showing that uncertainty variables explain and predict the shape of the credit-spread curve, particularly at longer forecasting horizons. The influence of macroeconomic uncertainty features varies across coefficients, resulting in heterogeneous impacts on the long-term rate, slope, and curvature of the credit-spread term structure. Third, we show that machine-learning algorithms provide a robust and parsimonious solution for capturing potential nonlinear interactions and heterogeneous risk-factor effects across firms. Notably, boosted trees outperform other models in both accuracy and stability, particularly during periods of market stress. By integrating a novel estimation strategy, a comprehensive set of predictors, and machine-learning tools, our study contributes to a more nuanced understanding of credit-spread dynamics and offers practical insights for researchers and risk managers.

The chapter is structured as follows. Section 4.2 describes the data sources. Section 4.3 presents the credit spread estimation and data analysis. Section 4.4 outlines the methodological framework for both the linear and machine-learning models. Section 4.5 presents our empirical findings, including predictive performance, variable importance, and robustness. Section 4.6 concludes.

4.2 Data Description

Our study integrates a multifaceted dataset encompassing bond transaction details, bond features, equity trade information, corporate balance sheets, U.S. default rates, indices of economic uncertainty, and various systematic factors.

We source our bond transaction data from the enhanced historical Trade Reporting and Compliance Engine (TRACE) database, maintained by the Financial Industry Regulatory Authority (FINRA). This database provides comprehensive information on Over-The-Counter (OTC) transactions involving U.S. dollar–denominated fixed-income securities,

complete with intraday timestamps.[2] Our analysis spans from July 2002, when TRACE data first became available, through June 2020. We apply the data-cleaning methodology outlined in Dick-Nielsen, 2009, 2014, which involves removing duplicates, cancellations, and corrected transactions. To prevent double-counting, we eliminate one side of agency transactions as per Dick-Nielsen, 2014. We further exclude equity-linked notes, trades executed under exceptional conditions, and commissioned trades (i.e., transactions inclusive of commission fees). Ultimately, we retain only end-of-day standard trades executed on the secondary market. For each retained trade, we record the CUSIP, execution date and time, trade volume, clean price[3], yield, trade side, and whether the trade was conducted by a customer or a dealer.

We integrate the TRACE transaction data with the Fixed Income Securities Database (FISD) to acquire bond features, using the CUSIP as the key to merge the two datasets. To ensure consistency in pricing, we focus on vanilla bonds, i.e., those without complex features (Berndt et al., 2020; Elton et al., 2001; Eom et al., 2004) such as foreign currencies, private placements, derivatives, sinking-fund provisions, convertibility, exchangeability, unit deals, Rule 144A securities, Yankee bonds, Canadian bonds, or asset-backed and mortgage-backed enhancements. We also remove all corporate floating-rate debt. From the FISD, we extract detailed bond characteristics, including maturity, bond type, seniority, coupon size, payment frequency, and credit ratings from S&P, Fitch, and Moody's, as well as issuing-firm sector.

To calculate the mid-close price, we focus on bonds with at least one bid and one ask price available per day. Out of the $19,951,603$ end-of-day bond prices in our dataset, we isolate $1,823,265$ vanilla bonds, representing 9.14% of the database. After consolidating all data sources, excluding bonds lacking a credit rating, and cleansing anomalies (e.g., negative tenors), we end up with $1,147,390$ bonds. We further narrow our analysis to bonds with maturities exceeding three months. For estimating the credit-spread curve, we employ

[2]The enhanced TRACE database includes additional transaction details beyond the standard database, such as actual volume data and indicators of the buying and selling parties. In contrast, the standard database caps volume data for trades exceeding $1million for high - yield bonds and$5 million for investment-grade bonds and only disseminates the selling side of inter-dealer trades.

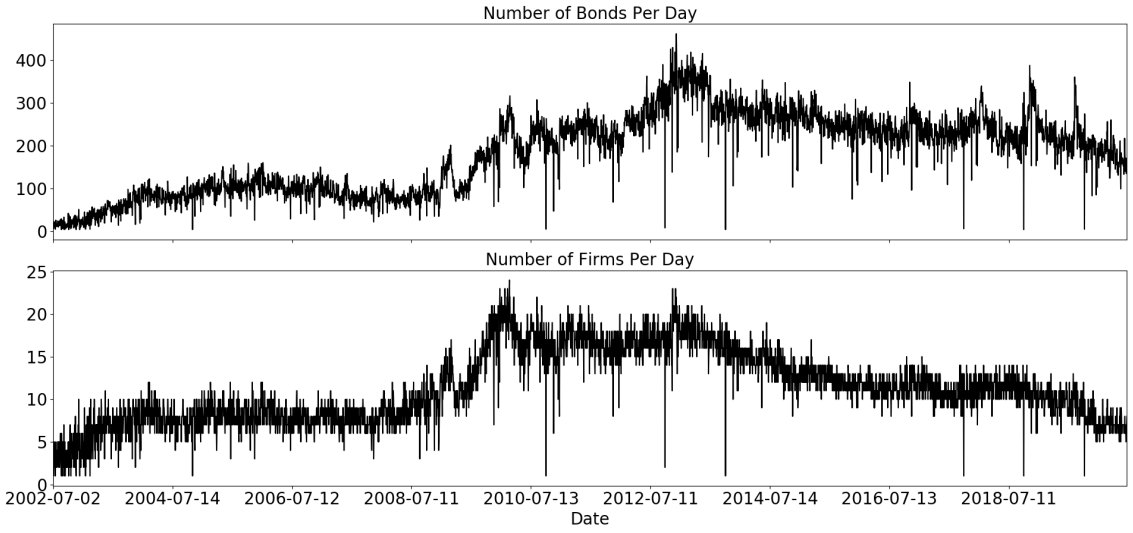
[3]The price of a coupon bond without accrued interest, as quoted on financial news sites.

the Nelson-Siegel framework, which requires optimizing four coefficients. We restrict to issuers that have a minimum of four bonds available per day for each issue and seniority level, to facilitate Nelson-Siegel optimization. Consequently, $813,154$ bonds contribute to the construction of $52,564$ zero-coupon credit-spread curves.

Figure 4.1 shows the number of bonds and issuers over time. On average, there are 181 bonds and 11 firms per day throughout the period. These numbers are lower before 2009 and increase thereafter. We categorize bonds into four seniority classes: unsecured, senior secured, senior subordinated, and subordinate; and eight industry sectors: industrials, consumer discretionary, consumer staples, health care, financials, information technology, communication services, and utilities.

We use the Center for Research in Security Prices (CRSP) database to acquire daily stock prices and calculate market capitalization[4] for each firm; Compustat to compute quarterly book-to-market values; and the U.S. zero-coupon yield term structure from Gürkaynak

Figure 4.1. Time Series of the Number of Bonds and Firms per Day



The top panel displays the daily number of bonds, while the bottom panel shows the daily count of firms. The time series spans from July 1, 2002, to June 30, 2020.

[4]Market capitalization is determined as the product of the stock price and the number of shares outstanding.

et al., 2007 to compute credit spreads.

Long-established firms are often perceived as less uncertain due to their extensive market history, while newer firms face greater uncertainty owing to limited information availability. We measure firm age from the IPO date in Compustat or, when absent, from the initial listing in the CRSP database (Pástor & Veronesi, 2003; Zhang, 2006). We also calculate firm-specific volatility as the standard deviation of daily equity returns over a three-month rolling window, providing a dynamic assessment of market-based volatility.

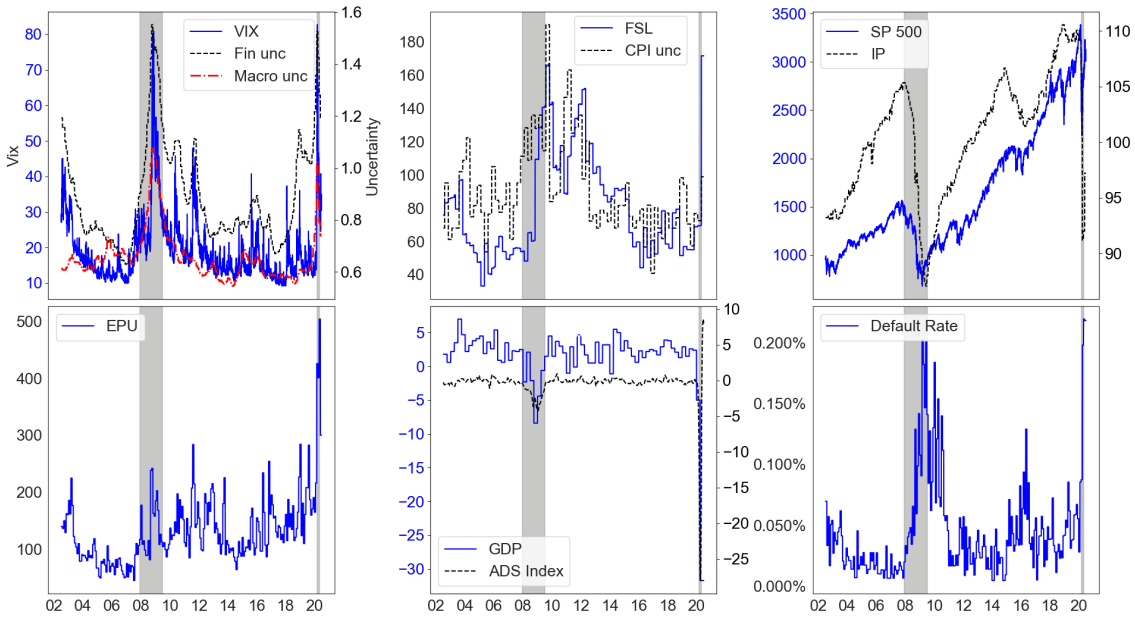
Credit spreads are influenced by a range of aggregate variables, such as changes in interest rates, business-cycle conditions, and market volatility, as noted by Collin-Dufresne et al., 2001. Therefore, we incorporate macroeconomic indicators that reflect the business cycle, including quarterly Gross Domestic Product (GDP) and monthly Industrial Production (IP), as well as the monthly recession indicators based on the National Bureau of Economic Research (NBER) from the FRED database. We also consider the Aruoba-Diebold—cotti (ADS) Business Conditions Index (Aruoba et al., 2009), which offers a daily snapshot of economic conditions. This index integrates various macroeconomic series such as the U.S. Treasury yield-curve term premium, weekly initial unemployment insurance claims, monthly non-farm payrolls, and quarterly real GDP. The ADS Index is disseminated by the Federal Reserve Bank of Philadelphia.

Credit spread is a function of default probability and recovery rate. To approximate default probability, we calculate the monthly U.S. default rate using data from Moody's Default & Rating Analytics database, defined as the number of companies in default divided by the total number of companies monitored each month. As a proxy for recovery rate, we utilize uncertainty measures linked to recovery rates in prior research (e.g., Gambetti et al., 2019). Specifically, we employ the CPI uncertainty measure and the policy-related macroeconomic uncertainty indices for federal and state/local (FSL) government purchases proposed by Baker et al., 2016. These indices are based on the dispersion of economic forecasters' expectations, sourced from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters, regarding future levels of the consumer price index, federal expenditures, and state and local government spending.

Additionally, we include the Economic Policy Uncertainty (EPU) Index from Baker et

al., 2016, which quantifies policy uncertainty based on the frequency of newspaper coverage, using a normalized index derived from articles in ten major newspapers. These three series are accessible via the website maintained by Baker et al., 2016. We also integrate the financial and macroeconomic uncertainty measures from Jurado et al., 2015 and Ludvigson et al., 2021, which aggregate the volatilities of the unpredictable components of numerous economic indicators. Lastly, we incorporate stock market volatility measures from the VIX index and the S&P 500 index, both available on the CBOE website.

Figure 4.2. Dynamic of Uncertainty Measures and Macrovariables from July 2002 to June 2020



The grey shaded areas represent NBER-defined recession periods: from December 2007 to June 2009, associated with the subprime crisis, and from March to April 2020, corresponding to the onset of the Covid-19 pandemic. The VIX index measures stock market volatility. Fin unc. refers to the financial uncertainty measure, and macro unc. denotes the macroeconomic uncertainty measure; both are from Jurado et al., 2015 and Ludvigson et al., 2021. EPU is the Economic Policy Uncertainty measure developed by Baker et al., 2016. CPI unc. indicates the inflation uncertainty measure, and FSL unc. signifies the uncertainty related to federal and state/local government purchases, both of which are also from Baker et al., 2016. IP stands for U.S. Industrial Production, GDP for U.S. Gross Domestic Product, and SP500 for the S&P 500 index. The ADS Index, which reflects economic conditions, is defined by Aruoba et al., 2009.

Figure 4.2 illustrates the trends of selected uncertainty proxies and macroeconomic variables across four phases from July 2002 to June 2020: a return-to-stability phase (post-dot-com recovery); a heightened uncertainty, volatility, and downturn during the subprime crisis; a period of relative improvement with ongoing market uncertainty; and the onset of the Covid-19 pandemic with its heightened uncertainty and economic challenges.

4.3 Zero-Coupon Credit Spread Curve Estimation

This section describes the bond-price model and the estimation procedure. The term structure of interest rates is the sum of the risk-free rate (government yield) curve and the credit spread curve, both modelled using Nelson and Siegel, 1987.

4.3.1 Estimation Method

The price of a coupon-bearing bond can be represented as a function of the price of a zero-coupon bond, as follows:

$$B_{i,t}^{model} = \sum_{j=1}^{m_i} c_{i,j} P\left(u_{i,j}\right) = \sum_{j=1}^{m_i} c_{i,j} e^{-y(u_{i,j})u_{i,j}}, \tag{4.1}$$

where $P\left(u_{i,j}\right)$ denotes the zero-coupon bond price acting as the discount factor; m_i is the number of remaining coupon payments of bond i; $u_{i,j}$ is the time-to-maturity for the j^{th} coupon payment of bond i; and $y(u_{i,j})$ is the zero-coupon yield at each point $u_{i,j}$. The j^{th} coupon payment for bond i, $c_{i,j}$, is calculated as

$$c_{i,j} = 100 \left(\frac{c_i}{f_i} \mathbb{I}_{j<m_i} + \left(1 + \frac{c_i}{f_i}\right) \mathbb{I}_{j=m_i} \right),$$

where 100 represents the face value, c_i is the coupon rate, and f_i is the number of coupon payments per year for bond i.

Our objective is to estimate the zero-coupon spread curve to better understand the risks associated with each issuer. We model the term structure of the yield curve as a combination of the risk-free zero-coupon bond yield and the credit-spread term structure, expressed

as:

$$y(\theta_{i,t}, u_{i,j}) = y(\theta_t^{gov}, u_{i,j}) + y(\theta_{i,t}^{spread}, u_{i,j}), \tag{4.2}$$

where θ_t^{gov} represents the set of coefficients for the government's zero-coupon yield curve at time t, and $\theta_{i,t}^{spread}$ denotes the set of coefficients for the zero-coupon spread curve of issuer i at time t. Applying Equation (4.2) to Equation (4.1) yields

$$B_{i,t}^{model}(\theta_t^{gov}, \theta_{i,t}^{spread}) = \sum_{j=1}^{m_i} c_{i,j} e^{-y(\theta_t^{gov}, u_{i,j})u_{i,j}} e^{-y(\theta_{i,t}^{spread}, u_{i,j})u_{i,j}}. \tag{4.3}$$

In practice, we first estimate the government yield curve and subsequently determine each issuer's credit-spread curve.

The zero-coupon yield y is modelled using the Nelson-Siegel (NS) framework of Nelson and Siegel, 1987, specified as

$$\begin{aligned} y\left(\theta_t, t\right) &= \beta_{0,t} + \beta_{1,t}\phi_1(t, \tau_t) + \beta_{2,t}\phi_2(t, \tau_t), \\ \phi_1(t, \tau_t) &= \frac{1 - e^{-t/\tau_t}}{t/\tau_t}, \\ \phi_2(t, \tau_t) &= \phi_1(t, \tau_t) - e^{-t/\tau_t}. \end{aligned} \tag{4.4}$$

$\theta_t = \{\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \tau_t\}$ represents the vector of time-specific unknown coefficients. $\beta_{0,t}$ denotes the long-term yield, $\beta_{1,t}$ represents the slope of the yield curve between the short-term and long-term yields (so that $\beta_{0,t} + \beta_{1,t}$ approximates the short-term yield). $\beta_{2,t}$ captures the curvature of the yield curve, and τ_t indicates the location of the curve's peak. This same Nelson-Siegel framework is applied to estimate the zero-coupon spread curve.

4.3.2 U.S. Treasury Yield Curve Estimation

4.3.2.1 Methodology

In this study, we reconstruct the Nelson-Siegel coefficients for U.S. Treasury zero-coupon yield curves, which are unobservable, by leveraging the daily estimates provided by Gürkaynak et al., 2007[5], derived from the Svensson model (Svensson, 1994). From a forecasting

[5]https://www.federalreserve.gov/econres/feds/the-us-treasury-yield-curve-1961-to-the-present.htm

standpoint, directly relying on the Svensson coefficients of Gürkaynak et al., 2007 can be challenging for two reasons: first, the six coefficients exhibit strong interdependencies (e.g., τ_1 and τ_2 are particularly difficult to disentangle); second, their estimates are not always stable over time. Consequently, we adopt the four-coefficient Nelson-Siegel model and ensure coefficient stability over time. To implement this, our approach involves a reverse-engineering procedure that optimizes the set of coefficients θ_t^{gov} by minimizing the root mean squared relative error between our fitted yield curves and those reported by Gürkaynak et al., 2007:

$$\hat{\theta}_t^{gov} = \underset{\theta^{gov}}{\operatorname{argmin}} \left\{ \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(\frac{y(j,\theta^{gov}) - y^{Sv,gov}(j)}{y^{Sv,gov}(j)}\right)^2} + p\left(\theta_t^{gov}\right) \right\}, \tag{4.5}$$

such that

$\beta_{0,t}^{gov} > 0,$

$\tau^{gov} \in (0, T].$

$y^{Sv,gov}$ is the U.S. Treasury zero-coupon yield curve of Gürkaynak et al., 2007, $\theta_t^{gov} = \{\beta_{0,t}^{gov}, \beta_{1,t}^{gov}, \beta_{2,t}^{gov}, \tau_t^{gov}\}$, and $T = 30$.

Given known challenges of the Nelson-Siegel model, such as its susceptibility to produce nearly identical yield curves from different coefficient sets and the existence of multiple local optima (Gauthier & Simonato, 2012; Sundaram & Das, 2011), we employ a numerical optimization strategy that incorporates a penalty function. This penalty function ensures that the coefficients are both economically interpretable and temporally consistent by mitigating excessive day-to-day fluctuations. Specifically, we define

$$p\left(\theta_t^X\right) = \sum_{l=0}^{2}\left(\beta_{l,t}^X - \beta_{l,t}^{*X}\right)^2 + \sum_{l=0}^{2}\left(\beta_{l,t}^X - \beta_{l,t-1}^X\right)^2 + \left(\frac{\tau_t^X - \tau_{t-1}^X}{\tau_{t-1}^X}\right)^2, \tag{4.6}$$

with $X \in \{gov, spread\}$. The first summation of Equation (4.6) penalizes deviations from target coefficients $\beta_{0,t}^{*gov}$, $\beta_{1,t}^{*gov}$, and $\beta_{2,t}^{*gov}$, which capture the behaviour of the U.S. Treasury zero-coupon yield curves reported by Gürkaynak et al., 2007. Assuming coefficient values should be relatively stable from day to day, the second and third summations promote stability by penalizing changes in the model's coefficients. Because the government yield

curve is derived from observable yields, the first penalty term exerts a relatively smaller influence on its estimation; for issuer credit-spread curves, however, which rely on bond prices, this component becomes more significant. Details on the construction of $\beta_{l,t}^{*gov}$ for $l = 0, 1, 2$ are provided in Appendix C.1.1. This penalized optimization framework ensures that our reconstructed yield curves align closely with observed data while maintaining a parsimonious and robust structure that is both economically meaningful and consistent over time for forecasting.

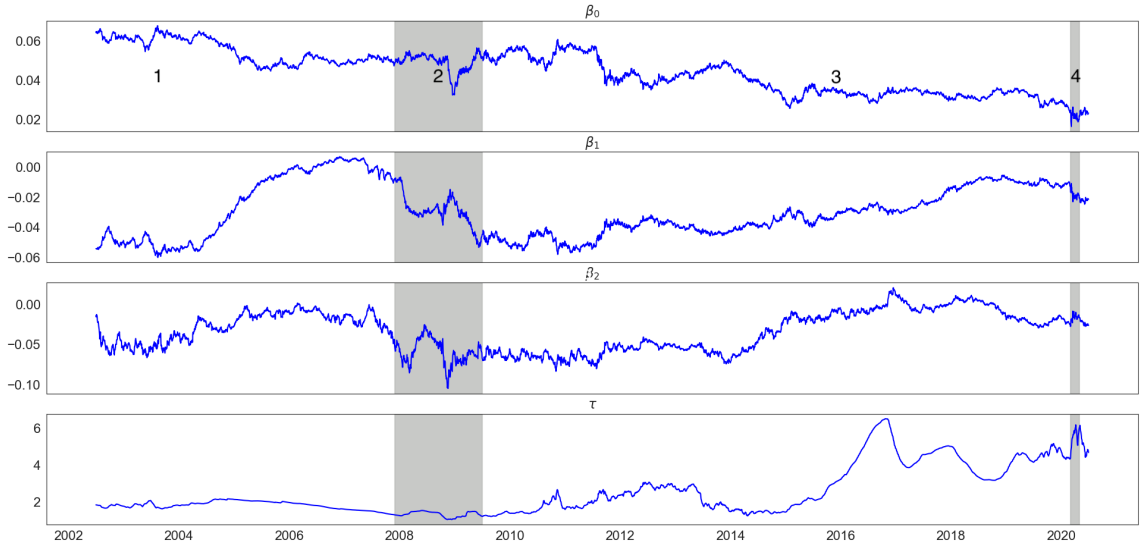
4.3.2.2 Estimation Results

Figure 4.3 depicts the time series of the estimated Nelson-Siegel coefficients $\hat{\theta}_t^{gov}$ from 2002 to 2020. The long-term yield exhibits a general downtrend over this period. In the prelude to the subprime crisis, the slope coefficient approaches zero, indicative of a predominantly flat or occasionally inverted yield curve, as corroborated by Figure 4.4.

The temporal evolution of the government yield curve's shape, shown in Figure 4.4, aligns with four distinct economic phases identified by macroeconomic dynamics. First, in the aftermath of the dot-com bubble, the yield curve is notably steep. Second, as the subprime crisis looms, the curve flattens or inverts, with elevated short-term yields reflecting heightened economic risks. The crisis itself precipitates a drop in short-term yields and a return to a steep curve, driven largely by quantitative easing measures. Third, from 2014 to 2020, a period marked by uncertainty yet favourable economic conditions, the long-term yield declines while the short-term yield rises, transitioning the curve from normal to flat. Fourth, the onset of the Covid-19 pandemic inverts the curve once more.

The correlation matrix for the Nelson-Siegel coefficients of the U.S. zero-coupon yield curve, presented in Table 4.1, reveals notable interdependencies among the coefficients. There is a strong negative correlation between the long-term yield coefficient ($\hat{\beta}_0^{\text{gov}}$) and the other coefficients, and a strong positive correlation between the slope coefficient ($\hat{\beta}_1^{\text{gov}}$) and the curvature coefficient ($\hat{\beta}_2^{\text{gov}}$). These patterns suggest that movements in the long-term yield are typically inversely related to changes in slope and curvature, while slope and curvature tend to move together, illustrating the interconnectedness of the yield curve's shape in response to evolving economic conditions.

Figure 4.3. Estimated Nelson-Siegel Coefficients of the U.S. Zero-Coupon Treasury Yield Curve



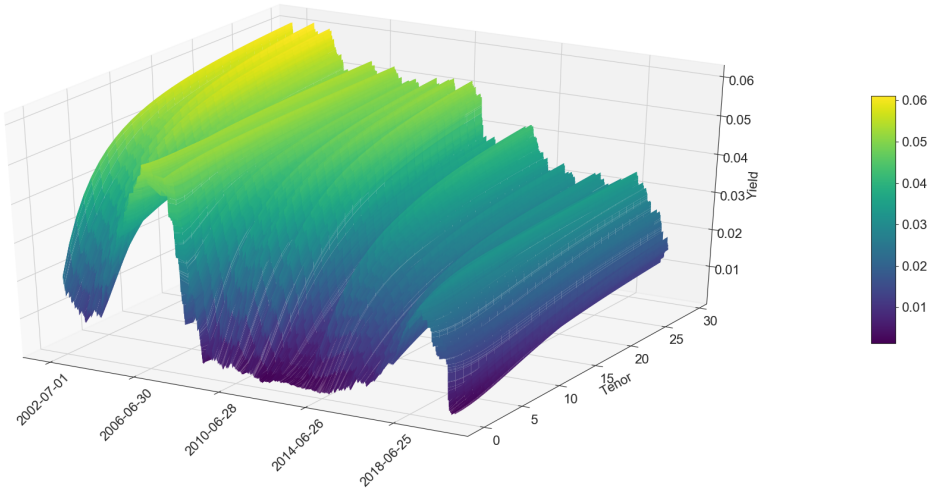
Model (4.4) is fitted on the daily Fed yield curves of Gürkaynak et al., 2007 by minimizing Equation (4.5) from July 1, 2002 to June 30, 2020. The grey shaded areas correspond to the NBER-recession from December 2007 to June 2009 and from March 2020 to April 2020. The numbers in the top panel correspond to the four distinct economic phases that composed our database.

Table 4.1. Lower Triangle Correlation Matrix Between the Nelson-Siegel Coefficients of the U.S. Zero-Coupon Yield Curves from July, 2002, to June, 2020

	β_0	β_1	β_2	τ
β_0	1			
β_1	-0.45	1		
β_2	-0.50	0.65	1	
τ	-0.70	0.28	0.58	1

Model (4.4) is fitted on the daily Fed yield curves of Gürkaynak et al., 2007 by minimizing Equation (4.5) from July 1, 2002 to June 30, 2020.

Figure 4.4. Estimated U.S. Zero-Coupon Treasury Yield Curve



Model (4.4) is fitted on the daily Fed yield curves of Gürkaynak et al., 2007 by minimizing Equation (4.5) from July 1, 2002 to June 30, 2020. The grey shaded areas correspond to the NBER-recession from December 2007 to June 2009 and from March 2020 to April 2020.

4.3.3 Firm-Specific Zero-Coupon Spread Curve Estimation

4.3.3.1 Methodology

In contrast to the estimation of government curve coefficients, where we calibrated the yield structure to Gürkaynak et al., 2007, we estimate the firm-specific zero-coupon spread curve directly from bond prices. The clean price, ($B_{i,t}^{clean}$), excludes accrued interest since the last coupon payment. The dirty price, ($B_{i,t}^{obs}$), includes accrued interest and is computed as:

$$\begin{aligned} B_{i,t}^{obs} &= B_{i,t}^{clean} + \text{accrued interest at time } t, \\ &= B_{i,t}^{clean} + 100\frac{c_i}{f_i}\frac{360(t-u_{i,j-1})}{360(u_{i,j}-u_{i,j-1})}, \end{aligned} \tag{4.7}$$

where $(t-u_{i,j-1})$ represents the fraction of year since the last coupon payment, and $(u_{i,j}-u_{i,j-1})$ is the fraction of year between consecutive coupon payments.

To estimate the firm-specific coefficients of the zero-coupon spread curve, we minimize the following loss function:

$$\hat{\theta}_{i,t}^{spread} = \underset{\theta_{i,t}^{spread}}{\operatorname{argmin}} \left\{ \sqrt{\frac{1}{N_{i,t}} \sum_{j=1}^{N_{i,t}} \left(\frac{B_{j,t}^{model}(\hat{\theta}_t^{gov}, \theta_{i,t}^{spread}) - B_{j,t}^{obs}}{B_{j,t}^{obs}} \right)^2} + p\left(\theta_{i,t}^{spread}\right) \right\}, \quad (4.8)$$

such that

$\beta_0^{spread} > 0,$

$\tau^{spread} \in (0, u_{j,last}].$

$N_{i,t}$ is the number of bonds issued by firm i at time t, $B_{j,t}^{model}(\hat{\theta}_t^{gov}, \theta_{i,t}^{spread})$ is the model-derived price of the j^{th} bond defined by Equation (4.3), $B_{j,t}^{obs}$ is the observed dirty price defined by Equation (4.7), and $p\left(\theta_{i,t}^{spread}\right)$ is the penalty function defined in Equation (4.6) ensuring the coefficients' economic interpretability and temporal stability. The penalty function is applied to the spread coefficients $\theta_{i,t}^{spread} = \{\beta_{0,i,t}^{spread}, \beta_{1,i,t}^{spread}, \beta_{2,i,t}^{spread}, \tau_{i,t}^{spread}\}$, with $u_{j,last}$ indicating the time-to-maturity of the last coupon payment for bond j. The target coefficients $\beta_{l,i,t}^{*spread}$ are determined using the long-term yield, slope, and curvature derived from the yield-to-maturity curve of issuer i at time t (c.f., Appendix C.1.2).

4.3.3.2 Estimation Results

To assess the accuracy of our model, we calculate the pricing error between the estimated price from Equation (4.3) and the observed dirty price:

$$e_{i,t} = B_{i,t}^{model}(\hat{\theta}_{i,t}^{gov}, \hat{\theta}_{i,t}^{spread}) - B_{i,t}^{obs} \quad (4.9)$$

The results in Table 4.2 indicate that model precision improves for shorter bond maturities. On average, the MAE is 96 cents, with a median of 57 cents for a face value of $\$100$, representing less than 1% of face value. This error remains below $\$1$ for bonds maturing in less than ten years, which account for 75.85% of the dataset, and exceeds $\$1$ for longer maturities. The mean is approximately double the median, reflecting the influence of outliers.[6]

[6]To ensure robustness, we re-estimated the model after excluding bonds with errors beyond the 99.99th percentile. This adjustment did not significantly alter the MAE distribution, suggesting that the estimated

Table 4.2. Absolute Error Between Observed and Estimated Prices by Maturity Bucket

	count	mean	std	min	25%	50%	75%	95%	99%	max
$T \leq 1$	$73,457$	0.41	0.71	0	0.08	0.21	0.48	1.44	3.02	21.40
$1 < T \leq 5$	$365,192$	0.63	0.79	0	0.16	0.40	0.83	1.96	3.60	25.68
$5 < T \leq 10$	$178,121$	0.99	0.94	0	0.33	0.75	1.38	2.75	4.21	17.31
$10 < T \leq 20$	$144,885$	1.72	1.59	0	0.58	1.28	2.34	4.92	7.60	19.30
$20 < T \leq 30$	$51,499$	1.82	1.72	0	0.55	1.30	2.54	5.39	7.69	16.61
Overall	$813,154$	0.96	1.17	0	0.22	0.57	1.26	3.19	5.79	25.68

Errors correspond to the difference between estimated and observed prices for a face value of $\$100$ (Equation (4.9)) from July 1, 2002, to June 30, 2020. Estimated prices use the NS coefficients of the U.S. Treasury yield curve (Equation (4.5)) and the firm-specific spread curve coefficients (Equation (4.8)).

4.3.4 Preliminary Data Analysis of the Zero-Coupon Credit Spread Curves

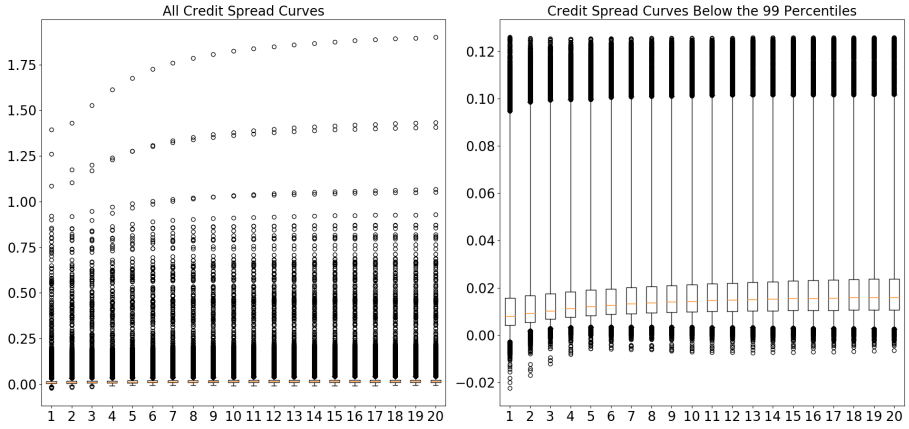
Figure 4.5 depicts the distribution of credit spread curves, highlighting outliers particularly during the subprime crisis. 50% percent of the credit spreads lie between 0.8887% and 2.173%, with an upward trend for longer tenors. The largest credit spread was recorded by Ford on November 21, 2008. Credit spreads began to spike in September 2008, peaking by the end of November 2008, coinciding with the worst days of the crisis (c.f., Figure 4.6). During this period, major automotive companies, such as General Motors, Chrysler, and Ford, were on the brink of insolvency, averted through government intervention.

Credit-spread curves, differentiated by credit rating and sector in Figures 4.7a and 4.7b, show that higher-rated firms have lower spreads due to reduced risk,[7] while firms in financial distress, especially during the subprime crisis, such as Ford, exhibit higher spreads. Less economically sensitive sectors (e.g., industrials and health care), have lower spreads, while finance and consumer discretionary sectors show higher spreads, with Ford's spreads during the 2008 crisis illustrating the latter's economic vulnerability.

coefficients are not disproportionately influenced by extreme values.

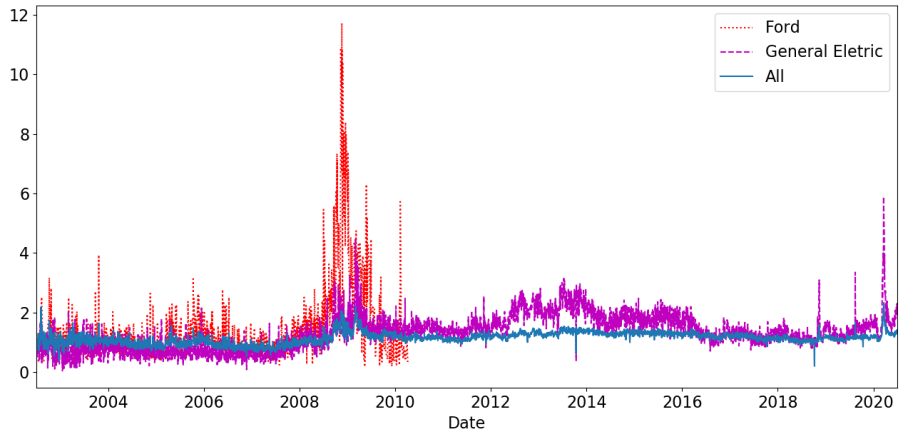
[7]The prime credit-spread curve includes General Electric up to early 2009.

Figure 4.5. Boxplot of the Zero-Coupon Credit Spreads from July 2002 to June 2020



The x-axis corresponds to tenor (in years). ZC CS are computed using end-of-day vanilla bond prices from TRACE (July 1, 2002–June 30, 2020). Curves are estimated by minimizing the RMSRE plus a penalty function. $52,564$ curves were estimated across 75 issuers.

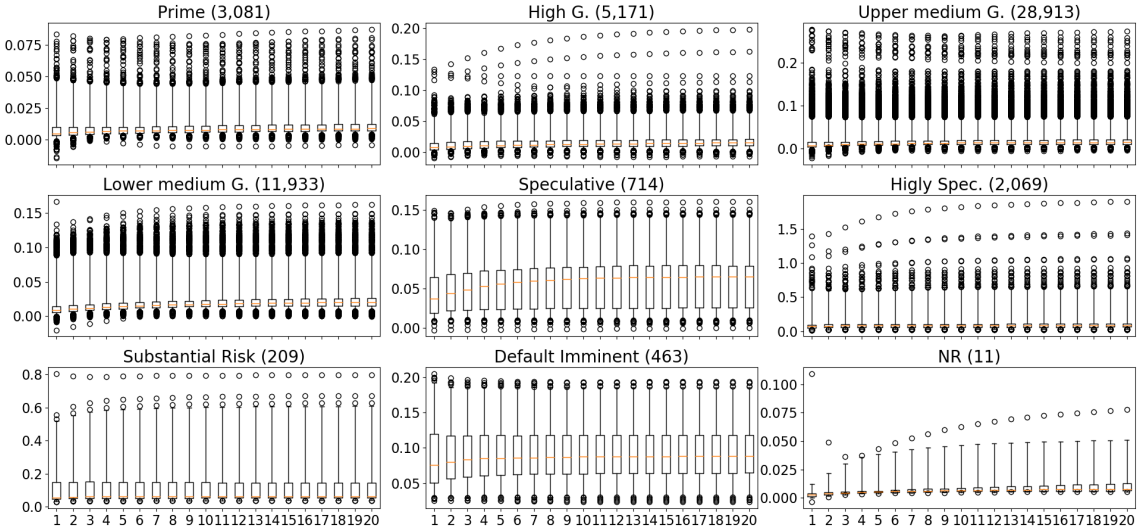
Figure 4.6. Time Series of RMSE Between Observed and Estimated Price



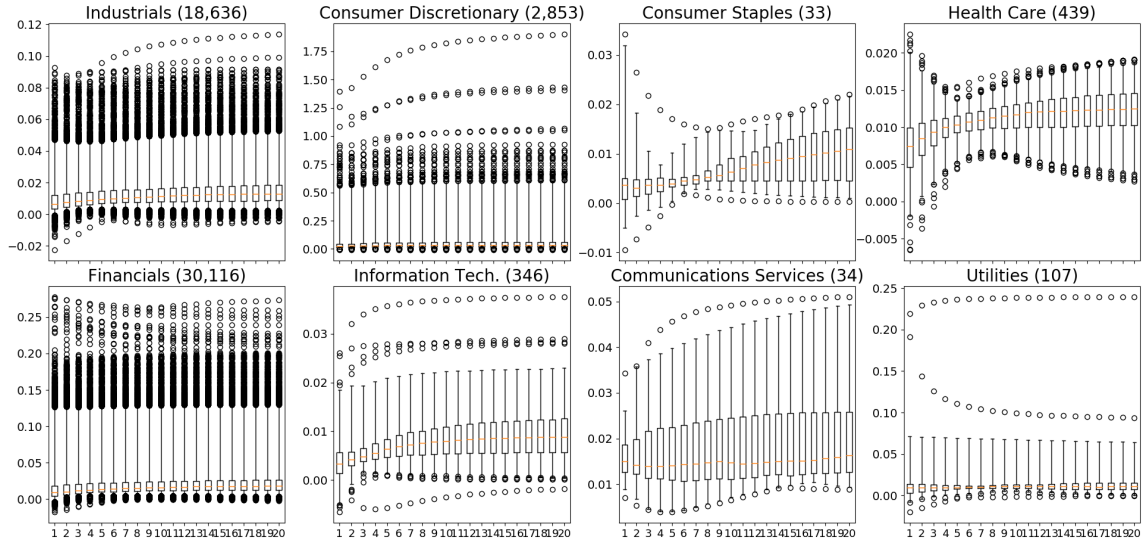
RMSE of pricing errors. Estimated bond prices are computed using Equation (4.9).The blue curve shows the RMSE across all issuers; the red dotted curve (magenta dashed curve) is for Ford (General Electric).

Figure 4.7. Boxplot of the Zero-Coupon Credit Spreads from July 2002 to June 2020

by Credit Rating



by Sectors



The x-axis corresponds to tenor (in years). The top (bottom) panel displays the distribution of zero-coupon credit spreads by rating (sector). Numbers in parentheses in each subtitle indicate the total count of credit-spread curves in that category. Curves are computed using end-of-day mid dirty prices of vanilla corporate bonds from TRACE for July 2002–June 2020. Coefficients are estimated via the Nelson-Siegel model by minimizing the RMSRE plus a penalty function for interpretable and stable estimates (Equation (4.8)).

4.4 Models

In this section, we first categorize the explanatory variables into three major groups. We then introduce the two classes of predictive models employed in our analysis: linear models and ensemble-based models. Next, we discuss our forecasting framework and the performance metrics used to evaluate model accuracy. Finally, we present the SHAP feature-importance measure, which quantifies each variable's marginal contribution to the predictive model.

4.4.1 Features

The explanatory variables presented in Section 4.2 can be classified into three groups.

The first group comprises firm-specific variables: firm age, bankruptcy indicator, seniority, sector dummies, and rating dummies. These features are denoted by the vector $\boldsymbol{X}_{i,t}$.

The second group comprises equity, macroeconomic, and uncertainty variables. Table 4.3 reveals strong correlations among several indicators, notably a 0.79 correlation between the VIX and the financial-uncertainty measure, underscoring how market sentiment aligns with broader stability perceptions. Similarly, the S&P 500 Index exhibits a 0.78 correlation with Industrial Production, suggesting a link between stock-market performance and industrial activity. Lastly, the financial-uncertainty measure has a 0.72 correlation with the macroeconomic-uncertainty measure from Ludvigson et al., 2021, reflecting the intertwined nature of financial and economic uncertainties.

To mitigate multicollinearity, we select the variables most highly correlated with the Nelson-Siegel coefficients for inclusion: namely, the VIX, Industrial Production, and the macroeconomic-uncertainty measure. A rank-plot analysis (see Appendix C.2) indicates minimal relationships among the remaining variables. Consequently, this second group includes equity prices, equity trading volume, equity returns, bid-ask spreads, high-low spreads, market capitalization, book-to-market ratio, VIX, federal-and-state/local purchases uncertainty, firm-specific uncertainty, CPI uncertainty, Economic Policy Uncertainty, macroeconomic uncertainty measure, default rates, Industrial Production, GDP, and the ADS In-

dex. These variables are denoted by the vector $\boldsymbol{Z_{i,t}}$.

The third group consists of the issuer's own lagged coefficients, grouped in $\boldsymbol{\theta_{i,t-1}^{spread}}$, defined as

$$\boldsymbol{\theta_{i,t-1}^{spread}} = \left[\beta_{0,i,t-1}, \beta_{1,i,t-1}, \beta_{2,i,t-1}, \tau_{i,t-1}\right]'.$$

Taken together, these three sets of explanatory variables serve as the inputs for our empirical analysis.

Table 4.3. Lower-Triangle Correlation Matrix Between the Uncertainty Measures and the Macrovariables from July 2002 to June 2020

	Default rate	EPU	FSL unc.	CPI unc.	VIX	Fin unc.	IP	GDP	SP500 index	Macro unc.	ADS index
Default Rate	1										
EPU	0.43	1									
FSL unc.	0.44	0.26	1								
CPI unc.	0.21	-0.01	0.46	1							
VIX	0.49	0.45	0.36	0.40	1						
Fin unc.	0.65	0.53	0.38	0.41	**0.79**	1					
IP	-0.45	-0.01	-0.53	-0.44	-0.46	-0.34	1				
GDP	-0.55	-0.60	-0.23	-0.13	-0.45	-0.49	0.17	1			
SP500 index	0.01	0.34	-0.27	-0.42	-0.33	-0.06	**0.78**	-0.18	1		
Macro unc.	0.55	0.21	0.18	0.45	0.68	**0.72**	-0.41	-0.47	-0.31	1	
ADS index	-0.29	-0.40	0.04	-0.04	-0.45	-0.43	0.12	0.40	-0.04	-0.47	1

EPU stands for the Economic Policy Uncertainty measure of Baker et al., 2016. FSL unc. stands for the uncertainty relative to both federal and state/local purchases of Baker et al., 2016. CPI unc. stands for the inflation uncertainty measure of Baker et al., 2016. The VIX index measures stock market volatility. The ADS index reflects the economic state condition and is defined by Aruoba et al., 2009. Fin unc. refers to the financial uncertainty measure, and macro unc. denotes the macroeconomic uncertainty measure; both are derived from the work of Jurado et al., 2015 and Ludvigson et al., 2021. IP stands for U.S. Industrial Production, and GDP for U.S. Gross Domestic Product.

4.4.2 Models Specification

The identification of the best class of models is undertaken through a benchmark study. We compare seven models belonging to the classes of linear and rule-based machine learning methods.

4.4.2.1 Linear Models

A linear representation of the model is encompass in

$$\boldsymbol{\theta}_{\boldsymbol{i,t+h}}^{\boldsymbol{spread}} = \alpha_0 + \boldsymbol{\alpha}\boldsymbol{X}_{\boldsymbol{i,t}} + \boldsymbol{\Gamma}\boldsymbol{Z}_{\boldsymbol{i,t}} + \boldsymbol{\Psi}\boldsymbol{\theta}_{\boldsymbol{i,t}}^{\boldsymbol{spread}} + \varepsilon_{t+h},$$

where h is the forecast horizon; α_0 is a constant; $\boldsymbol{\alpha}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$ are parameter vectors for firm-specific characteristics $\boldsymbol{X}_{\boldsymbol{i,t}}$, financial, macroeconomic, and uncertainty indicators $\boldsymbol{Z}_{\boldsymbol{i,t}}$, and the issuer's own lagged coefficients $\boldsymbol{\theta}_{\boldsymbol{i,t}}^{\boldsymbol{spread}}$, respectively.

We consider four linear models: ordinary least squares (OLS), lasso regression, ridge regression, and elastic net regression. Each can be defined as a special case of the following minimization problem:

$$\begin{aligned}\text{argmin}_{\alpha_0,\boldsymbol{\alpha},\boldsymbol{\Gamma}} \|\boldsymbol{\theta}_{\boldsymbol{i,t+h}}^{\boldsymbol{spread}} - \alpha_0 - \boldsymbol{\alpha}\boldsymbol{X}_{\boldsymbol{i,t}} - \boldsymbol{\Gamma}\boldsymbol{Z}_{\boldsymbol{i,t}} - \boldsymbol{\Psi}\boldsymbol{\theta}_{\boldsymbol{i,t}}^{\boldsymbol{spread}}\|_2^2 \\ + \lambda\left((1-\gamma)\|\boldsymbol{\alpha}+\boldsymbol{\Gamma}\|_2^2 + \gamma\|\boldsymbol{\alpha}+\boldsymbol{\Gamma}\|_{\mathbf{1}}\right). \qquad (4.10)\end{aligned}$$

$\lambda > 0$ and $\gamma \in [0,1]$ are two penalty factors. In particular, $\lambda = 0$ defines the standard linear regression (OLS). When $\lambda > 0$ the model become a penalized model and threes models can occurs:

- if $\gamma = 0$, Eq. (4.10) corresponds to the ridge model;
- if $\gamma = 1$, Eq. (4.10) corresponds the lasso model;
- if $\gamma \in]0,1[$, Eq. (4.10) corresponds the elastic net model.

These regularization techniques help reduce model complexity and prevent overfitting.

4.4.2.2 Rule-Based Models

An ensemble model is a predictive framework that combines multiple "weak learners" (in this case, decision trees) into a single aggregated predictor. A decision tree itself is a supervised learning model that uses a tree-like structure to make predictions by iteratively splitting the training dataset into subsets based on selected feature thresholds. The growth of each tree is guided by minimizing the Mean Squared Error (MSE).

We consider three nonlinear regression models derived from tree-based ensembles: bagging trees, random forests, and boosted trees. Although each method aggregates multiple decision trees into one final predictor, they differ in their sources of randomness and in whether the trees are trained in parallel or sequentially.

The bagging trees model (Breiman, 1996) is a combination of multiple decision trees trained in parallel on different observations of the dataset; thus, each tree sees a slightly different subset of the training data. The final prediction is the average of the individual trees' predictions. The random forest model (Breiman, 2001) is a combination of multiple decision trees trained in parallel on different observations of the dataset and on different subset of features (Ho, 1998). This decorrelation among trees diminishes the chance that any single feature will dominate the prediction, thereby improving predictive accuracy and reducing the overall variance. By contrast, the boosted trees model is a combination of multiple decision trees trained sequentially such that each new tree is trained to reduce the residual errors produced by the existing ensemble. This approach often yields high accuracy, but it requires careful hyper-parameter tuning. The list of hyper-parameters and their values for each model can be found in Appendix C.3.

4.4.3 Pipeline, Prediction, and Model Assessments

The explanatory variables are scaled using the min-max transformation. In line with standard practice, we allocate the first 75% of the sample (ending January 7, 2016) to the in-sample (train) set and the remaining 25% (from February 8, 2016 onward) to the out-of-sample (test) set. Following De Prado, 2018, we enforce a one-month "purging period" between these sets to prevent data leakage, that is, unintended information overlap between

the training and testing samples. Specifically, leakage may occur when either the features or labels at the boundary of the training period closely resemble those in the test period. Consequently, the in-sample set consists of all the credit spread curves from July 1, 2002, to January 7, 2016, totalling $39,386$ curves and covers the post-dot-com bubble era and the global financial crisis. The out-of-sample set consists of all the credit spread curves from February 8, 2016, to June 30, 2020, totalling $12,900$ curves and encompass a period of uncertainty that finish with the covid-19 pandemic.

Hyper-parameters are tuned using 10-fold time-series split cross-validation, where the training set is partitioned into 10 sequential sub-blocks. At each iteration, a new block serves as a validation set, while the preceding blocks is added to the in-sample set.[8] A one-month gap is included between these sub-samples to further safeguard against data leakage. The RMSE is used as cost function.

All models are implemented, trained, and validated in Python using `scikit-learn`. We adopt a two-stage procedure. In the first stage, we estimate the four Nelson-Siegel coefficients jointly using the `MultiOutputRegressor` class. Let

$$\boldsymbol{\theta}^{\boldsymbol{spread}}_{\boldsymbol{i,t+h|t}} = \mathbb{E}^{\mathbb{P}}_t \left[\boldsymbol{\theta}^{\boldsymbol{spread}}_{\boldsymbol{i,t+h}} \middle| \mathcal{F}_t \right] \tag{4.11}$$

denote the h-day-ahead predicted Nelson-Siegel coefficients conditional on the information set $\mathcal{F}_t$ for issuer i. The loss function in this multi-output setting is defined as the mean squared error across the four coefficients:

$$MSE = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left| \boldsymbol{\theta}^{\text{spread}}_{\boldsymbol{i,t+h}} - \boldsymbol{\theta}^{\text{spread}}_{\boldsymbol{i,t+h|t}} \right|^2, \tag{4.12}$$

where $N = \sum_{t=1}^{T} N_t$. Estimating the coefficients jointly via a single loss function allows us to capture their interdependencies.

In the second stage, to forecast the credit spread itself, we use the predicted Nelson-Siegel coefficients and compute the RMSE between predicted and realized spreads, such

[8] In others word, at the first iteration, the first block is used as in-sample set, and the second block is used as validation set. At the second iteration, the first two blocks are used as in-sample set, and the third block is used as validation set, and so on.

that:

$$RMSE(h, model) = \sqrt{\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{j=1}^{20}\left(y\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h}}^{\boldsymbol{spread}}, j\right) - y\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h|t}}^{\boldsymbol{spread}}, j\right)\right)^2}, \quad (4.13)$$

where the model minimizing this loss function is selected as the best performer. In addition, we employ three further precision metrics (Mean Absolute Error, R^2, and pseudo R^2) to provide a comprehensive assessment of forecasting performance.

MAE is defined as

$$MAE(h, model) = \frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{j=1}^{20}\left|y\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h}}^{\boldsymbol{spread}}, j\right) - y\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h|t}}^{\boldsymbol{spread}}, j\right)\right|. \quad (4.14)$$

The R^2 statistic can be defined in numerous way. As mentioned in Kvalseth, 1985 the different specifications of R^2 statistics generally yield different values except for linear models with an intercept term. Following Campbell and Thompson, 2008, we define the out-of-sample R^2 statistic as

$$R_{OS}^2(h, model) = 1 - \frac{\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{j=1}^{20}\left(y\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h}}^{\boldsymbol{spread}}, j\right) - y\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h|t}}^{\boldsymbol{spread}}, j\right)\right)^2}{\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{j=1}^{20}\left(y\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h}}^{\boldsymbol{spread}}, j\right) - y\left(\overline{\boldsymbol{\theta}_{\boldsymbol{i,t+h}}^{\boldsymbol{spread}}}, j\right)\right)^2}, \quad (4.15)$$

where $\overline{\theta_{i,t+h}^{spread}}$ is the historical average coefficient of the out-of-sample set. As mentioned in Campbell and Thompson, 2008, a positive R_{OS}^2 indicates that the predictive model outperforms the historical average (in terms of mean squared error), with larger R_{OS}^2 values reflecting stronger predictive performance.

Following Ferrari and Cribari-Neto, 2004 and Bellotti et al., 2021, we also compute the pseudo-R^2 (R_p^2), defined as the squared sample correlation between the predicted and the actual credit spreads:

$$R_p^2(h, model) = \rho^2_{\boldsymbol{y}\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h|t}}^{\boldsymbol{spread}}\right), \boldsymbol{y}\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h}}^{\boldsymbol{spread}}\right)}, \quad (4.16)$$

where $\boldsymbol{y}\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h|t}}^{\boldsymbol{spread}}\right)$ denotes the predicted credit spread at time $t+h$ knowing the information at time t for issuer i:

$$\boldsymbol{y}\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h|t}}^{\boldsymbol{spread}}\right) = y\left(\boldsymbol{\theta}_{\boldsymbol{i,t+h|t}}^{\boldsymbol{spread}}, \mathcal{T}\right), \quad (4.17)$$

where $\mathcal{T} \in \{1, 2, 3, ...20\}$ representing the set of maturities. This measure captures how well the predicted curve co-moves with the realized spread across all maturities, regardless of scale. Collectively, these metrics offer a robust framework for comparing and validating the predictive accuracy of the various models examined.

4.4.4 Feature Importance Assessment

Predicted credit spreads are computed by integrating information from multiple features. To evaluate each variable's individual contribution to predictive accuracy, we employ SHAP (SHapley Additive exPlanations) values (Lundberg & Lee, 2017), which measure feature importance by decomposing a model's prediction into feature-specific contributions.

SHAP values extend the concept of Shapley values from cooperative game theory, originally formulated to fairly allocate the total gains of a coalition among its members. In the context of predictive modelling, each "player" in the coalition corresponds to a feature, and the "value" to be divided is the model prediction. For each feature, the SHAP value is calculated by determining its marginal contribution to the prediction when added to every possible subset of features. This involves computing the difference between the model's prediction with and without the feature in question. By examining every possible combination of features and quantifying how the prediction changes when a specific feature is included or excluded, SHAP values provide a theoretically justified measure of each feature's importance.capturing both direct effects and interaction effects.

Let F represent the complete set of features, and consider a specific observation for firm i at time t. The SHAP value for feature $s \in F$, denoted $\phi_{s,i,t}$, is defined by

$$\phi_{s,i,t} = \sum_{S \subseteq F \setminus s} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[y\big(\theta^{\text{spread}}_{i,t+h|t} \mid S \cup s(x_{i,t,S\cup s})\big) - y\big(\theta^{\text{spread}}_{i,t+h|t} \mid S(x_{i,t,S})\big) \right], \tag{4.18}$$

where $x_{i,t,S}$ denotes the values of the feature subset S for firm i at time t, and $y(\theta^{\text{spread}}_{i,t+h|t} \mid S(x_{i,t,S}))$ is the predicted credit spread using a model trained on the subset S only. By averaging across all possible subsets S that exclude s, the SHAP value $\phi_{s,i,t}$ isolates the fair contribution of feature s for this particular observation.

To measure overall feature importance, we aggregate SHAP values by average absolute contribution:

$$\Phi_s = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N_t} |\phi_{s,i,t}|, \tag{4.19}$$

where higher value of Φ_s indicates greater predictive influence. In practice, we compute these values using the `shap` package in Python.

4.5 Results and Discussion

In this section, we analyze the precision metrics of the predicted coefficients and the feature importances of the explanatory variables over different time horizons ($h \in 1, 5$).

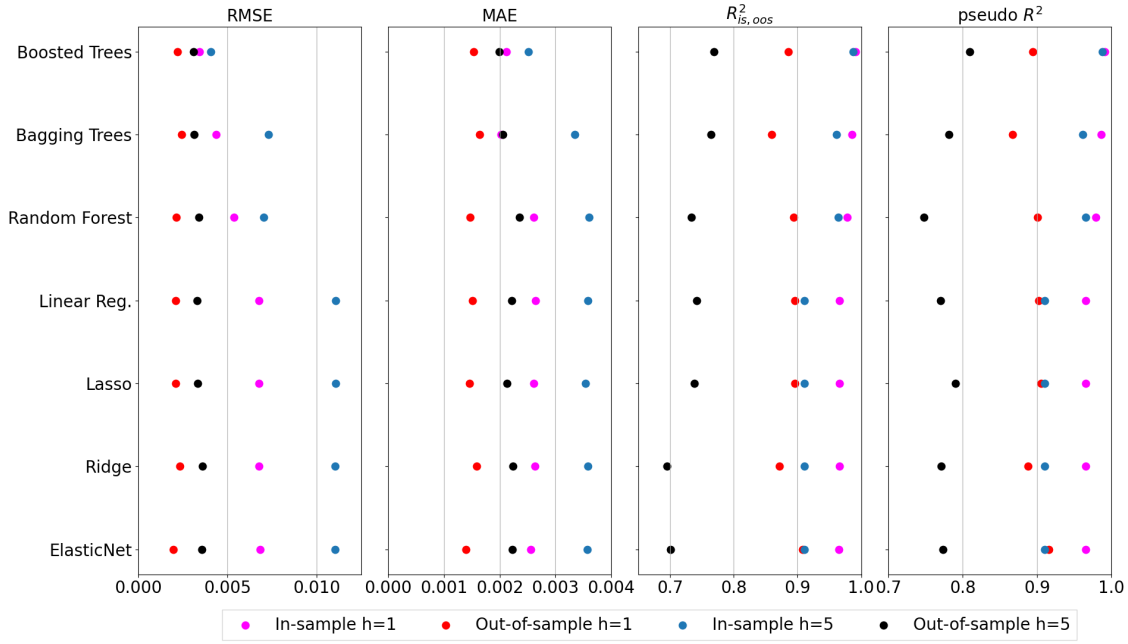
4.5.1 Models Performance

Figure 4.8 visually documents the in-sample and the out-of-sample predictive performance of the seven credit spread models evaluated using RMSE, MAE, R^2, and R^2_p. Each model is assessed both at the one-day ($h = 1$) and five-day ($h = 5$) forecast horizons. The pink (blue) dots represent the one-day ahead (five-day ahead) in-sample forecasts, whereas the red (black) dots represent the one-day ahead (five-day ahead) out-of-sample forecasts.

Contrary to expectations, out-of-sample RMSE and MAE values are lower than their in-sample counterparts. This is attributable to the large credit spreads in the in-sample dataset during the subprime crisis (Section 4.3.4). These extreme values—concentrated within a short period and reflecting high systemic risk—pose challenges for the models, which struggle to capture abrupt spikes in credit spreads, thereby inflating in-sample error metrics. Figure 4.9 illustrates this by showing the histogram of RMSE aggregated by month and firm at the one-day ahead forecast horizon. All models except boosted trees exhibit extreme in-sample errors, driven by Ford's large credit spreads between October 2008 and March 2009. Similar patterns occur at the five-day ahead horizon. Despite these spikes, the models predict unseen data with reasonable accuracy.

Figure 4.10 shows RMSE over time for each model. The boosted trees is the most stable, closely followed by random forests. Other models show larger divergences between

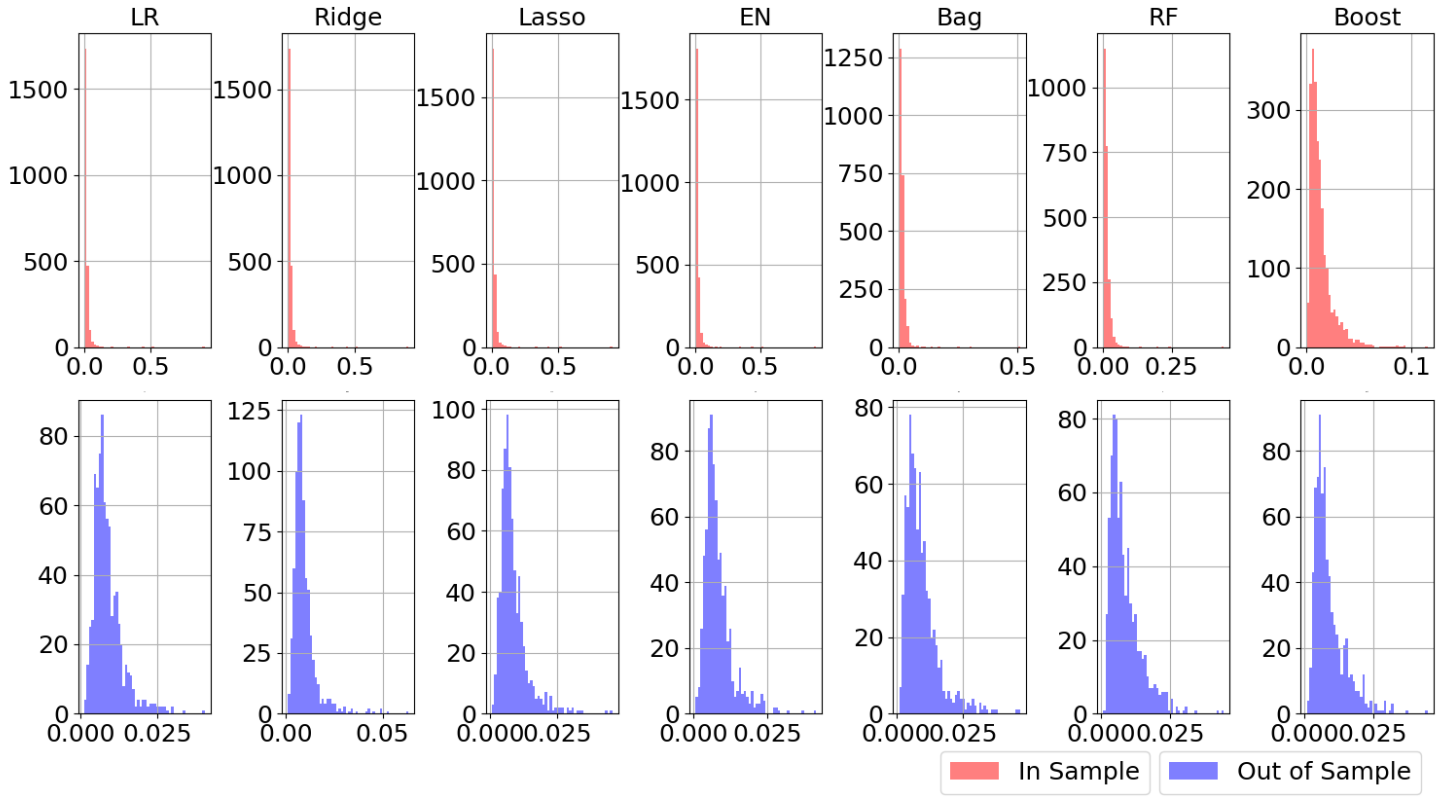
Figure 4.8. Comparison of Credit Spread Model Performances Across Metrics (In Sample and Out of Sample) - One-Day Ahead ($h = 1$) and Five-Day Ahead ($h = 5$)



The pink (blue) dots represent the one-day head (five-day-ahead) in-sample forecasts, whereas the red (black) dots represent the one-day ahead (five-day ahead) out-of-sample forecasts. The first 75% dates constitute the train set (from July 1, 2002, to January 7, 2016) and the remaining 25% of the date constitute the test set (January 8, 2016 to June 30, 2020) with a one-month purging period between the train set and the test set. The precision metrics are defined in Equation (4.13), (4.14), (4.15), and (4.16).

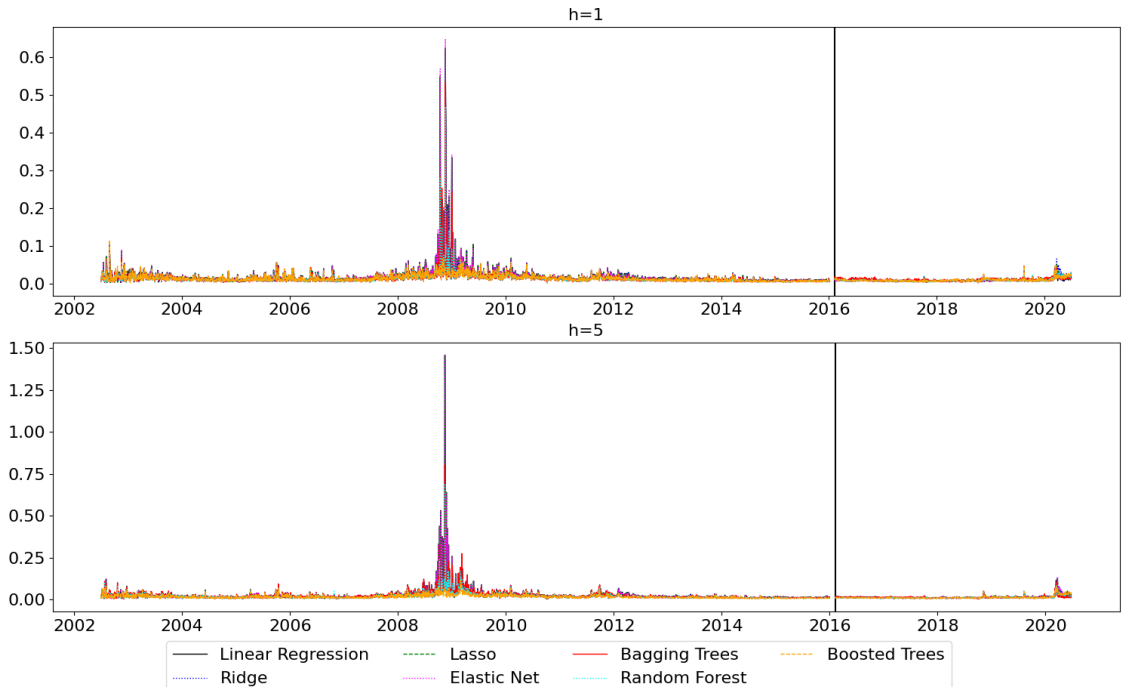
in-sample and out-of-sample performance during the 2008 crisis. Appendix C.4 reports that removing Ford from the sample narrows the gap between in-sample and out-of-sample metrics (Figure C.3).

Figure 4.9. Histogram of RMSE Aggregated by Month and Firm Between Predicted and Observed Credit Spread Curves at One-Day Ahead Forecast Horizon



The RMSE are aggregated by month and firm, resulting in 2,374 in-sample credit spreads curves, and 771 out-of-sample curves. The first 75% date constitute the train set (from July 1, 2002, to January 7, 2016) and the remaining 25% of the date constitute the test set (January 8, 2016 to June 30, 2020) with a one-month purging period between them.

Figure 4.10. Time Series of the RMSE Between Predicted and Observed Credit Spread - One-Day Ahead ($h = 1$) and Five-Day Ahead ($h = 5$)



The first 75% date constitutes the train set (from July 1, 2002, to January 7, 2016) and the remaining 25% of the date constitutes the test set (January 8, 2016 to June 30, 2020) with a one-month purging period between the train set and the test set.

Overall, ensemble tree-based methods outperform linear models by achieving lower RMSE and MAE, as well as higher R^2 and R^2_p. This suggests their ability to capture nonlinear relationships and complex features interactions. In particular, the boosted trees model minimizes the discrepancy between out-of-sample and in-sample performance metrics and achieves the lowest RMSE across the two forecast horizons. Except for the out-of-sample R^2 at five-day ahead, it also delivers the highest R^2_{OOS} and R^2_p. At five-day ahead, however, linear regression, lasso, and random forest models exhibit similar out-of-sample performance.

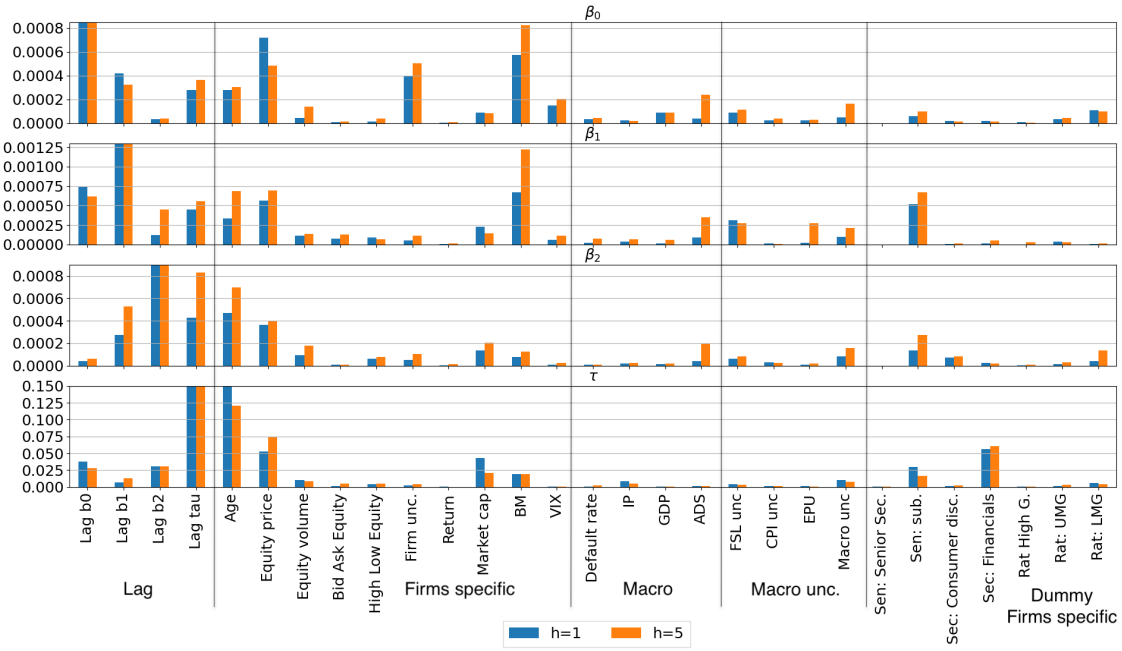
Regularized linear approaches (lasso, ridge, elastic net) do not improve results over basic linear regression, which is consistent with near-zero penalty estimates (see Appendix

C.3). All models perform better at one-day ahead forecast horizon than five-day ahead, highlighting the increasing challenge of making longer-term predictions.

4.5.2 Feature Importance Results: SHAP values

SHAP values are employed to interpret the boosted trees model, with out-of-sample results illustrated in Figure 4.11.

Figure 4.11. SHAP Values of the Predicted Nelson-Siegel Coefficients at $t + h$



SHAP-based decomposition of feature contributions to the prediction of the four Nelson-Siegel coefficients at two forecast horizons ($h = 1$ and $h = 5$) under the boosted-trees model. Each subplot corresponds to one coefficient, and the bars show how much each feature contributes to its predicted value, computed using Equation (4.19).

As expected, each Nelson-Siegel coefficient draws significant explanatory power from its own historical values and those of the other three coefficients, highlighting both the persistence and the interconnected nature of yield curve dynamics. The y-axis are capped: β_0's SHAP value for one-day ahead and five-day ahead are 0.00595 and 0.00526 respectively;

β_1's SHAP value for one-day ahead and five-day ahead are 0.00525 and 0.00437 respectively; β_2's SHAP value for one-day ahead and five-day ahead are 0.00886 and 0.00759 respectively; τ's SHAP value for one-day ahead and five-day ahead are 2.8901 and 2.8248 respectively. However, these lagged coefficients are not the sole drivers; additional features also contribute to the predictions to a lesser extent.

Firm-specific characteristics, such as firm age, equity price, and book-to-market ratio, exhibit substantial explanatory power across all four coefficients. Measures capturing liquidity (e.g., market capitalization, equity trading volume) also appear relevant. Interestingly, firm-specific uncertainty has a significant impact on the long-term yield level of the credit spread, and this impact increases as the forecast horizon lengthens.

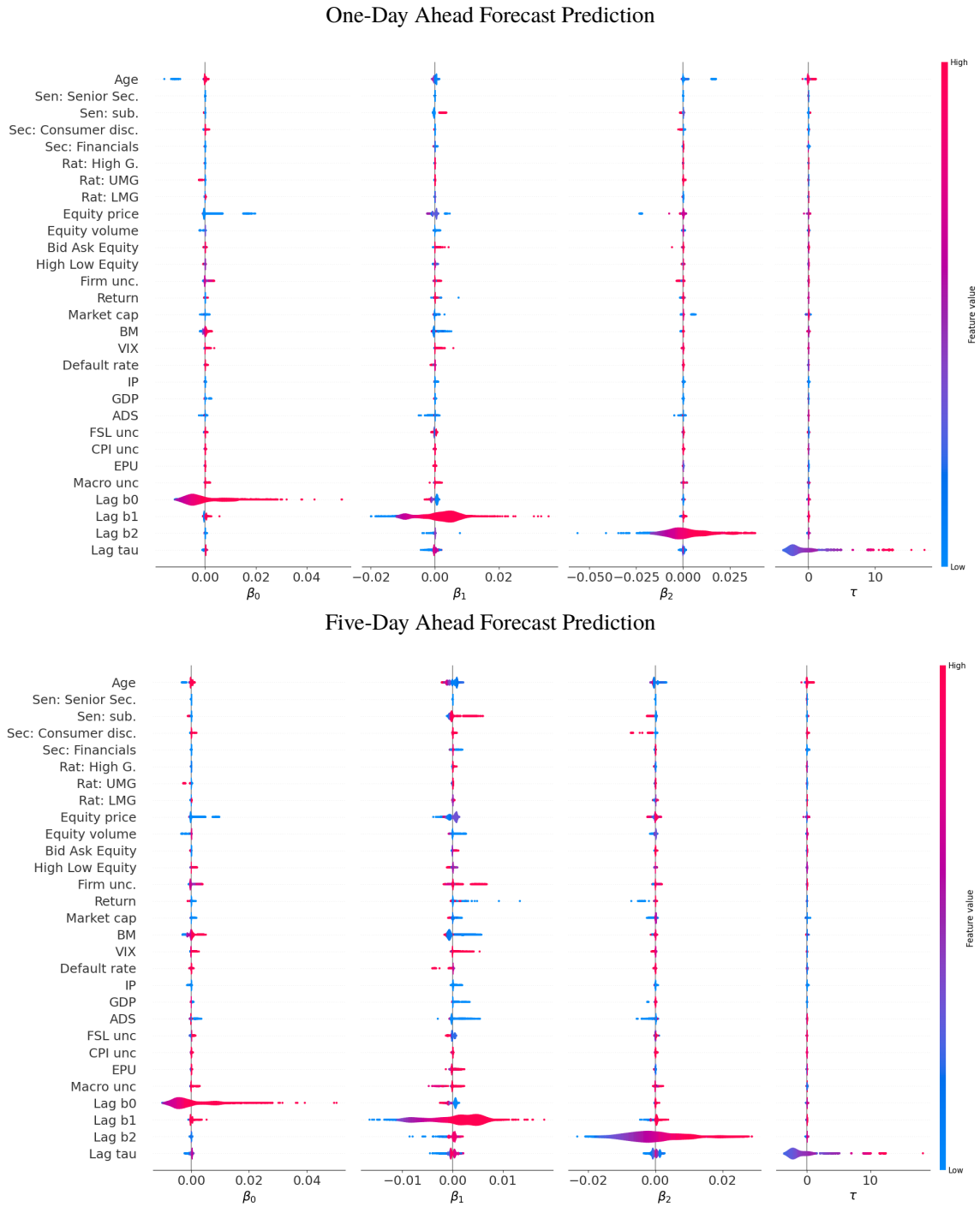
Macroeconomic and uncertainty measures (e.g., macroeconomic uncertainty, FSL, VIX, ADS, EPU) further enrich the predictive landscape. Although their individual contributions may be smaller than those of lagged coefficients and firm-level features, their importance tends to grow as the forecasting horizon extends. This aligns with the intuition that short-term uncertainty is lower, while uncertainty increases over longer horizons. Among the macroeconomic indicators, the ADS Index has greater SHAP values than the other variables. Because the ADS Index aggregates several macro variables, such as monthly Industrial Production and quarterly real GDP, this result suggests that a composite indicator may have more predictive power than its individual components. Moreover, the total contribution of the macroeconomic uncertainty variables exceeds that of the macrovariables, indicating that uncertainty measures are more predictive than macroeconomic levels.

Sector classifications (e.g., "Sec: Financials") and seniority features (e.g., "Sen: Subordinated") also emerge as contributors, demonstrating that the coefficients respond, in part, to credit-quality signals associated with specific economic and bond-market segments.

Figure 4.12 presents violin plots of SHAP values, visualizing the distribution and variability of feature contributions. The y-axis lists the features, and the x-axis shows SHAP values, indicating the magnitude and direction of each feature's influence on the Nelson-Siegel coefficients. The color bar reflects feature values. The empirical evidences suggest that high values of macroeconomic uncertainty tend to increase β_0 and β_1, steepening the credit-spread curve. Low values of firm-specific indicators tend to decrease β_0.

Subordinated bonds tend to increase β_1 and decrease β_0 and β_2. Firms in the consumer-discretionary sector exhibit higher β_0 and β_1 but lower β_2, consistent with the sector's economic sensitivity. From a macroeconomic perspective, these results confirm that macro-financial conditions shape credit spread term-structure dynamics.

Figure 4.12. Violin Plot of the SHAP Value by Nelson-Siegel Coefficients



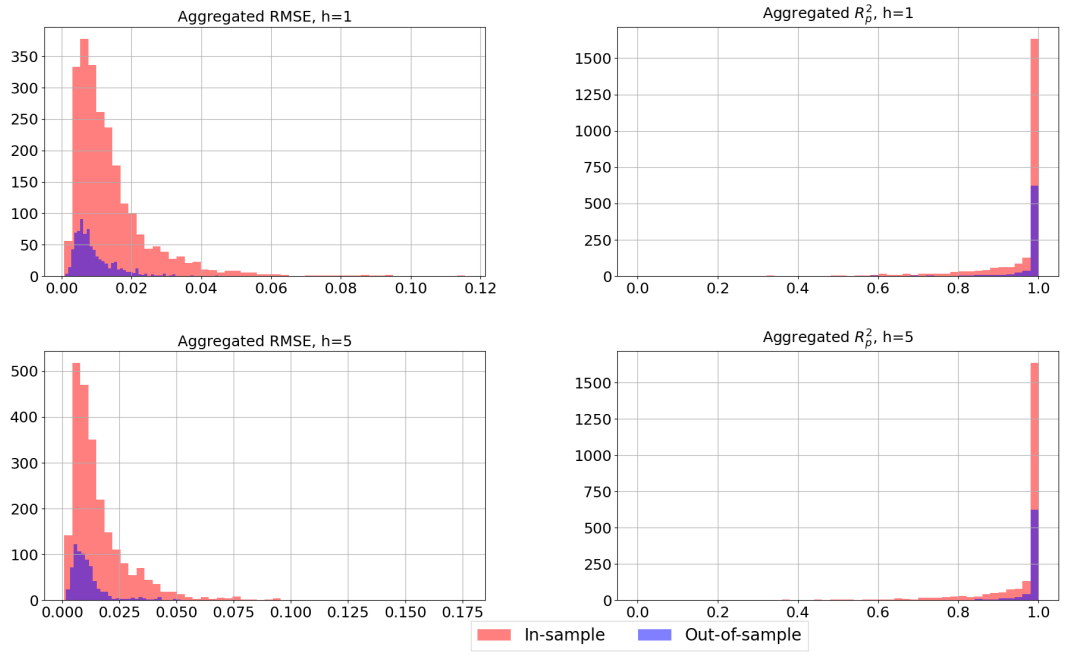
SHAP-based decomposition of feature contributions to the prediction of the four Nelson-Siegel coefficients under the boosted trees model.

Overall, different sets of variables become relevant depending on the Nelson-Siegel coefficient and forecast horizon. Longer-term forecasts incorporate more macroeconomic and uncertainty-driven features. This complexity underlines the importance of using a broad and diverse set of predictors to fully capture credit-spread curve dynamics.

4.5.3 Robustness

Figure 4.13 presents the histogram of the RMSE and pseudo R_p^2 aggregated by firm and by month for the boosted trees model. The distribution of these performance metrics does not exhibit pronounced outliers, thereby underscoring the model's robustness.

Figure 4.13. Histogram of RMSE and R_p^2 for the Boosted Trees Model



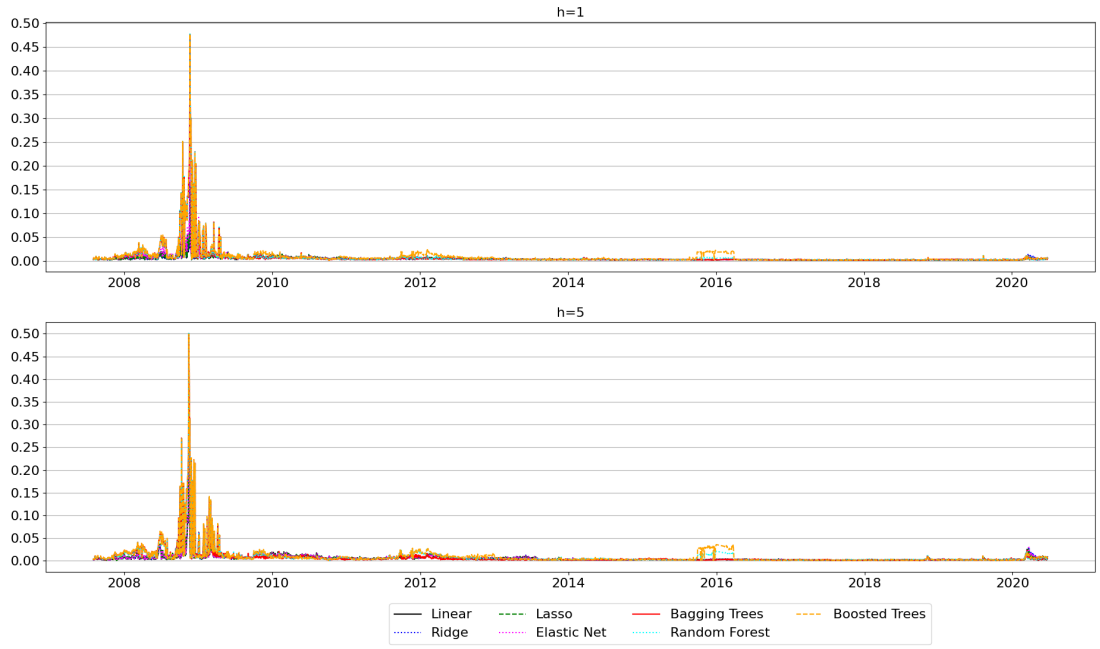
RMSE and R_p^2 are aggregated by firm and by month. For one-day-ahead forecasts, there are $2,374$ in-sample curves and 771 out-of-sample curves. For five-day-ahead forecasts, there are $2,347$ in-sample curves and 768 out-of-sample curves.

To assess the robustness of each model, we employ 13 expanding rolling windows and examine the performance metrics within each window. This approach highlights how con-

sistently each model performs over time. The first window spans five consecutive years as the in-sample period, followed by one year of out-of-sample data, with a one-month purging gap in between. The initial in-sample period runs from July 2002 to June 2008, with its out-of-sample period from August 2008 through July 2009. At each subsequent iteration, the in-sample period is extended by one additional year. Consequently, the second in-sample period covers July 2002 through June 2009, and its out-of-sample period runs from August 2009 to July 2010, and so forth.

Figure 4.14 reports the daily RMSE across the 13 out-of-sample windows. Overall performance remains relatively stable, except during the subprime crisis, when credit spreads for certain firms surged sharply. The RMSE spikes in 2008 align with earlier observations. Smaller fluctuations also occur toward the end of 2015 and the beginning of 2016 for the boosted trees model, likely reflecting challenges in selecting hyper-parameters that avoid

Figure 4.14. Daily Out-of-Sample RMSE



The daily RMSE is computed on expanding, out-of-sample rolling windows. The in-sample training set starts with five consecutive years (July 2002-June 2008) and expands by one year at each iteration. RMSE is defined in Equation (4.13).

overfitting.

To gauge the contribution of additional features, we compare the predictive performance of two boosted-trees models. Model $M1$ includes all available features, whereas Model $M2$ relies exclusively on lagged coefficients. Since this comparison focuses solely on predictive performance, we do not necessarily expect $M1$ to outperform $M2$ on every metric. As shown in Table 4.4, $M2$ achieves a lower RMSE than $M1$, but $M1$ outperforms M2 in MAE, R^2, and R^2_p. Overall, these findings imply that incorporating firm-specific characteristics, macroeconomic indicators, and uncertainty measures enhances the model's predictive power.

Table 4.4. Comparison of the Boosted Trees Performance Metrics between $M1$ and $M2$

Desc.	Model	RMSE	MAE	R^2	R^2_p
$h = 1$ - IS	$M1$	0.00344	0.00213	99.12%	99.12%
	$M2$	0.00240	0.00508	98.86%	98.19%
$h = 1$ - OS	$M1$	0.00222	0.00154	88.59%	89.39%
	$M2$	0.00135	0.00198	90.72%	91.87%
$h = 5$ - IS	$M1$	0.00409	0.00251	98.77%	98.78%
	$M2$	0.00377	0.00943	93.50%	94.11%
$h = 5$ - OS	$M1$	0.00313	0.00199	76.91%	80.92%
	$M2$	0.00220	0.00323	75.50%	79.51%

$h = 1$ and $h = 5$ denote one- and five-day-ahead forecast horizons, respectively. The in-sample (IS) period is July 1, 2002–January 7, 2016; the out-of-sample (OS) period is January 8, 2016–June 30, 2020. Model $M1$ uses all available variables; Model $M2$ uses only lagged coefficients. Performance metrics are defined in Equations (4.13), (4.14), (4.15), and (4.16).

4.6 Conclusion

This study contributes to the credit-spread modelling literature by integrating macroeconomic uncertainty indicators and statistical-learning techniques into the predictive framework for the term structure of credit spreads. Our findings highlight that the historical

values of the credi spread term-structure coefficients are not the sole drivers; firm-specific variables and macroeconomic uncertainty measures also shape credit spreads.

We propose a penalized loss function that improves the stability and economic interpretability of Nelson-Siegel credit-spread coefficients. We then compare seven models across two statistical-learning approaches: linear models and ensemble models. Our results show that the boosted-trees model outperforms others in predicting the credit-spread term structure. Using SHAP values, we demonstrate that the coefficients' historical values play a major role in prediction, but they are not the only drivers. Incorporating firm-specific variables and macroeconomic uncertainty measures enhances both explanatory power and predictive accuracy. The marginal contribution of features varies by Nelson-Siegel coefficient and forecast horizon: firm age, equity price, and book-to-market ratio are the most important firm-specific variables, while the impact of macroeconomic uncertainty measures increases with the forecast horizon. Among these, the firm-specific uncertainty measure and the macroeconomic-uncertainty measure of Ludvigson et al., 2021 are the most important for predicting the credit-spread term structure.

Future research could extend this framework by incorporating alternative uncertainty measures or by applying neural-network or deep-learning approaches to further improve predictive accuracy.

References

Altman, E. I., Brady, B., Resti, A., & Sironi, A. (2005). The link between default and recovery rates: Theory, empirical evidence, and implications. *Journal of Business*, *78*(6), 2203–2228.

Altman, E. I., & Kalotay, E. A. (2014). Ultimate recovery mixtures. *Journal of Banking & Finance*, *40*, 116–129.

Amato, J. D., & Luisi, M. (2006). *Macro factors in the term structure of credit spreads*. Citeseer.

Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, *50*(4), 745–787.

Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, *131*(4), 1593–1636.

Bellotti, A., Brigo, D., Gambetti, P., & Vrins, F. (2021). Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting*, *37*(1), 428–444.

Berndt, A., Duffie, D., & Zhu, Y. (2020). Across-the-curve credit spread indices.

Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, *34*(2), 1046–1089.

Bloom, N. (2014). Fluctuations in uncertainty. *Journal of Economic Perspectives*, *28*(2), 153–76.

Boudreault, M., Gauthier, G., & Thomassin, T. (2013). Recovery rate risk and credit spreads in a hybrid credit risk model. *The Journal of Credit Risk*, *9*(3), 3.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531.

Carbonneau, A. (2021). Deep hedging of long-term financial derivatives. *Insurance: Mathematics and Economics*, *99*, 327–340.

Collin-Dufresne, P., Goldstein, R. S., & Martin, J. S. (2001). The determinants of credit spread changes. *The Journal of Finance*, *56*(6), 2177–2207.

Cooper, I., & Priestley, R. (2009). Time-varying risk premiums and the output gap. *The Review of Financial Studies*, *22*(7), 2801–2833.

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

Dick-Nielsen, J. (2009). Liquidity biases in trace. *The Journal of Fixed Income*, *19*(2), 43–55.

Dick-Nielsen, J. (2014). How to clean enhanced trace data. *Available at SSRN 2337908*.

Duffee, G. R. (1998). The relation between treasury yields and corporate bond yield spreads. *The Journal of Finance*, *53*(6), 2225–2241.

Duffie, D., & Singleton, K. J. (1999). Modeling term structures of defaultable bonds. *The review of financial studies*, *12*(4), 687–720.

Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. (2001). Explaining the rate spread on corporate bonds. *Journal of Finance*, *56*(1), 247–277.

Eom, Y. H., Helwege, J., & Huang, J.-z. (2004). Structural models of corporate bond pricing: An empirical analysis. *The Review of Financial Studies*, *17*(2), 499–544.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Finance*.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, *51*(1), 55–84.

Feng, G., He, J., & Polson, N. G. (2018). Deep learning for predicting asset returns. *arXiv preprint arXiv:1804.09314*.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, *31*(7), 799–815.

Gambetti, P., Gauthier, G., & Vrins, F. (2019). Recovery rates: Uncertainty certainly matters. *Journal of Banking & Finance*, *106*, 371–383.

Gambetti, P., Roccazzella, F., & Vrins, F. (2022). Meta-learning approaches for recovery rate prediction. *Risks*, *10*(6), 124.

Gauthier, G., & Simonato, J.-G. (2012). Linearized nelson–siegel and svensson models for the estimation of spot interest rates. *European Journal of Operational Research*, *219*(2), 442–451.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, *54*(8), 2291–2304.

Hartmann-Wendels, T., Miller, P., & Töws, E. (2014). Loss given default for leasing: Parametric and nonparametric estimations. *Journal of Banking & Finance*, *40*, 364–375.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, *20*(8), 832–844.

Jarrow, R. (2001). Default parameter estimation using market prices. *Financial Analysts Journal*, *57*(5), 75–92.

Jarrow, R. A., Lando, D., & Turnbull, S. M. (1997). A markov model for the term structure of credit risk spreads. *The review of financial studies*, *10*(2), 481–523.

Joslin, S., Priebsch, M., & Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, *69*(3), 1197–1233.

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216.

Knight, F. H. (1921). *Risk, uncertainty and profit* (Vol. 31). Houghton Mifflin.

Kvalseth, T. O. (1985). Cautionary note about r2. *The American Statistician*, *39*(4), 279–285.

Kwan, S. H. (1996). Firm-specific information and the correlation between individual stocks and bonds. *Journal of financial economics*, *40*(1), 63–80.

Litterman, R. (1991). Common factors affecting bond returns. *Journal of fixed income*, 54–61.

Litterman, R., & Iben, T. (1991). Corporate bond valuation and the term structure of credit spreads. *Journal of portfolio management*, *17*(3), 52.

Liu, W., Fan, H., & Xia, M. (2022). Multi-grained and multi-layered gradient boosting decision tree for credit scoring. *Applied Intelligence*, 1–17.

Liu, W., Fan, H., Xia, M., & Xia, M. (2022). A focal-aware cost-sensitive boosted tree for imbalanced credit scoring. *Expert Systems with Applications*, *208*, 118158.

Longstaff, F. A., & Schwartz, E. S. (1995). A simple approach to valuing risky fixed and floating rate debt. *The Journal of Finance*, *50*(3), 789–819.

Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, *28*(1), 161–170.

Ludvigson, S. C., Ma, S., & Ng, S. (2021). Uncertainty and business cycles: Exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, *13*(4), 369–410.

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, *22*(12), 5027–5067.

Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. *CoRR*, *abs/1705.07874*.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, *29*(2), 449–470.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Nelson, C., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, *60*(4), 473–89.

Pástor, L., & Veronesi, P. (2003). Stock valuation and learning about profitability. *Journal of Finance*, *58*(5), 1749–1789.

Qi, M., & Zhao, X. (2011). Comparison of modeling methods for loss given default. *Journal of Banking & Finance*, *35*(11), 2842–2855.

Sundaram, R. K., & Das, S. R. (2011). *Derivatives: Principles and practice*. McGraw-Hill Irwin New York, NY.

Svensson, L. (1994). *Estimating and interpreting forward interest rates: Sweden 1992 - 1994* (NBER Working Papers No. 4871). National Bureau of Economic Research, Inc.

Zhang, X. F. (2006). Information uncertainty and stock returns. *Journal of Finance*, *61*(1), 105–137.

Chapter 5

Concluding Remarks

In this thesis, we explore credit risk in the bond market through the lens of Contingent Convertible debt and the prediction of the term structure of credit spreads.

The first essay provides insight about the integration of Contingent Convertible debt into banks' capital structures and the impact from an equity holder's perspective, employing a reduced-form model and dynamic optimization. Although CoCos are expensive debt instruments due to higher coupon payments rewarding the underlying risk of default and the possible zero chance of recovery on default, they offer significant benefits. CoCos reduce the cost associated with standard debt and the overall cost of debt by acting as precautionary buffers that delay the occurrence of default. From the viewpoint of equity holders, integrating CoCos into banks' capital structures increases aggregate shareholder value, particularly through enhanced dividend payouts during financial distress or periods characterized by low returns.

The second essay proposes a methodology to predict the term structure of credit spreads and bond prices by addressing challenges related to varying bond-trading frequencies across issuers. While conventional models adequately estimate the term structure of credit spreads for frequently traded bonds, predicting the credit spreads remains challenging due to the instability of term-structure coefficients. We tackle this by introducing an augmented loss function incorporating penalty terms. Given limited data for issuers with infrequently traded bonds, we develop a two-step clustering methodology: we identify clusters of high-

frequency issuers using a Gaussian Mixture Model, and we associate low-frequency issuers with these clusters based on issuer-specific characteristics. Our results reveal heterogeneity in predictive performance across clusters, where clusters with more observations yield more accurate price predictions, and inversely.

In the third essay, the prediction of the term structure of credit spreads is investigated through the integration of macroeconomic uncertainty indicators and statistical learning methodologies. Our findings demonstrate the importance of firm-specific variables to explain credit spread dynamics. As the prediction horizon increases, the explanatory power of macroeconomic uncertainty indicators intensifies, reflecting greater uncertainty over extended periods. Additionally, ensemble statistical learning models show superior predictive performance over traditional linear models, underscoring their effectiveness in credit-risk modelling.

As demonstrated throughout this thesis, the analysis of credit risk remains critically important. Recent events involving CoCos have provoked debate regarding their inherent risks and regulatory implications, forcing regulators to reassess governance frameworks aimed at mitigating associated systemic risks. Ultimately, incorporating statistical learning techniques into credit-risk research offers valuable opportunities to revisit established questions and develop innovative, more robust solutions.

Bibliography

Albul, B., Jaffee, D. M., & Tchistyi, A. (2010). Contingent convertible bonds and capital structure decisions.

Altman, E. I., Brady, B., Resti, A., & Sironi, A. (2005). The link between default and recovery rates: Theory, empirical evidence, and implications. *Journal of Business*, *78*(6), 2203–2228.

Altman, E. I., & Kalotay, E. A. (2014). Ultimate recovery mixtures. *Journal of Banking & Finance*, *40*, 116–129.

Altman, E. I., & Kishore, V. M. (1996). Almost everything you wanted to know about recoveries on defaulted bonds. *Financial Analysts Journal*, *52*(6), 57–64.

Amato, J. D., & Luisi, M. (2006). *Macro factors in the term structure of credit spreads*. Citeseer.

Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, *50*(4), 745–787.

Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, *131*(4), 1593–1636.

Bellotti, A., Brigo, D., Gambetti, P., & Vrins, F. (2021). Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting*, *37*(1), 428–444.

Berndt, A., Duffie, D., & Zhu, Y. (2020). Across-the-curve credit spread indices.

Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, *34*(2), 1046–1089.

Bishop, C. M. (2006a). Mixture model and em. In *Pattern recognition and machine learning* (pp. 423–459). springer.

Bishop, C. M. (2006b). *Pattern recognition and machine learning*. springer.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685.

Bloom, N. (2014). Fluctuations in uncertainty. *Journal of Economic Perspectives*, *28*(2), 153–76.

Bolton, P., & Samama, F. (2012). Capital access bonds: Contingent capital with an option to convert. *Economic Policy*, *27*(70), 275–317.

Boudreault, M., Gauthier, G., & Thomassin, T. (2013). Recovery rate risk and credit spreads in a hybrid credit risk model. *The Journal of Credit Risk*, *9*(3), 3.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Brigo, D., Garcia, J., & Pede, N. (2015). Coco bonds pricing with credit and equity calibrated first-passage firm value models. *International Journal of Theoretical and Applied Finance*, *18*(03), 1550015.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531.

Carbonneau, A. (2021). Deep hedging of long-term financial derivatives. *Insurance: Mathematics and Economics*, *99*, 327–340.

Chen, N., Glasserman, P., Nouri, B., & Pelger, M. (2013). Cocos, bail-in, and tail risk. *Working Paper, Office of Financial Research, U.S. Treasury Department,Washington, DC*.

Chen, N., Glasserman, P., Nouri, B., & Pelger, M. (2017). Contingent capital, tail risk, and debt-induced collapse. *The Review of Financial Studies*, *30*(11), 3921–3969.

Cheridito, P., & Xu, Z. (2015). A reduced-form contingent convertible bond model with deterministic conversion intensity. *Journal of Risk*, *17*(3).

Chung, T.-K., & Kwok, Y.-K. (2016). Enhanced equity-credit modelling for contingent convertibles. *Quantitative Finance*, *16*(10), 1511–1527.

Collin-Dufresne, P., Goldstein, R. S., & Martin, J. S. (2001). The determinants of credit spread changes. *The Journal of Finance*, *56*(6), 2177–2207.

Cooper, I., & Priestley, R. (2009). Time-varying risk premiums and the output gap. *The Review of Financial Studies*, *22*(7), 2801–2833.

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

De Spiegeleer, J., Höcht, S., Marquet, I., & Schoutens, W. (2017). Coco bonds and implied cet1 volatility. *Quantitative Finance*, *17*(6), 813–824.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Dick-Nielsen, J. (2009). Liquidity biases in trace. *The Journal of Fixed Income*, *19*(2), 43–55.

Dick-Nielsen, J. (2014). How to clean enhanced trace data. *Available at SSRN 2337908*.

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

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. *Advances in neural information processing systems*, 155–161.

Duffee, G. R. (1998). The relation between treasury yields and corporate bond yield spreads. *The Journal of Finance*, *53*(6), 2225–2241.

Duffie, D. (2010). *A contractual approach to restructuring financial institutions* (K. E. Scott, G. P. Shultz, & J. B. Taylor, Eds.). Hoover Institution, Stanford University.

Duffie, D., & Lando, D. (2001). Term structures of credit spreads with incomplete accounting information. *Econometrica*, *69*(3), 633–664.

Duffie, D., & Singleton, K. J. (1999). Modeling term structures of defaultable bonds. *The review of financial studies*, *12*(4), 687–720.

Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. (2001). Explaining the rate spread on corporate bonds. *Journal of Finance*, *56*(1), 247–277.

Eom, Y. H., Helwege, J., & Huang, J.-z. (2004a). Structural models of corporate bond pricing: An empirical analysis. *The Review of Financial Studies*, *17*(2), 499–544.

Eom, Y. H., Helwege, J., & Huang, J.-z. (2004b). Structural models of corporate bond pricing: An empirical analysis. *The Review of Financial Studies*, *17*(2), 499–544.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Finance*.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, *51*(1), 55–84.

Feng, G., He, J., & Polson, N. G. (2018). Deep learning for predicting asset returns. *arXiv preprint arXiv:1804.09314*.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, *31*(7), 799–815.

Flannery, M. J. (2005). *No pain, no gain? effecting market discipline via reverse convertible debentures* (H. S. Scoot, Ed.). Oxford University Press Oxford, UK.

Flannery, M. J. (2014). Maintaining adequate bank capital. *Journal of Money, Credit and Banking*, *46*(s1), 157–180.

Flannery, M. J. (2016). Stabilizing large financial institutions with contingent capital certificates. *Quarterly Journal of Finance*, *6*(02), 1650006.

Gambetti, P., Gauthier, G., & Vrins, F. (2019). Recovery rates: Uncertainty certainly matters. *Journal of Banking & Finance*, *106*, 371–383.

Gambetti, P., Roccazzella, F., & Vrins, F. (2022). Meta-learning approaches for recovery rate prediction. *Risks*, *10*(6), 124.

Gauthier, G., & Simonato, J.-G. (2012). Linearized nelson–siegel and svensson models for the estimation of spot interest rates. *European Journal of Operational Research*, *219*(2), 442–451.

Gieseck, A., & Largent, Y. (2016). The impact of macroeconomic uncertainty on activity in the euro area. *Jahrbuch für Wirtschaftswissenschaften*, *67*(1), 25.

Glasserman, P., & Nouri, B. (2012). Contingent capital with a capital-ratio trigger. *Management Science*, *58*(10), 1816–1833.

Goldberger, A. S. (1964). *Econometric theory.* New York: John Wiley & Sons.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, *54*(8), 2291–2304.

Hartmann-Wendels, T., Miller, P., & Töws, E. (2014). Loss given default for leasing: Parametric and nonparametric estimations. *Journal of Banking & Finance*, *40*, 364–375.

Hilscher, J., & Raviv, A. (2014). Bank stability and market discipline: The effect of contingent capital on risk taking and default probability. *Journal of Corporate Finance*, *29*, 542–560.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, *20*(8), 832–844.

Hoang, D., & Wiegratz, K. (2023). Machine learning methods in finance: Recent applications and prospects. *European Financial Management*, *29*(5), 1657–1701.

Jarrow, R. (2001). Default parameter estimation using market prices. *Financial Analysts Journal*, *57*(5), 75–92.

Jarrow, R. A., Lando, D., & Turnbull, S. M. (1997). A markov model for the term structure of credit risk spreads. *The review of financial studies*, *10*(2), 481–523.

Joslin, S., Priebsch, M., & Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, *69*(3), 1197–1233.

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216.

Knight, F. H. (1921). *Risk, uncertainty and profit* (Vol. 31). Houghton Mifflin.

Kvalseth, T. O. (1985). Cautionary note about r2. *The American Statistician*, *39*(4), 279–285.

Kwan, S. H. (1996). Firm-specific information and the correlation between individual stocks and bonds. *Journal of financial economics*, *40*(1), 63–80.

Litterman, R. (1991). Common factors affecting bond returns. *Journal of fixed income*, 54–61.

Litterman, R., & Iben, T. (1991). Corporate bond valuation and the term structure of credit spreads. *Journal of portfolio management*, *17*(3), 52.

Liu, W., Fan, H., & Xia, M. (2022). Multi-grained and multi-layered gradient boosting decision tree for credit scoring. *Applied Intelligence*, 1–17.

Liu, W., Fan, H., Xia, M., & Xia, M. (2022). A focal-aware cost-sensitive boosted tree for imbalanced credit scoring. *Expert Systems with Applications*, *208*, 118158.

Longstaff, F. A., & Schwartz, E. S. (1995). A simple approach to valuing risky fixed and floating rate debt. *The Journal of Finance*, *50*(3), 789–819.

Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, *28*(1), 161–170.

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Advances in neural information processing systems*, 431–439.

Ludvigson, S. C., Ma, S., & Ng, S. (2021). Uncertainty and business cycles: Exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, *13*(4), 369–410.

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, *22*(12), 5027–5067.

Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. *CoRR*, *abs/1705.07874*.

McDonald, R. L. (2013). Contingent capital with a dual price trigger. *Journal of Financial Stability*, *9*(2), 230–241.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, *29*(2), 449–470.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Nazemi, A., & Fabozzi, F. J. (2024). Interpretable machine learning for creditor recovery rates. *Journal of Banking & Finance*, *164*, 107187.

Nelson, C., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, *60*(4), 473–89.

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777–787.

Nunes, M., Gerding, E., McGroarty, F., & Niranjan, M. (2019). A comparison of multitask and single task learning with artificial neural networks for yield curve forecasting. *Expert Systems with Applications*, *119*, 362–375.

Pástor, L., & Veronesi, P. (2003). Stock valuation and learning about profitability. *Journal of Finance*, *58*(5), 1749–1789.

Pennacchi, G. (2010). A structural model of contingent bank capital. *FRB of Cleveland Working Paper No. 10-04*.

Qi, M., & Zhao, X. (2011). Comparison of modeling methods for loss given default. *Journal of Banking & Finance*, *35*(11), 2842–2855.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199–222.

Sundaram, R. K., & Das, S. R. (2011). *Derivatives: Principles and practice*. McGraw-Hill Irwin New York, NY.

Sundaresan, S., & Wang, Z. (2015). On the design of contingent capital with a market trigger. *The Journal of Finance*, *70*(2), 881–920.

Svensson, L. (1994a). *Estimating and interpreting forward interest rates: Sweden 1992 - 1994* (NBER Working Papers No. 4871). National Bureau of Economic Research, Inc.

Svensson, L. (1994b). *Estimating and interpreting forward interest rates: Sweden 1992 - 1994* (NBER Working Papers No. 4871). National Bureau of Economic Research, Inc.

Tobback, E., Martens, D., Van Gestel, T., & Baesens, B. (2014). Forecasting loss given default models: Impact of account characteristics and the macroeconomic state. *Journal of the Operational Research Society*, *65*(3), 376–392.

Vapnik, V. (2015). *The nature of statistical learning theory*. Springer science & business media.

Zhang, X. F. (2006). Information uncertainty and stock returns. *Journal of Finance*, *61*(1), 105–137.

Chapter A

Appendices of *Contingent Convertible Debt: The Impact on Equity Holders*

A.1 The Floating Rates

Since pricing is performed under the risk-neutral measure, the first step is to determine the risk factor $\mathbb{Q}-$dynamics.

A.1.1 The $\mathbb{Q}-$Dynamics of the Debt Ratio

The model uncertainty is captured by the noise series $\left\{\varepsilon_t^{\mathbb{P}}\right\}_{t\in\mathbb{N}}$. The link between the $\mathbb{P}$ probability measure and some risk neutral probability measure $\mathbb{Q}$ is achieved through a Radon–Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$. The latter is based on the Girsanov theorem,

$$\frac{\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathcal{F}_{t+1}}}{\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathcal{F}_t}} = \exp\left(\gamma\varepsilon_{t+1}^{\mathbb{P}} - \frac{\gamma^2}{2}\right),$$

where the γ parameter is interpreted as the price of risk. The conditional risk-neutral moment generating function (MGF) of $\varepsilon_{t+1}^{\mathbb{P}}$ is

$$M_t^{\mathbb{Q}}(x) = \mathrm{E}^{\mathbb{Q}}\left[\mathrm{e}^{-x\varepsilon_{t+1}^{\mathbb{P}}}\middle|\,\mathcal{F}_t\right] = \mathrm{E}^{\mathbb{P}}\left[\exp\left(\gamma\varepsilon_{t+1}^{\mathbb{P}} - \frac{\gamma^2}{2}\right)\mathrm{e}^{-x\varepsilon_{t+1}^{\mathbb{P}}}\middle|\,\mathcal{F}_t\right] = \exp\left(\frac{1}{2}x^2 - x\gamma\right),$$

which corresponds to the MGF of a Gaussian random variable with expectation $-\gamma$ and variance 1. Thus, we construct a risk-neutral noise term centred at zero:

$$\varepsilon_{t+1}^{\mathbb{Q}} = \varepsilon_{t+1}^{\mathbb{P}} + \gamma.$$

Under the risk-neutral measure $\mathbb{Q}$, the asset return satisfies

$$R_{t+1} = m_{t+1} + \sigma_{t+1}\varepsilon_{t+1}^{\mathbb{P}} = m_{t+1} + \sigma_{t+1}\left(\varepsilon_{t+1}^{\mathbb{Q}} - \gamma\right) = \underbrace{m_{t+1} - \gamma\sigma_{t+1}}_{r_{t+1}} + \sigma_{t+1}\varepsilon_{t+1}^{\mathbb{Q}} = R_{t+1}^{\mathbb{Q}}.$$

Therefore, $m_{t+1} = r_{t+1} + \gamma\sigma_{t+1}$ and the risk-neutral debt ratio dynamics are

$$\begin{aligned} X_{t+1}^{\mathbb{Q}} &= \frac{(1+\delta_{t+1})(1+\eta_{t+1})X_t}{\left(1+R_{t+1}^{\mathbb{Q}}\right) - (\mu_{t+1}-\eta_{t+1})X_t} \qquad (A.1) \\ &= \frac{(1+\delta_{t+1})(1+\eta_{t+1})X_t}{\left(1+R_{t+1}^{\mathbb{Q}}\right) - (r_{t+1}(1-z_t)(1-y_t) + b(X_t,y_t)z_t(1-y_t) + c(X_t)y_t)X_t}\mathbf{1}_{\tau_C>t+1} \\ &\quad + \frac{(1+\delta_{t+1})(1+\eta_{t+1})X_t}{\left(1+R_{t+1}^{\mathbb{Q}}\right) - (r_{t+1}(1-z_t)(1-y_t) + b(X_t,y_t)z_t(1-y_t) + y_t)X_t}\mathbf{1}_{\tau_C=t+1} \\ &\quad + \frac{(1+\delta_{t+1})(1+\eta_{t+1})X_t}{\left(1+R_{t+1}^{\mathbb{Q}}\right) - (r_{t+1}(1-z_t) + b(X_t,0)z_t)X_t}\mathbf{1}_{\tau_C<t+1}. \end{aligned}$$

The conversion decision depends on the debt ratio, given that there is no dividend payment, $\delta_{t+1} = 0$, no change in debt structure, $\eta_{t+1} = 0$, and with the CoCo interest payment. More precisely,

$$X_{t+1}^{\mathbb{Q},C} = \frac{X_t}{1+R_{t+1}^{\mathbb{Q}} - (r_{t+1}(1-z_t)(1-y_t) + b(X_t,y_t)z_t(1-y_t) + c(X_t)y_t)X_t}$$

and

$$G_{t+1}^{\mathbb{Q}} = g\left(X_{t+1}^{\mathbb{Q},C}\right). \qquad (A.2)$$

The default intensity is computed after the conversion (if the latter has not already occurred):

$$X_{t+1}^{\mathbb{Q},D} = \frac{(1+\eta_{t+1})X_t}{\left(1+R_{t+1}^{\mathbb{Q}}\right) - (r_{t+1}(1-z_t)(1-y_t) + b(X_t,y_t)z_t(1-y_t) + y_t)X_t}\mathbf{1}_{\tau_C>t}$$

$$+\frac{X_t}{1+R^{\mathbb{Q}}_{t+1}-\left(r_{t+1}\left(1-z_t\right)+b\left(X_t,0\right)z_t\right)X_t}\mathbf{1}_{\tau_C\leq t},$$

and

$$H^{\mathbb{Q}}_{t+1}=h\left(X^{\mathbb{Q},D}_{t+1}\right). \tag{A.3}$$

A.1.2 Credit Sensitive Debt

Lemma 2. *Assume that the time $t+1$ value of the standard risky debt is*

$$B_{(t+1)^-}\mathbf{1}_{\tau_D>t}=B_t\left(\left(1+b\left(X_t,y_t\right)\right)\mathbf{1}_{\tau_D>t+1}+\rho_D\mathbf{1}_{\tau_D=t+1}\right)\mathbf{1}_{\tau_D>t}.$$

Then, the interest rate $b\left(X_t,y_t\right)$ of the standard debt satisfies

$$b\left(X^{\mathbb{Q}}_t,0\right)\mathbf{1}_{\tau_C\leq t}=\left[r_t+\left(1-\rho_D+r_t\right)\left(\frac{1}{\mathbb{Q}_t\left[\tau_D>t+1\,|\tau_D>t,\tau_C\leq t\right]}-1\right)\right]\mathbf{1}_{\tau_C\leq t} \tag{A.4}$$

and

$$b\left(X^{\mathbb{Q}}_t,y_t\right)\mathbf{1}_{\tau_C>t}=\left[r_t+\left(1-\rho_D+r_t\right)\left(\frac{1}{\mathbb{Q}_t\left[\tau_D>t+1\,|\tau_D>t,\tau_C>t\right]}-1\right)\right]\mathbf{1}_{\tau_C>t}, \tag{A.5}$$

where r_t is the risk-free rate and ρ_D represents the recovery rate.

See proof in Appendix A.2.1. The risk-neutral survival probabilities are provided by Lemma 6.

A.1.3 Convertible Contingent Debt

Lemma 3. *Assume that the time $t+1$ value of the convertible contingent debt is*

$$C_t\left(\left(1+c\left(X_t\right)\right)\mathbf{1}_{\tau_C>t+1}+\left(1+\rho_C c\left(X_t\right)\right)\mathbf{1}_{\tau_C=t+1}\right)\mathbf{1}_{\tau_D>t+1}.$$

It follows that the convertible contingent debt interest rate satisfies

$$\begin{aligned} & c\left(X_t\right)\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t} \\ = \; & \frac{\left(1+r_t\right)\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t}-\mathbb{Q}_t\left[\tau_D>t+1|\,\tau_C>t,\tau_D>t\right]}{\mathbb{Q}_t\left[\tau_C>t+1|\,\tau_C>t,\tau_D>t\right]+\rho_C\mathbb{Q}_t\left[\tau_D>t+1\textit{ and }\tau_\alpha=t+1|\,\tau_C>t,\tau_D>t\right]}\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t}. \end{aligned} \tag{A.6}$$

Remarks If $\rho_C = 1$, then $c(X_t) = b\left(X_t^{\mathbb{Q}}, y_t\right)$, where $\rho_D = 0$.
See proof in Appendix A.2.2. The risk-neutral conditional probabilities are detailed in Lemmas 4–6.

A.1.4 Conditional Probabilities

The proofs of the following lemmas are available in the Appendix A.2.5.

Lemma 4 (No conversion risk-neutral probability)**.**

$$\begin{aligned}\mathbb{Q}_t\left[\tau_C > t+1 \middle| \tau_\alpha > t, \tau_D > t\right] &= \mathbb{Q}_t\left[\tau_\alpha > t+1 \text{ and } \tau_D > t+1 \middle| \tau_\alpha > t, \tau_D > t\right] \\ &= \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}}\mathbf{1}_{\tau_C > t} - G_{t+1}^{\mathbb{Q}}\right)\middle| \tau_\alpha > t, \tau_D > t\right],\end{aligned}$$

where $G_{t+1}^{\mathbb{Q}}$ and $H_{t+1}^{\mathbb{Q}}$ are defined by Equations (A.2) and (A.3).

Lemma 5 (Conversion and survival risk-neutral probability)**.**

$$\begin{aligned}&\mathbb{Q}_t\left[\tau_\alpha = t+1 \text{ and } \tau_D > t+1 \middle| \tau_\alpha > t, \tau_D > t\right] \\ = \;& \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}}\mathbf{1}_{\tau_C > t}\right)\left(1 - \exp\left(-G_{t+1}^{\mathbb{Q}}\right)\right)\middle| \tau_C > t, \tau_D > t\right],\end{aligned}$$

where $G_{t+1}^{\mathbb{Q}}$ and $H_{t+1}^{\mathbb{Q}}$ are defined at Equations (A.2) and (A.3).

Lemma 6 (Survival risk-neutral probabilities)**.**

$$\begin{aligned}\mathbb{Q}_t\left[\tau_D > t+1 \middle| \tau_D > t, \tau_\alpha > t\right] &= \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}}\mathbf{1}_{\tau_C > t}\right)\middle| \tau_D > t, \tau_\alpha > t\right], \\ \mathbb{Q}_t\left[\tau_D > t+1 \middle| \tau_D > t, \tau_\alpha \le t\right] &= \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}}\mathbf{1}_{\tau_C \le t}\right)\middle| \tau_D > t, \tau_\alpha \le t\right],\end{aligned}$$

where $H_{t+1}^{\mathbb{Q}}$ is defined at Equation (A.3).

A.1.5 Approximations

A.1.5.1 Approximation of $b(x, 0)$

From Equation (A.1), we note that $X_{t+1}^{\mathbb{Q},D}$ can be viewed as a function of $R_{t+1}^{\mathbb{Q}}$. To emphasize this relation, we write $X_{t+1}^{\mathbb{Q},D} = X_{t+1}^{\mathbb{Q},D}\left(R_{t+1}^{\mathbb{Q}}\right)$. Since $R_{t+1}^{\mathbb{Q}}$ is centered at r_{t+1}, a Taylor

expansion of $\exp\left(-h\left(X_{t+1}^{\mathbb{Q},D}\right)\right)$ around $R_{t+1}^{\mathbb{Q}} = r_{t+1}$ leads to

$$\begin{aligned}
&\mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-h\left(X_{t+1}^{\mathbb{Q},D}\right)\right)\right]\mathbf{1}_{\tau_D>t} \cong \exp\left(-h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\right)\right)\\
&-\exp\left(-h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\right)\right)h'\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\right)\frac{\partial X_{t+1}^{\mathbb{Q},D}}{\partial R_{t+1}^{\mathbb{Q}}}\left(r_{t+1}\right)\underbrace{\mathrm{E}_t^{\mathbb{Q}}\left[\left(R_{t+1}^{\mathbb{Q}}-r_{t+1}\right)\right]}_{=0},
\end{aligned}$$

that is,

$$\begin{aligned}
&\mathbb{Q}_t\left[\tau_D > t+1\,|\tau_D > t, \tau_\alpha \le t\right]\\
\cong\ &\exp\left(-h\left(\frac{X_t}{1+r_{t+1}-\left(r_{t+1}\left(1-z_t\right)+b\left(X_t,0\right)z_t\right)X_t}\right)\right)\mathbf{1}_{\tau_C\le t}.
\end{aligned}$$

Placing it back in Equation (A.4),

$$b\left(x,0\right) \cong r_t + (1-\rho_D+r_t)\left(\exp\left(h\left(\frac{x}{1+r_{t+1}-\left(r_{t+1}\left(1-z_t\right)+b\left(x,0\right)z_t\right)x}\right)\right)-1\right).$$

A.1.5.2 Approximation of $b\left(x,y\right)$ and $c\left(x\right)$

Similarly, let

$$\begin{aligned}
X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t} &= \frac{\left(1-y_t\right)X_t}{\left(1+r_{t+1}\right)-\left(r_{t+1}\left(1-z_t\right)\left(1-y_t\right)+b\left(X_t,y_t\right)z_t\left(1-y_t\right)+y_t\right)X_t},\\
X_{t+1}^{\mathbb{Q},C}\left(r_{t+1}\right) &= \frac{X_t}{1+r_{t+1}-\left(r_{t+1}\left(1-z_t\right)\left(1-y_t\right)+b\left(X_t,y_t\right)z_t\left(1-y_t\right)+c\left(X_t\right)y_t\right)X_t}.
\end{aligned}$$

From Lemmas 4–6, we derive the following approximations:

$$\begin{aligned}
\mathbb{Q}_t\left[\tau_D > t+1\,|\tau_D > t, \tau_\alpha > t\right] &\cong \mathrm{e}^{-h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t}\right)},\\
\mathbb{Q}_t\left[\tau_C > t+1|\,\tau_C > t, \tau_D > t\right] &\cong \mathrm{e}^{-h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t}\right)-g\left(X_{t+1}^{\mathbb{Q},C}\left(r_{t+1}\right)\right)},\\
\mathbb{Q}_t\left[\tau_\alpha = t+1 \text{ and } \tau_D > t+1|\,\tau_\alpha > t, \tau_D > t\right] &\cong \mathrm{e}^{-h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t}\right)}\left(1-\mathrm{e}^{-g\left(X_{t+1}^{\mathbb{Q},C}\left(r_{t+1}\right)\right)}\right).
\end{aligned}$$

Placing it back in Equations (A.5) and (A.6) leads to

$$\begin{aligned}
b\left(X_t,y_t\right) &\cong r_t + (1-\rho_D+r_t)\left(\mathrm{e}^{h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t}\right)}-1\right),\\
c\left(X_t\right) &\cong \frac{\left(1+r_t\right)-\mathrm{e}^{-h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t}\right)}}{\mathrm{e}^{-h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t}\right)-g\left(X_{t+1}^{\mathbb{Q},C}\left(r_{t+1}\right)\right)}+\rho_C\mathrm{e}^{-h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t}\right)}\left(1-\mathrm{e}^{-g\left(X_{t+1}^{\mathbb{Q},C}\left(r_{t+1}\right)\right)}\right)}\\
&= \frac{\left(1+r_t\right)\exp\left(h\left(X_{t+1}^{\mathbb{Q},D}\left(r_{t+1}\right)\mathbf{1}_{\tau_C>t}\right)\right)-1}{\left(1-\rho_C\right)\exp\left(-g\left(X_{t+1}^{\mathbb{Q},C}\left(r_{t+1}\right)\right)\right)+\rho_C}.
\end{aligned}$$

A.2 Proofs

A.2.1 Standard Bond Floating Coupon Rate

Proof of Lemma 2. The time t price of a credit-sensitive debt is denoted B_t. At time $t+1$, there is an interest rate payment of $B_t b\left(X_t^{\mathbb{Q}}, y_t\right)$ if no default occurs and there is a recovery of $\rho_D B_{(t+1)^-}$ in case of default. We set the interest rate $b\left(X_t^{\mathbb{Q}}, y_t\right)$ such that the next period debt value, before the dividend payment, remains constant, that is, $B_{(t+1)^-} = B_t$. Interestingly, the floating rate is affected by the presence (or absence) of the CoCo debt instrument. We therefore study two cases: $\tau_C > t$ and $\tau_C \leq t$.

If $\tau_C \leq t$, then the standard risky bond pricing corresponds to the classic case:

$$\begin{aligned}
& B_t \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t} \\
= \; & \mathrm{E}_t^{\mathbb{Q}}\left[\frac{\left(B_t + B_t b\left(X_t^{\mathbb{Q}}, 0\right)\right)\mathbf{1}_{\tau_D > t+1} + \rho_D B_t \mathbf{1}_{\tau_D = t+1}}{1+r_t}\right]\mathbf{1}_{\tau_C \leq t}\mathbf{1}_{\tau_D > t} \\
= \; & B_t \mathrm{E}_t^{\mathbb{Q}}\left[\frac{\left(1 + b\left(X_t^{\mathbb{Q}}, 0\right)\right)\mathbf{1}_{\tau_D > t+1} + \rho_D\left(1 - \mathbf{1}_{\tau_D > t+1}\right)}{1+r_t}\right]\mathbf{1}_{\tau_C \leq t}\mathbf{1}_{\tau_D > t} \\
= \; & B_t \frac{\rho_D + \left(1 - \rho_D + b\left(X_t^{\mathbb{Q}}, 0\right)\right)\mathbb{Q}_t\left[\tau_D > t+1 \,|\tau_D > t, \tau_C \leq t\right]}{1+r_t}\mathbf{1}_{\tau_C \leq t}\mathbf{1}_{\tau_D > t}.
\end{aligned}$$

Since

$$\frac{\rho_D + \left(1 - \rho_D + b\left(X_t^{\mathbb{Q}}, 0\right)\right)\mathbb{Q}_t\left[\tau_D > t+1 \,|\tau_D > t, \tau_C \leq t\right]}{1+r_t}\mathbf{1}_{\tau_C \leq t}\mathbf{1}_{\tau_D > t} = \mathbf{1}_{\tau_C \leq t}\mathbf{1}_{\tau_D > t},$$

$$\begin{aligned}
& b\left(X_t^{\mathbb{Q}}, 0\right)\mathbf{1}_{\tau_C \leq t}\mathbf{1}_{\tau_D > t} \\
= \; & \left[r_t + (1 - \rho_D + r_t)\left(\frac{1}{\mathbb{Q}_t\left[\tau_D > t+1 \,|\tau_D > t, \tau_C \leq t\right]} - 1\right)\right]\mathbf{1}_{\tau_C \leq t}\mathbf{1}_{\tau_D > t}.
\end{aligned}$$

Similarly, if $\tau_C > t$, then the risky debt interest rate is

$$\begin{aligned} & b\left(X_t^{\mathbb{Q}}, y_t\right)\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t} \\ = & \left[r_t + (1-\rho_D+r_t)\left(\frac{1}{\mathbb{Q}_t\left[\tau_D > t+1\,|\tau_D > t, \tau_C > t\right]} - 1\right)\right]\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t}. \end{aligned}$$

□

A.2.2 Convertible Bond Floating Coupon Rate

Proof of Lemma 3. The goal is to choose the coupon rate $c\left(X_t\right)$ so that the convertible contingent debt value at time $t+1$ is the same as its value at time t, that is $C_{(t+1)^-} = C_t$. Assume that $\tau_C > t$ and that $\tau_D > t$, since, otherwise, the convertible instrument is worth 0. In that case, the time t value of the convertible contingent debt value satisfies

$$\begin{aligned} & C_t\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t} \\ = & \mathrm{E}_t^{\mathbb{Q}}\left[\frac{C_t + c\left(X_t\right)C_t}{1+r_t}\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha>t+1} + \frac{C_t + \rho_C c\left(X_t\right)C_t}{1+r_t}\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha=t+1}\right]\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t}, \end{aligned}$$

which is equivalent to

$$\begin{aligned} & \mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t} \\ = & \mathrm{E}_t^{\mathbb{Q}}\left[\frac{1 + c\left(X_t\right)}{1+r_t}\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha>t+1} + \frac{1 + \rho_C c\left(X_t\right)}{1+r_t}\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha=t+1}\right]\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t} \\ = & \frac{1}{1+r_t}\begin{pmatrix} \left(1 + c\left(X_t\right)\right)\mathrm{E}_t^{\mathbb{Q}}\left[\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha>t+1}|\,\tau_C > t, \tau_D > t\right] \\ + \left(1 + \rho_C c\left(X_t\right)\right)\mathrm{E}_t^{\mathbb{Q}}\left[\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha=t+1}|\,\tau_C > t, \tau_D > t\right] \end{pmatrix}\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t}. \end{aligned}$$

Consequently,

$$\begin{aligned} & c\left(X_t\right)\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t} \\ = & \frac{\left(1+r_t\right)\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t} - \mathrm{E}_t^{\mathbb{Q}}\left[\mathbf{1}_{\tau_D>t+1}|\,\tau_C > t, \tau_D > t\right]}{\mathrm{E}_t^{\mathbb{Q}}\left[\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha>t+1}|\,\tau_C > t, \tau_D > t\right] + \rho_C\mathrm{E}_t^{\mathbb{Q}}\left[\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha=t+1}|\,\tau_C > t, \tau_D > t\right]}\mathbf{1}_{\tau_C>t}\mathbf{1}_{\tau_D>t}. \end{aligned}$$

Lastly, note that $\mathbf{1}_{\tau_D>t+1}\mathbf{1}_{\tau_\alpha>t+1} = \mathbf{1}_{\tau_C>t+1}$. □

A.2.3 The Equity Value Variation

From Equations (2.7) and (2.8), it follows that

$$\begin{aligned} & E_{t+1} - E_t \\ = \; & A_{t+1} - D_{t+1} - E_t \\ = \; & \frac{(1+R_{t+1})\,A_t - (\mu_{t+1} - \eta_{t+1})\,D_t}{1+\delta_{t+1}} - D_t\,(1+\eta_{t+1}) - (A_t - D_t) \\ = \; & \left(\frac{R_{t+1} - \delta_{t+1}}{1+\delta_{t+1}} - \frac{\mu_{t+1} + \eta_{t+1}\delta_{t+1}}{1+\delta_{t+1}} X_t\right) A_t. \end{aligned}$$

Therefore,

$$\mathrm{E}_t^{\mathbb{P}}\left[E_{t+1}|_{\delta_{t+1}=0} - E_t\right] = \mathrm{E}_t^{\mathbb{P}}\left[R_{t+1} - \mu_{t+1}X_t\right] A_t = \left(m_{t+1} - \mathrm{E}_t^{\mathbb{P}}\left[\mu_{t+1}\right] X_t\right) A_t \mathbf{1}_{\tau_C \le t},$$

where

$$\begin{aligned} & \mathrm{E}_t^{\mathbb{P}}\left[\mu_{t+1}\right] \\ = \; & \left(r_{t+1}\,(1-z_t) + b\,(X_t, 0)\,z_t\right)\mathbf{1}_{\tau_C \le t} \\ & + \left(r_{t+1}\,(1-z_t)\,(1-y_t) + b\,(X_t, y_t)\,z_t\,(1-y_t)\right)\mathbb{P}_t\left[\tau_C = t+1\,|\tau_C > t\right]\mathbf{1}_{\tau_C > t} \\ & + \left(r_{t+1}\,(1-z_t)\,(1-y_t) + b\,(X_t, y_t)\,z_t\,(1-y_t) + c\,(X_t)\,y_t\right)\mathbb{P}_t\left[\tau_C > t+1\,|\tau_C > t\right]\mathbf{1}_{\tau_C > t}. \end{aligned}$$

A.2.4 The Value of Expected Discounted Dividends at T

Proof of Lemma 1. Since $u > T$, we can substitute $X_u = X_T$, $A_u = A_T$, $\mu_u = \mu_{T+1}$, $\delta_u^* = \delta_{T+1}^* = m_T - \mu_{T+1}X_T$, and $DF_{T+1,T+1+u} = (1 + w\,(X_T))^{-u}$ in Equation (2.21). Therefore,

$$\begin{aligned} & V\left(T+1, X_{T+1}^0\right)\mathbf{1}_{\tau_D > T+1} \\ \cong \; & \sum_{u=T+1}^{\infty} \mathrm{E}_{T+1}^{\mathbb{P}}\left[DF_{T+1,u}\,\delta_u A_u\left(1 - \frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C \le u}\right)\mathbf{1}_{\tau_D > u}\right]\mathbf{1}_{\tau_D > T+1} \\ = \; & \delta_{T+1}A_T\left(1 - \frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C \le T}\right)\sum_{u=T+1}^{\infty}\left(1 + w\,(X_T)\right)^{-(T+1-u)}\mathrm{E}_{T+1}^{\mathbb{P}}\left[\mathbf{1}_{\tau_D > u}\right]\mathbf{1}_{\tau_D > T+1}. \end{aligned}$$

Because the debt ratio remains constant over time, the default intensity is

$$H_u = H_T = \lambda_D + \left(\frac{1}{\theta_D}\max\left(X_T;0\right)\right)^{\beta_D}.$$

Therefore, the conditional survival probability $\mathrm{E}_{T+1}^{\mathbb{P}}\left[\mathbf{1}_{\tau_D>u}\right]\mathbf{1}_{\tau_D>T+1} = \exp\left(-\left(u-T-1\right)H_T\right)$ and

$$\begin{aligned} & V\left(T+1, X_{T+1}^0\right)\mathbf{1}_{\tau_D>T+1} \\ \cong\ & \delta_{T+1}A_T\left(1-\frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C\leq T}\right)\sum_{s=0}^{\infty}\left(\frac{\exp\left(-H_T\right)}{1+w\left(X_T\right)}\right)^s\mathbf{1}_{\tau_D>T+1}. \end{aligned}$$

The final result is obtained using the geometrical series property $\sum_{s=0}^{\infty}a^s = \left(1-a\right)^{-1}$, provided that $|a|<1$:

$$\begin{aligned} & V\left(T+1, X_{T+1}^0\right)\mathbf{1}_{\tau_D>T+1} \\ \cong\ & \delta_{T+1}A_T\left(1-\frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C\leq T}\right)\left(1+\frac{\exp\left(-H_T\right)}{1+w\left(X_T\right)-\exp\left(-H_T\right)}\right)\mathbf{1}_{\tau_D>T+1} \end{aligned}$$

and

$$\begin{aligned} & V\left(T, X_T^0, \delta_T\right)\mathbf{1}_{\tau_D>T} \\ \cong\ & \left\{\delta_T A_T\left(1-\frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C\leq T}\right)+\mathrm{E}_T^{\mathbb{P}}\left[\frac{V\left(T+1, X_{T+1}^0\right)}{1+w\left(X_T\right)}\mathbf{1}_{\tau_D>T+1}\right]\right\}\mathbf{1}_{\tau_D>T} \\ =\ & \left\{\delta_T A_T\left(1-\frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C\leq T}\right)+\mathrm{E}_T^{\mathbb{P}}\left[\frac{V\left(T+1, X_{T+1}^0\right)}{1+w\left(X_T\right)}\mathrm{E}_{T+1}^{\mathbb{P}}\left[\mathbf{1}_{\tau_D>T+1}\right]\right]\right\}\mathbf{1}_{\tau_D>T} \\ =\ & \left\{\delta_T A_T\left(1-\frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C\leq T}\right)+\mathrm{E}_T^{\mathbb{P}}\left[\frac{V\left(T+1, X_{T+1}^0\right)}{1+w\left(X_T\right)}\exp\left(-H_{T+1}\right)\right]\right\}\mathbf{1}_{\tau_D>T} \\ =\ & \left\{\delta_T A_T\left(1-\frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C\leq T}\right)+\frac{V\left(T+1, X_T^0\right)}{1+w\left(X_T\right)}\exp\left(-H_T\right)\right\}\mathbf{1}_{\tau_D>T} \\ =\ & A_T\left(1-\frac{y_0\alpha}{1-\alpha+y_0\alpha}\mathbf{1}_{\tau_C\leq T}\right)\left\{\delta_T+\delta_{T+1}\left(\frac{\exp\left(-H_T\right)}{1+w\left(X_T\right)-\exp\left(-H_T\right)}\right)\right\}\mathbf{1}_{\tau_D>T}. \end{aligned}$$

□

A.2.5 Proofs of Lemmas 4–6

Proof of Lemma 4. Note that $\{\tau_C > t-1\} = \{\tau_\alpha > t-1\} \cap \{\tau_D > t-1\}$. Once we condition on the time t debt ratio, the events $\{\tau_\alpha > t\}$ and $\{\tau_D > t\}$ are independent. Therefore,

$$\begin{aligned}
& \mathbb{Q}_t\left(\tau_\alpha > t \text{ and } \tau_D > t \,|\tau_\alpha > t-1, \tau_D > t-1\right) \\
= \;& \mathbb{Q}_t\left(\tau_D > t \,|\tau_\alpha > t-1, \tau_D > t-1\right) \mathbb{Q}_t\left(\tau_\alpha > t \,|\tau_\alpha > t-1, \tau_D > t-1\right) \\
= \;& \exp\left(-H_t^{\mathbb{Q}} \mathbf{1}_{\tau_C > t-1}\right) \exp\left(-G_t^{\mathbb{Q}}\right) \mathbf{1}_{\tau_\alpha > t-1} \mathbf{1}_{\tau_D > t-1}.
\end{aligned}$$

Finally, the law of iterated conditional expectation implies that

$$\begin{aligned}
& \mathbb{Q}_t\left(\tau_\alpha > t+1 \text{ and } \tau_D > t+1 \,|\tau_\alpha > t, \tau_D > t\right) \\
= \;& \mathrm{E}_t^{\mathbb{Q}}\left[\mathrm{E}_{t+1}^{\mathbb{Q}}\left[\mathbf{1}_{\tau_\alpha > t+1} \mathbf{1}_{\tau_D > t+1} | \, \tau_\alpha > t, \tau_D > t\right] \middle| \, \tau_\alpha > t, \tau_D > t\right] \\
= \;& \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}} \mathbf{1}_{\tau_C > t}\right) \exp\left(-G_{t+1}^{\mathbb{Q}}\right) \middle| \, \tau_\alpha > t, \tau_D > t\right].
\end{aligned}$$

□

Proof of Lemma 5. Given the time t debt ratio, the events $\{\tau_\alpha = t\}$ and $\{\tau_D > t\}$ are independent. Therefore,

$$\begin{aligned}
& \mathbb{Q}_t\left(\tau_\alpha = t \text{ and } \tau_D > t \,|\tau_\alpha > t-1, \tau_D > t-1\right) \\
= \;& \mathbb{Q}_t\left(\tau_D > t \,|\tau_\alpha > t-1, \tau_D > t-1\right) \mathbb{Q}_t\left(\tau_\alpha = t \,|\tau_\alpha > t-1, \tau_D > t-1\right) \\
= \;& \exp\left(-H_t^{\mathbb{Q}} \mathbf{1}_{\tau_C > t-1}\right) \left(1 - \exp\left(-G_t^{\mathbb{Q}}\right)\right) \mathbf{1}_{\tau_\alpha > t-1} \mathbf{1}_{\tau_D > t-1}
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{Q}_t\left(\tau_\alpha = t+1 \text{ and } \tau_D > t+1 \,|\tau_\alpha > t, \tau_D > t\right) \\
= \;& \mathrm{E}_t^{\mathbb{Q}}\left[\mathbf{1}_{\tau_\alpha = t+1} \mathbf{1}_{\tau_D > t+1} \,|\tau_\alpha > t, \tau_D > t\right] \\
= \;& \mathrm{E}_t^{\mathbb{Q}}\left[\mathrm{E}_{t+1}^{\mathbb{Q}}\left[\mathbf{1}_{\tau_\alpha = t+1} \mathbf{1}_{\tau_D > t+1} \,|\tau_\alpha > t, \tau_D > t\right] |\tau_\alpha > t, \tau_D > t\right] \\
= \;& \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}} \mathbf{1}_{\tau_C > t}\right) \left(1 - \exp\left(-G_{t+1}^{\mathbb{Q}}\right)\right) |\tau_\alpha > t, \tau_D > t\right].
\end{aligned}$$

□

Proof of Lemma 6. First, $\mathbb{Q}_t\left[\tau_D > t+1 | \, \tau_\alpha > t, \tau_D > t\right]$ is a consequence of Lemmas 4 and 5:

$$\begin{aligned}
& \mathbb{Q}_t\left[\tau_D > t+1 | \, \tau_\alpha > t, \tau_D > t\right] \\
= \; & \mathbb{Q}_t\left[\tau_\alpha > t+1 \text{ and } \tau_D > t+1 | \, \tau_\alpha > t, \tau_D > t\right] \\
& + \mathbb{Q}_t\left[\tau_\alpha = t+1 \text{ and } \tau_D > t+1 | \, \tau_\alpha > t, \tau_D > t\right] \\
= \; & \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}} \mathbf{1}_{\tau_C > t}\right) \middle| \, \tau_C > t, \tau_D > t\right].
\end{aligned}$$

We now compute $\mathbb{Q}_t\left[\tau_D > t+1 \, | \tau_D > t, \tau_\alpha \leq t\right]$. Since

$$\mathbb{Q}_t\left[\tau_D > t \, | \tau_\alpha \leq t-1, \tau_D > t-1\right] = \exp\left(-H_t^{\mathbb{Q}} \mathbf{1}_{\tau_C \leq t-1}\right) \mathbf{1}_{\tau_\alpha \leq t-1} \mathbf{1}_{\tau_D > t-1},$$

$$\begin{aligned}
\mathbb{Q}_t\left[\tau_D > t+1 \, | \tau_D > t, \tau_\alpha \leq t\right] &= \mathrm{E}_t^{\mathbb{Q}}\left[\mathbf{1}_{\tau_D > t+1} \, | \tau_D > t, \tau_\alpha \leq t\right] \\
&= \mathrm{E}_t^{\mathbb{Q}}\left[\mathrm{E}_{t+1}^{\mathbb{Q}}\left[\mathbf{1}_{\tau_D > t+1} \, | \tau_D > t, \tau_\alpha \leq t\right] | \tau_D > t, \tau_\alpha \leq t\right] \\
&= \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}} \mathbf{1}_{\tau_C \leq t}\right) \middle| \, \tau_D > t, \tau_\alpha \leq t\right].
\end{aligned}$$

□

A.3 Calibration of Default and Conversion Probabilities

We use the one-year default probability computed by Bloomberg in order to calibrate the one-year default probability. We also impose a default probability of 1.5% around the critical debt ratio imposed by the regulator. Using a log-linearization of Equation (2.19), the function H_{t+1} can be expressed such that

$$H_{t+1} = -\log(1 - \mathbb{P}_{t+1}\left(\tau_D = t+1 \, | \tau_C \leq t, \tau_D > t\right)).$$

Fixing the coefficient β_D, a log-linear regression can be done. We have

$$-\log\left(1 - \mathbb{P}_{t+1}(\tau_D = t+1 \, | \tau_C \leq t, \tau_D > t)\right) = \beta_0 + \beta_1 \left(X_{t+1}\right)^{\beta_D},$$

where $\beta_0 = \lambda$ and $\beta_1 = \left(\frac{1}{\theta_D}\right)^{\beta_D}$. A simple transformation of β_1 gives the estimated parameter θ_D:

$$\theta_D = \left(\frac{1}{\beta_1}\right)^{1/\beta_D}.$$

From Equation (2.14), we can isolate the parameter θ_C, such that

$$\theta_C = \frac{X_{t+1}^C}{\left(-\log(1 - \mathbb{P}_{t+1}(\tau_C = t+1 \,|\tau_C > t)) - H_{t+1}\mathbf{1}_{\tau_C > t+1}\right)^{1/\beta_C}}.$$

The conversion probability is calibrated such that there is a 10% conversion probability over the observed mean debt ratio and a 90% conversion probability at the trigger level imposed by the regulator. We are in the presence of two equations with two unknown parameters: θ_C and β_C. The observed mean debt ratio, denoted by $\bar{X}^{obs}$, corresponds to the average debt ratio from 2004 to 2017. The trigger level used, denoted by $\bar{X}_\alpha^{obs}$, is the average trigger level imposed by the regulator from 2004 to 2017. Solving the system of equation, we have

$$\theta_C = \exp\left\{\frac{\log(f_\alpha)\log(\bar{X}^{obs}) - \log(f_X)\log(\bar{X}_\alpha^{obs})}{\log(f_\alpha) - \log(f_X)}\right\},$$
$$\beta_C = \frac{\log(f_\alpha)\log(f_\alpha) - \log(f_\alpha)\log(f_X)}{\log(f_\alpha)\log(\bar{X}_\alpha^{obs}) - \log(f_\alpha)\log(\bar{X}^{obs})},$$

where

$$f_\alpha = -\log(1 - 0.9) - \left(\lambda + \left(\frac{\bar{X}_\alpha^{obs}}{\theta_C}\right)^{1/\beta_D}\right),$$
$$f_X = -\log(1 - 0.1) - \left(\lambda + \left(\frac{\bar{X}^{obs}}{\theta_C}\right)^{1/\beta_D}\right).$$

A.4 Finding the Optimal Dividend Rate Sequence

A.4.1 Post-Conversion Optimal Dividend Rates

We use the subscript C to indicate that the conversion occurred, that is, to indicate that $\tau_C \leq t$. Let $v_C\left(t, X_t^0, \delta_{t:\infty}\right) = A_t^{-1} V_C\left(t, X_t^0, \delta_{t:\infty}\right)$ and $v_C^*\left(t, X_t^0\right) = A_t^{-1} V_C\left(t, X_t^0, \delta_{t:\infty}^*\right)$,

where $\delta^*_{t:\infty}$ represents the optimal dividend rates from t onward when $\tau_C \leq t$. In that case, since

$$\frac{A_{t+1}}{A_t} = \frac{D_t}{A_t}\frac{D_{t+1}}{D_t}\frac{A_{t+1}}{D_{t+1}}\mathbf{1}_{\tau_C \leq t} = \frac{X_t}{X_{t+1}}\left(1+\eta_{t+1}\right)\mathbf{1}_{\tau_C \leq t} = \frac{X_t}{X_{t+1}},$$

Equation (2.22) becomes

$$\begin{aligned}\delta^*_t &= \arg\max_{\delta_t \in [0,\delta_t^{\max}]} A_t \left\{ \delta_t \left(1 - \frac{\left(1+\rho_C c\left(\alpha\right)\right) y_0 \alpha}{1-\alpha+\left(1+\rho_C c\left(\alpha\right)\right) y_0 \alpha} \mathbf{1}_{\tau_C \leq t} \right) \right. \\ &\quad \left. + \mathrm{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{v_C^*\left(t+1, X^0_{t+1}\right)}{1+w\left(X_t\right)} \mathbf{1}_{\tau_D > t+1} \right] \right\} \mathbf{1}_{\tau_D > t}.\end{aligned}$$

The conditional expectation is evaluated numerically. More precisely, let

$$0 = x_0 < x_1 < ... < x_n = 1$$

be a discretization of the pre-dividend debt ratio X^0_{t+1} support. For $i \in \{1, 2, ..., n\}$, $\xi_i = \left(x_{i-1} + x_i\right)/2$ is the mid-point of each interval. Assuming that $X^0_t = \xi_i$ and that the dividend rate for that particular state is $\delta_{t,i}$, Equations (2.10), (2.9), (2.6), (2.16), and (2.24) become respectively

$$\begin{aligned} X_{t,i} &= \left(1+\delta_{t,i}\right)\xi_i, \\ X_{t+1} &= \frac{\left(1+\delta_{t+1}\right)X_{t,i}}{\left(1+R_{t+1}\right) - \mu_{t+1,i}X_{t,i}}, \\ \mu_{t+1,i} &= r_{t+1}\left(1-z_t\right) + b\left(X_{t,i}, 0\right) z_t, \\ X^D_{t+1} &= \frac{X_{t,i}}{\left(1+R_{t+1}\right) - \mu_{t+1,i}X_{t,i}} = X_{t+1}|_{\delta_{t+1}=0}, \text{ and} \\ \delta^{\max}_{t,i} &= \min\left(\max\left(\frac{x^*_0 - \xi_i}{\xi_i}, 0 \right), m_{t+1} \right), \end{aligned}$$

where x^*_0 is the solution of $m_{t+1} - x^*_0\left[r_{t+1}\left(1-z_t\right) + b\left(x^*_0, 0\right) z_t\right] = 0$. As shown in the Appendix A.4.3, the transition probabilities are

$$\begin{aligned} \pi_{ij}\left(\delta_{t,i}\right) &= \mathbb{P}_t\left[x_{j-1} \leq X^0_{t+1} < x_j, \tau_D > t+1 \middle| \tau_C \leq t, X^0_t = \xi_i \right] \\ &\cong \left[\Phi\left(\varphi_{i,j-1}\left(\delta_{t,i}\right)\right) - \Phi\left(\varphi_{i,j}\left(\delta_{t,i}\right)\right)\right] \exp\left(-h\left(\xi_j\right)\right), \end{aligned} \tag{A.7}$$

where Φ is the cumulative distribution function of a standard normal random variable and

$$\varphi_{i,j}\left(\delta_{t,i}\right) = \frac{\left(x_j^{-1} + \mu_{t+1,i}\right) X_{t,i} - 1 - m_{t+1}}{\sigma_{t+1}}. \tag{A.8}$$

Lastly,

$$\begin{aligned} & \mathrm{E}_t^{\mathbb{P}}\left[\frac{X_{t,i}}{X_{t+1}}\frac{v_C^*\left(t+1, X_{t+1}^0\right)}{1+w\left(X_{t,i}\right)}\mathbf{1}_{\tau_D>t+1}\middle| X_t|_{\delta_t=0} = \xi_i, \tau_\alpha \le t, \tau_D > t\right] \\ \cong & \sum_{j=1}^{n}\frac{\left(1+\delta_{t,i}\right)\xi_i}{\left(1+\delta_{t+1,j}^*\right)\xi_j}\frac{v_C^*\left(t+1,\xi_j\right)}{1+w\left(\left(1+\delta_{t,i}\right)\xi_i\right)}\pi_{ij}\left(\delta_{t,i}\right), \end{aligned}$$

$$\begin{aligned} \delta_{t,i}^* = & \arg\max_{\delta_{t,i}\in\left[0,\delta_{t,i}^{\max}\right]} \delta_{t,i}\left(1 - \frac{\left(1+\rho_C c\left(\alpha\right)\right) y_0 \alpha}{1-\alpha+\left(1+\rho_C c\left(\alpha\right)\right) y_0 \alpha}\mathbf{1}_{\tau_C\le t}\right) \\ & + \sum_{j=1}^{n}\frac{\left(1+\delta_{t,i}\right)\xi_i}{\left(1+\delta_{t+1,j}^*\right)\xi_j}\frac{v_C^*\left(t+1,\xi_j\right)}{1+w\left(\left(1+\delta_{t,i}\right)\xi_i\right)}\pi_{ij}\left(\delta_{t,i}\right), \end{aligned} \tag{A.9}$$

and

$$\begin{aligned} v_C^*\left(t,\xi_i\right) = & \delta_{t,i}^*\frac{\left(1+\rho_C c\left(\alpha\right)\right) y_0 \alpha}{1-\alpha+\left(1+\rho_C c\left(\alpha\right)\right) y_0 \alpha}\mathbf{1}_{\tau_C\le t} \\ & + \sum_{j=1}^{n}\frac{\left(1+\delta_{t,i}^*\right)\xi_i}{\left(1+\delta_{t+1,j}^*\right)\xi_j}\frac{v_C^*\left(t+1,\xi_j\right)}{1+w\left(\left(1+\delta_{t,i}^*\right)\xi_i\right)}\pi_{ij}\left(\delta_{t,i}^*\right). \end{aligned} \tag{A.10}$$

A.4.2 Pre-Conversion Optimal Dividend Rates

We use the subscript NC to indicate that the CoCo debt is not yet converted, that is, to indicate that $\tau_C > t$. Let $v_{NC}\left(t, X_t^0, \delta_{t:\infty}\right) = A_t^{-1} V_{NC}\left(t, X_t^0, \delta_{t:\infty}\right)$ and $v_{NC}^{**}\left(t, X_t^0\right) = A_t^{-1} V_{NC}\left(t, X_t^0, \delta_{t:\infty}^{\text{opt}}\right)$, where $\delta_{t:\infty}^{\text{opt}}$ is the optimal dividend rate sequence:

$$\delta_{t+1}^{\text{opt}} = \delta_{t+1}^* \mathbf{1}_{\tau_C = t+1} + \delta_{t+1}^{**}\mathbf{1}_{\tau_C > t+1}.$$

In that case, Equation (2.22) becomes

$$\begin{aligned}
& \delta_t^{**} && \text{(A.11)}\\
= \; & \arg\max_{\delta_t \in [0,\delta_t^{\max}]} A_t \left\{ \delta_t + \mathrm{E}_t^{\mathbb{P}} \left[(1+\eta_{t+1}) \frac{X_t}{X_{t+1}} \frac{A_{t+1}^{-1} V_C\left(t+1, X_{t+1}, \delta^*_{t+1:\infty}\right)}{1+w\left(X_t\right)} \mathbf{1}_{\tau_C = t+1} \mathbf{1}_{\tau_D > t+1} \right] \right. \\
& \left. + \mathrm{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{A_{t+1}^{-1} V_{NC}\left(t+1, X^0_{t+1}, \delta^{\text{opt}}_{t+1:\infty}\right)}{1+w\left(X_t\right)} \mathbf{1}_{\tau_C > t+1} \mathbf{1}_{\tau_D > t+1} \right] \right\} \mathbf{1}_{\tau_D > t} \\
= \; & \arg\max_{\delta_t \in [0,\delta_t^{\max}]} A_t \left\{ \delta_t + (1-y_t)\, \mathrm{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{v_C^*\left(t+1, X^0_{t+1}\right)}{1+w\left(X_t\right)} \mathbf{1}_{\tau_C = t+1} \mathbf{1}_{\tau_D > t+1} \right] \right. && \text{(A.12)}\\
& \left. + \mathrm{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{v_{NC}^{**}\left(t+1, X^0_{t+1}\right)}{1+w\left(X_t\right)} \mathbf{1}_{\tau_C > t+1} \mathbf{1}_{\tau_D > t+1} \right] \right\} \mathbf{1}_{\tau_D > t}.
\end{aligned}$$

Assuming that $X_t^0 = \xi_i$ and that the dividend rate for that particular state is $\delta_{t,i}$, then Equations (2.10), (2.9), (2.6), (2.13), (2.16) and (2.24) become

$$\begin{aligned}
X_{t,i} &= (1+\delta_{t,i})\,\xi_i \\
X_{t+1} &= \frac{(1+\delta_{t+1})\,(1-y_t)\,X_{t,i}}{(1+R_{t+1}) - (\mu_{t+1,i} - c\left(X_{t,i}\right) y_t + y_t)\,X_{t,i}} \mathbf{1}_{\tau_\alpha = t+1} + \frac{(1+\delta_{t+1})\,X_{t,i}}{(1+R_{t+1}) - \mu_{t+1,i} X_{t,i}} \mathbf{1}_{\tau_\alpha > t+1}, \\
\mu_{t+1,i} &= r_{t+1}\left(1-z_t\right)\left(1-y_t\right) + b\left(X_{t,i}, y_t\right) z_t \left(1-y_t\right) + c\left(X_{t,i}\right) y_t, \\
X_{t+1}^C &= X^0_{t+1}\big|_{\tau_a > t+1} = \frac{X^0_{t+1}\big|_{\tau_a = t+1}}{(1-y_t) + (1-c\left(X_{t,i}\right))\, y_t\; X^0_{t+1}\big|_{\tau_a = t+1}}, \\
X_{t+1}^D &= X^0_{t+1}\big|_{\tau_a = t+1} = \frac{(1-y_t)\; X^0_{t+1}\big|_{\tau_a > t+1}}{1 - (1-c\left(X_{t,i}\right))\, y_t\; X^0_{t+1}\big|_{\tau_a > t+1}}, \text{ and} \\
\delta_{t,i}^{\max} &= \min\left(\max\left(\frac{x_0^{**} - \xi_i}{\xi_i}, 0 \right), m_{t+1} \right),
\end{aligned}$$

where x_0^{**} is the solution of $m_{t+1} - x_0^{**}\left[r_{t+1}\left(1-z_t\right)\left(1-y_t\right) + b\left(x_0^{**}, y_t\right) z_t \left(1-y_t\right) + c\left(x_0^{**}\right) y_t\right] = 0$. As shown in Appendix A.4.3, the transition probabilities are

$$\begin{aligned}
& \pi_{ij}^*\left(\delta_{t,i}\right) && \text{(A.13)}\\
= \; & \mathbb{P}_t\left[x_{j-1} \leq X^0_{t+1}\big|_{\tau_a = t+1} < x_j, \tau_\alpha = t+1, \tau_D > t+1 \,\middle|\, \tau_C > t,\, X_t\big|_{\delta_t = 0} = \xi_i \right] \\
\cong \; & \left(1 - \exp\left(-g\left(\frac{X^0_{t+1}\big|_{\tau_a = t+1}}{(1-y_t) + (1-c\left(X_{t,i}\right))\, y_t\; X^0_{t+1}\big|_{\tau_a = t+1}} \right) \right) \right) \exp\left(-h\left(\xi_j\right)\right)
\end{aligned}$$

$$\times \left[\Phi\left(\varphi^*_{i,j-1}\left(\delta_{t,i}\right)\right) - \Phi\left(\varphi^*_{i,j}\left(\delta_{t,i}\right)\right)\right],$$

with

$$\varphi^*_{i,j}\left(\delta_{t,i}\right) = \frac{\left[x_j^{-1}\left(1-y_t\right) + \mu_{t+1,i} + \left(1-c\left(X_{t,i}\right)\right)y_t\right]X_{t,i} - 1 - m_{t+1}}{\sigma_{t+1}} \tag{A.14}$$

and

$$\begin{aligned} & \pi^{**}_{ij}\left(\delta_{t,i}\right) \qquad\qquad (A.15)\\ = \;& \mathbb{P}_t\left[x_{j-1} \leq X^0_{t+1}\big|_{\tau_a > t+1} < x_j, \tau_\alpha > t+1, \tau_D > t+1 \,\middle|\, \tau_C > t,\, X_t|_{\delta_t=0} = \xi_i\right] \\ \cong \;& \exp\left(-g\left(\xi_j\right) - h\left(\frac{\left(1-y_t\right)\xi_j}{1-\left(1-c\left(X_{t,i}\right)\right)y_t\xi_j}\right)\right)\left[\Phi\left(\varphi^{**}_{i,j-1}\left(\delta_{t,i}\right)\right) - \Phi\left(\varphi^{**}_{i,j}\left(\delta_{t,i}\right)\right)\right], \end{aligned}$$

with

$$\varphi^{**}_{i,j}\left(\delta_{t,i}\right) = \frac{\left(x_j^{-1} + \mu_{t+1,i}\right)X_{t,i} - 1 - m_{t+1}}{\sigma_{t+1}}. \tag{A.16}$$

The conditional expectations are approximated with

$$\begin{aligned} & \mathrm{E}^{\mathbb{P}}_t\left[\frac{X_t}{X_{t+1}}\frac{v^*_C\left(t+1, X^0_{t+1}\right)}{1+w\left(X_t\right)}\mathbf{1}_{\tau_C=t+1}\mathbf{1}_{\tau_D>t+1}\,\middle|\, X_t|_{\delta_t=0} = \xi_i, \tau_C > t\right] \\ \cong \;& \sum_{j=1}^{n}\frac{\left(1+\delta_{t,i}\right)\xi_i}{\left(1+\delta^*_{t+1,j}\right)\xi_j}\frac{v^*_C\left(t+1,\xi_j\right)}{1+w\left(\left(1+\delta_{t,i}\right)\xi_i\right)}\pi^*_{ij}\left(\delta_{t,i}\right) \end{aligned}$$

and

$$\begin{aligned} & \mathrm{E}^{\mathbb{P}}_t\left[\frac{X_t}{X_{t+1}}\frac{v^{**}_{NC}\left(t+1, X^0_{t+1}\right)}{1+w\left(X_t\right)}\mathbf{1}_{\tau_C>t+1}\mathbf{1}_{\tau_D>t+1}\,\middle|\, X_t|_{\delta_t=0} = \xi_i, \tau_C > t\right] \\ \cong \;& \sum_{j=1}^{n}\frac{\left(1+\delta_{t,i}\right)\xi_i}{\left(1+\delta^{**}_{t+1,j}\right)\xi_j}\frac{v^{**}_{NC}\left(t+1,\xi_j\right)}{1+w\left(\left(1+\delta_{t,i}\right)\xi_i\right)}\pi^{**}_{ij}\left(\delta_{t,i}\right). \end{aligned}$$

Lastly,

$$\begin{aligned} & \delta^{**}_{t,i} \\ = \;& \arg\max_{\delta_{t,i}\in\left[0,\delta^{\max}_{t,i}\right]} A_t\left\{\delta_{t,i} + \left(1-y_t\right)\sum_{j=1}^{n}\frac{\left(1+\delta_{t,i}\right)\xi_i}{\left(1+\delta^*_{t+1,j}\right)\xi_j}\frac{v^*_C\left(t+1,\xi_j\right)}{1+w\left(\left(1+\delta_{t,i}\right)\xi_i\right)}\pi^*_{ij}\left(\delta_{t,i}\right)\right. \end{aligned}$$

$$+\sum_{j=1}^{n}\frac{(1+\delta_{t,i})\,\xi_i}{\left(1+\delta^{**}_{t+1,j}\right)\xi_j}\frac{v^{**}_{NC}\left(t+1,\xi_j\right)\mathbf{1}_{\tau_C>t+1}}{1+w\left(\left(1+\delta_{t,i}\right)\xi_i\right)}\pi^{**}_{ij}\left(\delta_{t,i}\right)\Bigg\}\mathbf{1}_{\tau_D>t}.$$

and

$$\begin{aligned}v^{*}_{NC}\left(t,\xi_i\right)\cong\delta^{**}_{t,i}&+(1-y_t)\sum_{j=1}^{n}\frac{\left(1+\delta^{**}_{t,i}\right)\xi_i}{\left(1+\delta^{*}_{t+1,j}\right)\xi_j}\frac{v^{*}_{C}\left(t+1,\xi_j\right)}{1+w\left(\left(1+\delta^{**}_{t,i}\right)\xi_i\right)}\pi^{*}_{ij}\left(\delta^{**}_{t,i}\right)\\&+\sum_{j=1}^{n}\frac{\left(1+\delta^{**}_{t,i}\right)\xi_i}{\left(1+\delta^{**}_{t+1,j}\right)\xi_j}\frac{v^{**}_{NC}\left(t+1,\xi_j\right)}{1+w\left(\left(1+\delta^{**}_{t,i}\right)\xi_i\right)}\pi^{**}_{ij}\left(\delta^{**}_{t,i}\right).\end{aligned}\tag{A.17}$$

A.4.3 Proofs

A.4.3.1 Proof of Equation (A.7)

$$\begin{aligned}&\mathbb{P}_t\left[x_{j-1}\le X^0_{t+1}<x_j\middle|\,\tau_C\le t,X^0_t=\xi_i\right]\\=\;&\mathbb{P}_t\left[x_{j-1}\le\frac{X_{t,i}}{(1+R_{t+1})-\mu_{t+1,i}X_{t,i}}<x_j\middle|\,\tau_C\le t,X^0_t=\xi_i\right]\\=\;&\mathbb{P}_t\left[x_j^{-1}<\frac{(1+R_{t+1})-\mu_{t+1,i}X_{t,i}}{X_{t,i}}\le x_{j-1}^{-1}\middle|\,\tau_C\le t,X^0_t=\xi_i\right]\\=\;&\mathbb{P}_t\left[\varphi_{i,j}\left(\delta_{t,i}\right)<\frac{R_{t+1}-m_{t+1}}{\sigma_{t+1}}\le\varphi_{i,j-1}\left(\delta_{t,i}\right)\middle|\,\tau_C\le t,X^0_t=\xi_i\right],\end{aligned}$$

where the $\varphi_{i,j}\left(\delta_{t,i}\right)$ are provided in Equation (A.8). The proof is completed by noting that

$$\left.\frac{R_{t+1}-m_{t+1}}{\sigma_{t+1}}\right|_{\mathcal{F}_t}$$

is a standard normal random variable. Finally,

$$\begin{aligned}&\mathbb{P}_t\left[x_{j-1}\le X^0_{t+1}<x_j,\tau_D>t+1\middle|\,\tau_C\le t,X^0_t=\xi_i\right]\\=\;&\mathbb{P}_t\left[\tau_D>t+1|\,\tau_C\le t,X^0_t=\xi_i,x_{j-1}\le X^0_{t+1}<x_j\right]\mathbb{P}_t\left[x_{j-1}\le X_{t+1}|_{\delta_{t+1}=0}<x_j\middle|\,\tau_C\le t,X^0_t=\xi_i\right]\\\cong\;&\exp\left(-h\left(\xi_j\right)\right)\mathbb{P}_t\left[x_{j-1}\le X_{t+1}|_{\delta_{t+1}=0}<x_j\middle|\,\tau_C\le t,X^0_t=\xi_i\right].\end{aligned}$$

A.4.3.2 Proof of Equation (A.13)

The probability that $X^0_{t+1}\big|_{\tau_a=t+1}$ lies between x_{j-1} and x_j is

$$\begin{aligned}
& \mathbb{P}_t\left[x_{j-1} \leq X^0_{t+1}\big|_{\tau_a=t+1} < x_j \middle| \tau_C > t, X^0_t = \xi_i\right] \\
= \; & \mathbb{P}_t\left[x_{j-1} \leq \frac{(1-y_t)\,X_{t,i}}{(1+R_{t+1}) - (\mu_{t+1,i} - c\,(X_{t,i})\,y_t + y_t)\,X_{t,i}} < x_j \middle| \tau_C > t, X^0_t = \xi_i\right] \\
= \; & \mathbb{P}_t\left[x_j^{-1} \leq \frac{(1+R_{t+1}) - (\mu_{t+1,i} - c\,(X_{t,i})\,y_t + y_t)\,X_{t,i}}{(1-y_t)\,X_{t,i}} < x_{j-1}^{-1} \middle| \tau_C > t, X^0_t = \xi_i\right] \\
= \; & \mathbb{P}_t\left[\varphi^*_{i,j}\left(\delta_{t,i}\right) < \frac{R_{t+1} - m_{t+1}}{\sigma_{t+1}} \leq \varphi^*_{i,j-1}\left(\delta_{t,i}\right) \middle| \tau_C > t, X^0_t = \xi_i\right] \\
= \; & \Phi\left(\varphi^*_{i,j-1}\left(\delta_{t,i}\right)\right) - \Phi\left(\varphi^*_{i,j}\left(\delta_{t,i}\right)\right),
\end{aligned}$$

where $\varphi^*_{i,j}\left(\delta_{t,i}\right)$ is defined at Equation (A.14).

The probability that $X^0_{t+1}\big|_{\tau_a=t+1}$ is contained between x_{j-1} and x_j while the conversion occurs without the firm default is

$$\begin{aligned}
& \mathbb{P}_t\left[x_{j-1} \leq X^0_{t+1}\big|_{\tau_a=t+1} < x_j, \tau_\alpha = t+1, \tau_D > t+1 \middle| \tau_C > t, X^0_t = \xi_i\right] \\
= \; & \mathbb{P}_t\left[\tau_\alpha = t+1 \middle| x_{j-1} \leq X^0_{t+1}\big|_{\tau_a=t+1} < x_j, \tau_C > t, X^0_t = \xi_i\right] \\
& \times \mathbb{P}_t\left[\tau_D > t+1 \middle| x_{j-1} \leq X^0_{t+1}\big|_{\tau_a=t+1} < x_j, \tau_C > t, X^0_t = \xi_i\right] \\
& \times \mathbb{P}_t\left[x_{j-1} \leq X^0_{t+1}\big|_{\tau_a=t+1} < x_j \middle| \tau_C > t, X^0_t = \xi_i\right] \\
\cong \; & \left(1 - \exp\left(g\left(\frac{X^0_{t+1}\big|_{\tau_a=t+1}}{(1-y_t) + (1-c\,(X_{t,i}))\,y_t\,X^0_{t+1}\big|_{\tau_a=t+1}}\right)\right)\right) \\
& \times \exp\left(-h\left(\xi_j\right)\right)\left[\Phi\left(\varphi^*_{i,j-1}\left(\delta_{t,i}\right)\right) - \Phi\left(\varphi^*_{i,j}\left(\delta_{t,i}\right)\right)\right].
\end{aligned}$$

A.4.3.3 Proof of Equation (A.15)

$$\begin{aligned}
& \mathbb{P}_t\left[x_{j-1} \leq X^0_{t+1}\big|_{\tau_a>t+1} < x_j \middle| \tau_C > t, X^0_t = \xi_i\right] \\
= \; & \mathbb{P}_t\left[x_{j-1} \leq \frac{X_{t,i}}{(1+R_{t+1}) - \mu_{t+1,i}X_{t,i}} < x_j \middle| \tau_C > t, X^0_t = \xi_i\right] \\
= \; & \mathbb{P}_t\left[x_j^{-1} < \frac{(1+R_{t+1}) - \mu_{t+1,i}X_{t,i}}{X_{t,i}} \leq x_{j-1}^{-1} \middle| \tau_C > t, X^0_t = \xi_i\right]
\end{aligned}$$

$$= \quad \mathbb{P}_t \left[\varphi_{i,j}^{**}\left(\delta_{t,i}\right) < \frac{R_{t+1} - m_{t+1}}{\sigma_{t+1}} \leq \varphi_{i,j-1}^{**}\left(\delta_{t,i}\right) \middle| \tau_C > t, X_t^0 = \xi_i \right],$$

where the $\varphi_{i,j}^{**}\left(\delta_{t,i}\right)$ are provided in Equation (A.16).

Chapter B

Appendices of *Predicting the Term Structure of Credit Spreads for High- and Low-Frequency Issuers: A Combined Nelson-Siegel and Clustering Approach*

B.1 The Linear Nelson-Siegel and Svenson Models of Gauthier and Simonato, 2012

This model facilitates the estimation of the Nelson-Siegel and Svensson zero-coupon (ZC) yield curves using coupon-bearing bond prices by:

1. Linearizing the model with respect to the level, slope, and curvature coefficients by applying a Taylor approximation to the discount factor.

2. Incorporating prior information to assist the estimation procedure through a Bayesian approach

Linearization is feasible when the term inside the exponential function from Equation (3.1) is close to zero. Because this condition typically does not hold in our context, we introduce an additional weight (ξ) reflecting the mean yield to satisfy this constraint. Thus,

the discount factor is rewritten as:

$$P\left(\theta_i, T\right) = e^{-\xi T} e^{(\xi - y^{ZC}(\theta,T))T}. \tag{B.1}$$

We linearize the model using the Taylor approximation $e^x \approx 1 + x$. The model is then estimated by replacing $y^{ZC}(\theta, T)$ with either the Nelson-Siegel formulation. A Bayesian approach based on the Theil–Goldberger estimator (Goldberger, 1964) is employed to incorporate prior information. Given that coefficients should exhibit limited day-to-day variation, prior information provides a valuable reference for today's coefficient values.

Ultimately, the model is linear in the level, slope, and curvature coefficients (the betas). Therefore, we create a grid of possible τ values and estimate a Bayesian linear model for each τ . The optimal τ is selected as the one minimizing the sum of squared errors.

To validate the linear approximation, we adopt a two-step estimation procedure. First, we estimate the linearized model using Generalized Least Squares (GLS). Further details are provided in the internet appendix of Gauthier and Simonato, 2012. In the second step, we estimate the fully nonlinear model using the linearized coefficients from the first step as initial values for optimization.

B.2 Linearized Nelson-Sigel ZC Spread Curve for Coupon-Bearing Bond Price With Prior Information

From Eq.(4.3), we introduce an additional weight ξ to get

$$B_i = \sum_{j=1}^{m_i} c_{i,j} e^{-y^{ZC}(\theta^{(Gov)}, t_{i,j}) t_{i,j}} e^{-\xi t_{i,j}} e^{(\xi - y^{ZC}(\theta^{(issuer)}, t_{i,j})) t_{i,j}}.$$

The last exponential is linearized to obtain

$$\begin{aligned} B_i &= \sum_{j=1}^{m_i} c_{i,j} e^{-y^{ZC}(\theta^{(Gov)}, t_{i,j}) t_{i,j}} e^{-\xi t_{i,j}} \left(1 + \xi t_{i,j} - y^{ZC}(\theta^{(issuer)}, t_{i,j}) t_{i,j}\right) \\ &= \sum_{j=1}^{m_i} \tilde{c}_{i,j} \psi_0(t_{i,j}) - \beta_0^{(2)} \sum_{j=1}^{m_i} \tilde{c}_{i,j} \psi_1(t_{i,j}) - \beta_1^{(2)} \sum_{j=1}^{m_i} \tilde{c}_{i,j} \psi_2(t_{i,j}) - \beta_2^{(2)} \sum_{j=1}^{m_i} \tilde{c}_{i,j} \psi_3(t_{i,j}), \end{aligned} \tag{B.2}$$

with

$$\begin{aligned}
\tilde{c}_{i,j} &= c_{i,j} e^{-y^{ZC}(\theta^{(Gov)}, t_{i,j}) t_{i,j}}, \\
\psi_0(t_{i,j}) &= e^{-\xi t_{i,j}} \left(1 + \xi t_{i,j}\right), \\
\psi_1(t_{i,j}) &= e^{-\xi t_{i,j}} t_{i,j}, \\
\psi_2(t_{i,j}) &= e^{-\xi t_{i,j}} \phi_1(t_{i,j}, \tau) t_{i,j}, \\
\psi_3(t_{i,j}) &= e^{-\xi t_{i,j}} \phi_2(t_{i,j}, \tau) t_{i,j}.
\end{aligned}$$

Fixing ξ and τ, Equation(B.2) can be solve by an OLS estimation. Assuming n bonds prices observed, the weighted bond prices may be written as a linear system $\boldsymbol{Y}_\xi = \boldsymbol{X}_{\xi,\tau}\boldsymbol{\theta}_{\xi,\tau} + \boldsymbol{e}$ where $\boldsymbol{X}_{\xi,\tau} = (X_1, X_2, X_3)$ with the $n \times 1$ vectors $X_i = \begin{pmatrix} \frac{1}{P_1} \sum_{j=1}^{m_1} \tilde{c}_{1,j} \psi_i(t_{1,j}) \\ \vdots \\ \frac{1}{P_n} \sum_{j=1}^{m_n} \tilde{c}_{n,j} \psi_i(t_{n,j}) \end{pmatrix}$ and $\boldsymbol{Y}_\xi = \begin{pmatrix} \frac{1}{P_1} \left[\sum_{j=1}^{m_1} \tilde{c}_{1,j} \psi_0(t_{1,j}) - P_1\right] \\ \vdots \\ \frac{1}{P_n} \left[\sum_{j=1}^{m_n} \tilde{c}_{n,j} \psi_0(t_{1,j}) - P_n\right] \end{pmatrix}$, for $i = 1, 2, 3$ where each elements of the system of equation are weighted with the inverse of the modified bond duration.

To incorporate the prior information, the system becomes

$$\begin{pmatrix} \boldsymbol{Y}_\xi \\ \theta_{prior} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{\xi,\tau} \\ \boldsymbol{R} \end{pmatrix} \boldsymbol{\theta}_{\xi,\tau} + \begin{pmatrix} \boldsymbol{e} \\ \boldsymbol{d} \end{pmatrix}. \tag{B.3}$$

θ_{prior} is an $r \times 1$ vector containing the prior values about the parameter; $\boldsymbol{R}$ is a $r \times 3$ matrix describing the linear combinations of the parameter; $\boldsymbol{d} = (d_1, d_2, d_3)^\top$ is a Gaussian random vector with $\mathbb{E}[\boldsymbol{d}] = 0$, $\mathbb{E}\left[\boldsymbol{d}\boldsymbol{e}^\top\right] = 0$, and diagonal covariance matrix of dimension r denoted by $\boldsymbol{\Sigma}_d = \mathbb{E}\left[\boldsymbol{d}\boldsymbol{d}^\top\right]$. In this framework, a GLS estimation is used instead of the OLS estimation, which leads to $\boldsymbol{\theta}_{\xi,\tau}^{Bayes} = \left(\tilde{\boldsymbol{X}}^\top \boldsymbol{\Omega}^{-1} \tilde{\boldsymbol{X}}\right)^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Omega}^{-1} \tilde{\boldsymbol{Y}}$ with $\tilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_{\xi,\tau} \\ \boldsymbol{R} \end{pmatrix}$, $\tilde{\boldsymbol{Y}} = \begin{pmatrix} \boldsymbol{Y}_\xi \\ \theta_{prior} \end{pmatrix}$, and $\boldsymbol{\Omega} = \begin{pmatrix} \sigma_e^2 \mathbb{I}_n & \boldsymbol{0}_{n\times r} \\ \boldsymbol{0}_{n\times r} & \boldsymbol{\Sigma}_d \end{pmatrix}$.

B.3 Time Series of the Coefficients Set for USD Corporate Issuers

The following graphics visually present the estimated Nelson-Siegel spread coefficients and the corresponding root mean squared relative errors (RMSRE) for each issuer over the sample period. For the six issuers considered, long-term yield spreads range between 1% and 3.5%, while the slope coefficient tends to remain negative. The τ coefficient is relatively stable over time. Notably, RBC, Walt Disney, and Verizon exhibit τ values between 1.5 and 3 years, whereas Goldman Sachs, Johnson & Johnson, and Amgen show higher and τ values ranging from 4 to 8 years.

Figure B.1. Time Series of the NS Coefficients for the Credit Spread Curve from 2016 to 2019 for RBC (id=270)

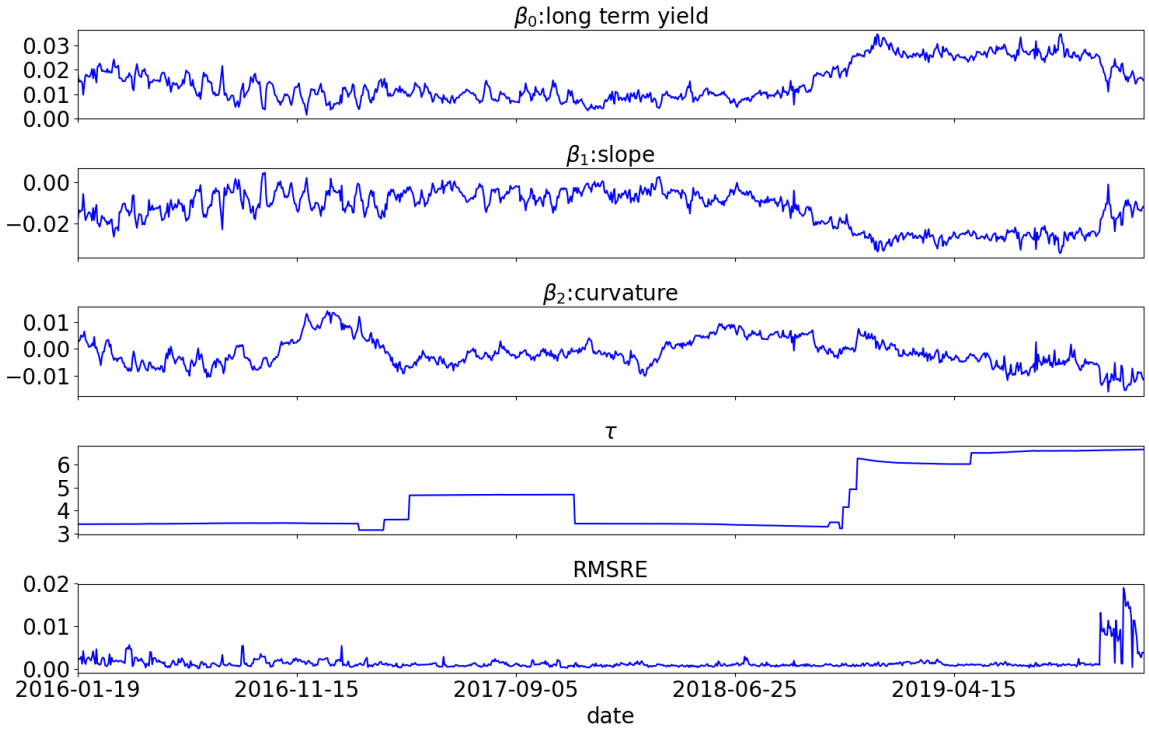


Figure B.2. Time Series of the NS Coefficients for the Credit Spread Curve from 2015 to 2019 for Goldman Sachs (id=580)

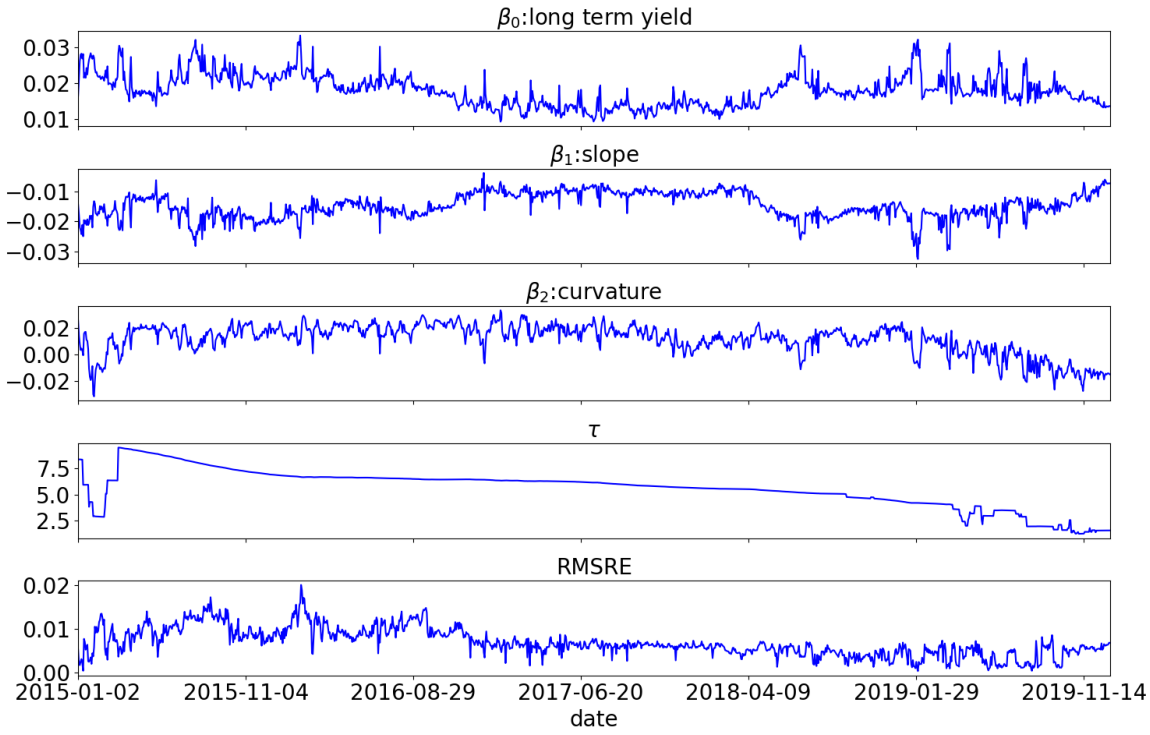


Figure B.3. Time Series of the NS Coefficients for the Credit Spread Curve from 2015 to 2019 for Johnson & Johnson (id=939)

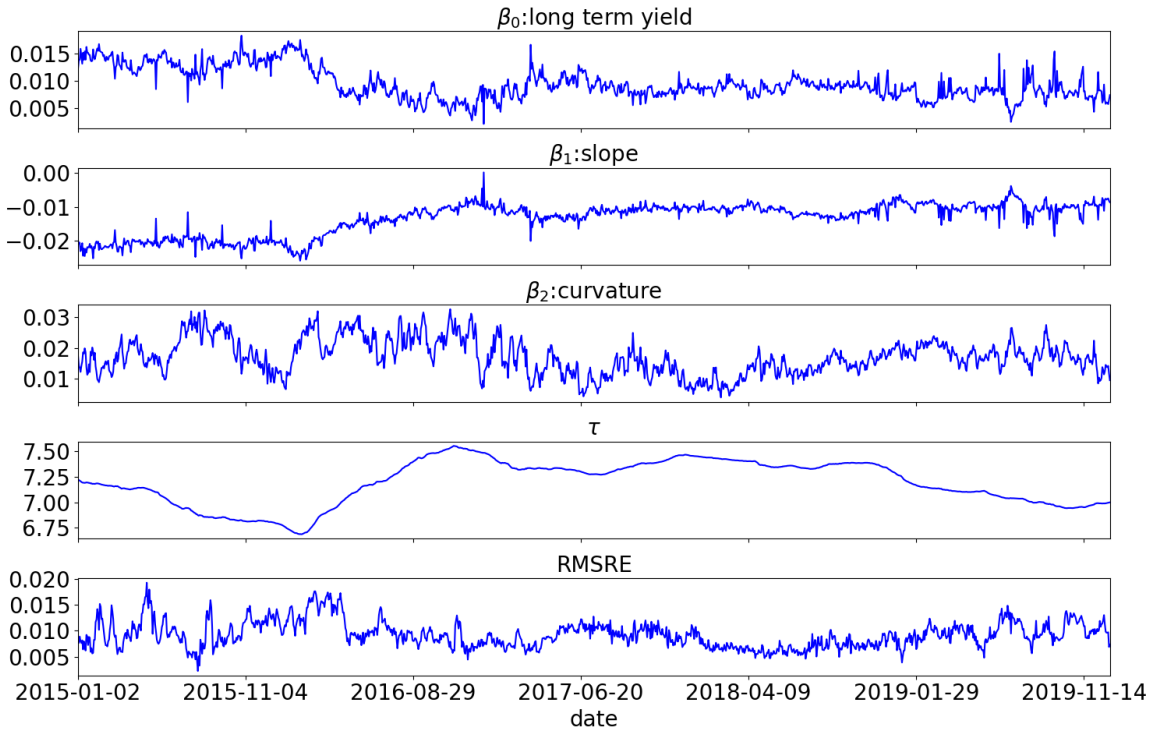


Figure B.4. Time Series of the NS Coefficients for the Credit Spread Curve from 2015 to 2019 for The Walt Disney Company (id=973)

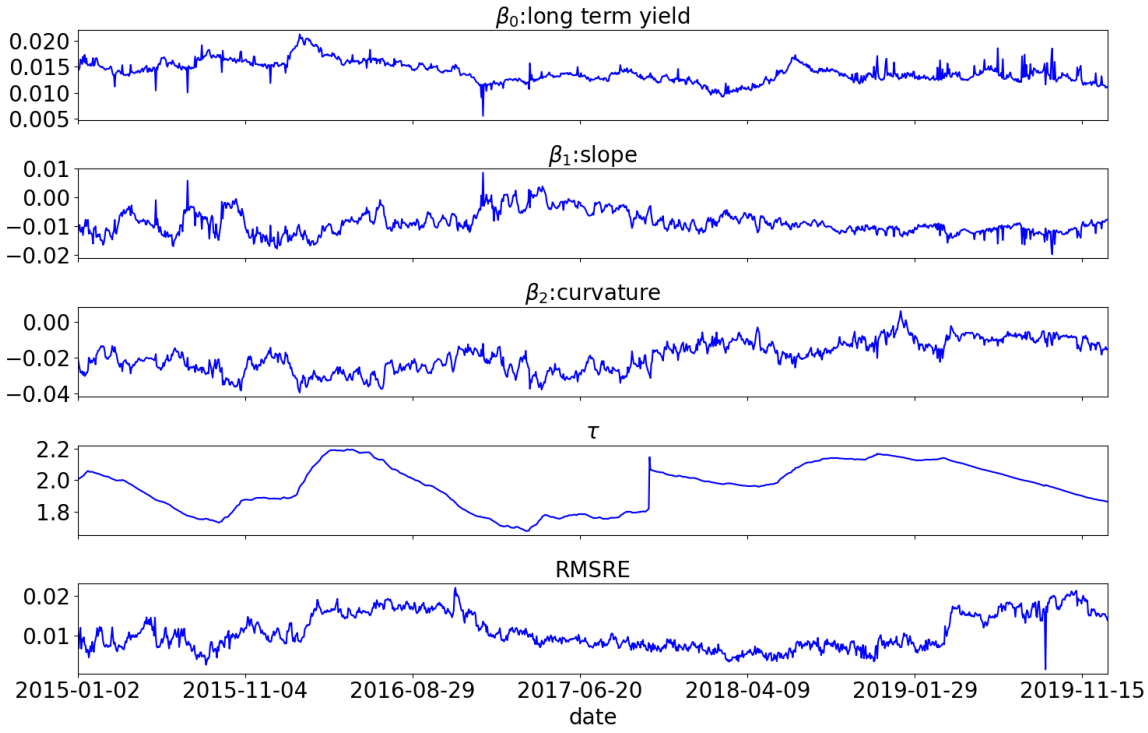


Figure B.5. Time Series of the NS Coefficients for the Credit Spread Curve from 2015 to 2019 for Amgen (id=979)

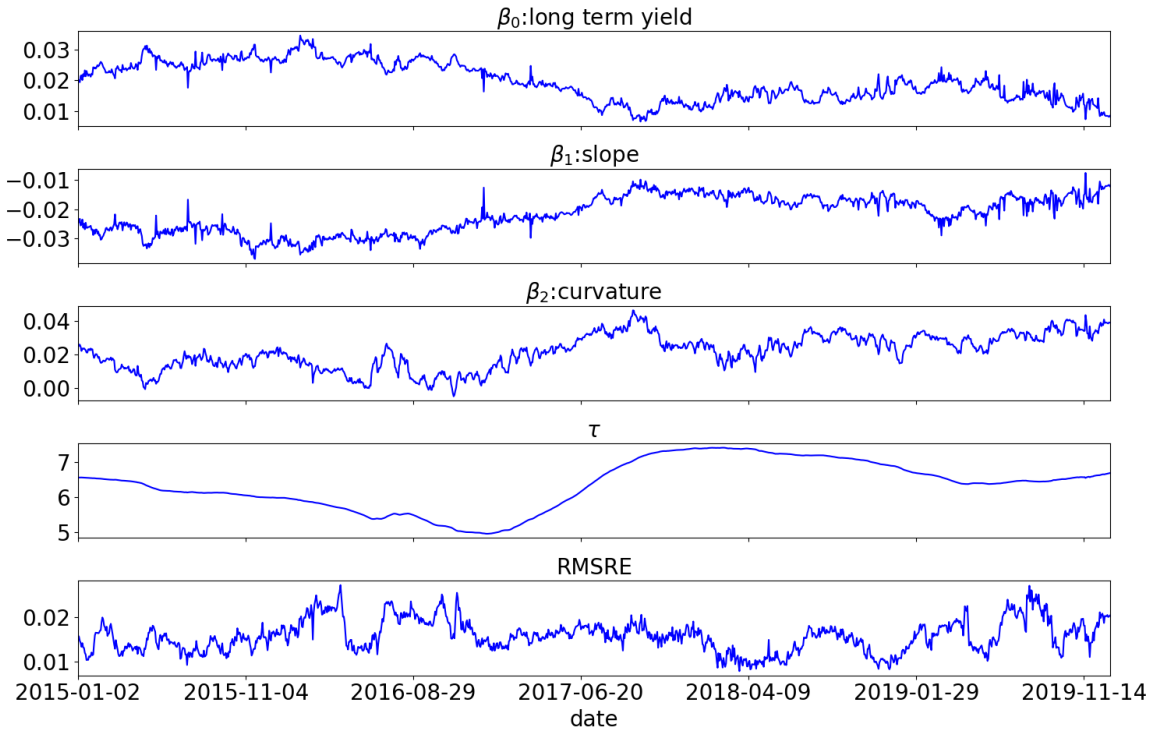
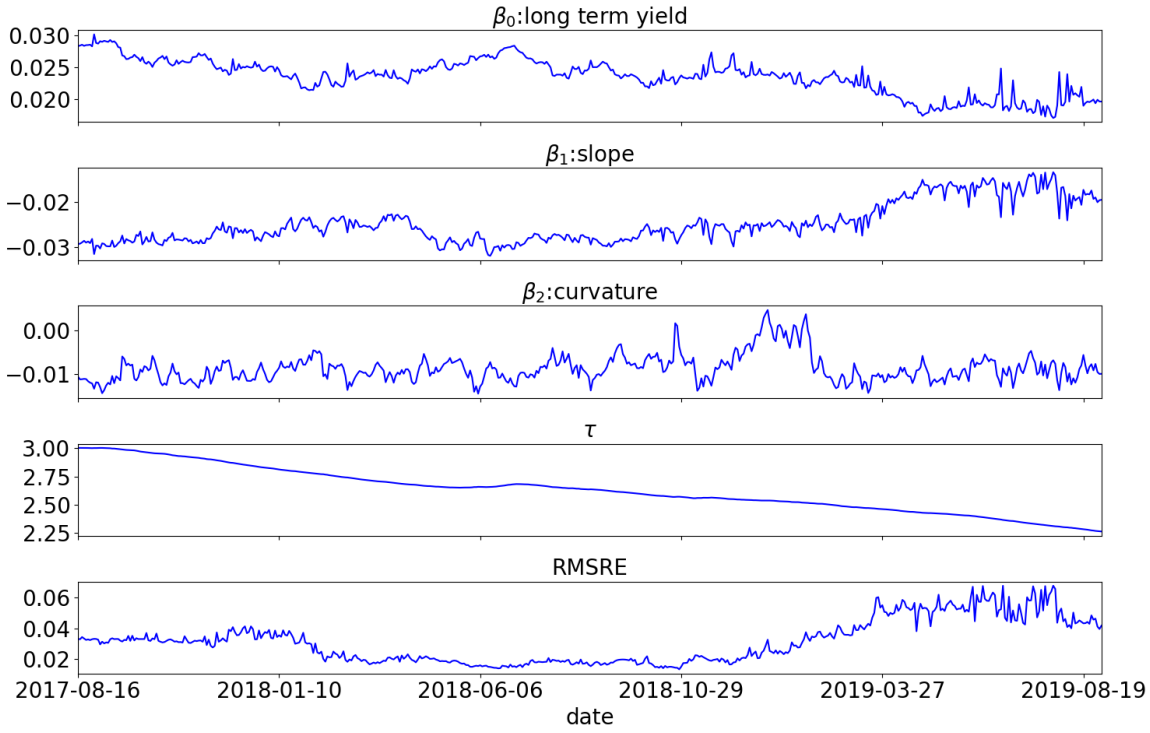


Figure B.6. Time Series of the NS Coefficients for the Credit Spread Curve from 2017 to 2019 for Verizon (id=995)



B.4 Prediction: Seeming Unrelated Regression

We consider a system of three equations:

$$\beta_{0,t} = \gamma_{0,0} + \sum_{i=1}^{AR} \gamma_{0,i}\beta_{0,t-i} + \varepsilon_{0,t}, \tag{B.4}$$

$$\beta_{1,t} = \gamma_{1,0} + \sum_{i=1}^{AR} \gamma_{1,i}\beta_{1,t-i} + \varepsilon_{1,t}, \tag{B.5}$$

$$\beta_{2,t} = \gamma_{2,0} + \sum_{i=1}^{AR} \gamma_{2,i}\beta_{2,t-i} + \varepsilon_{2,t}, \tag{B.6}$$

where $t = 2, ..., T$ is the time period, and the error terms $\varepsilon_{0,t}$, $\varepsilon_{1,t}$, and $\varepsilon_{2,t}$ are correlated.

The Seemingly Unrelated Regressions (SUR) model is estimated using Feasible Generalized Least Squares (FGLS). The first step is to estimate the system of equations using Ordinary Least Squares (OLS).

The system of Equation (B.5)-(B.6) can be rewritten in matrix form as:

$$Y = X\beta + \varepsilon \tag{B.7}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}_{6\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}, \tag{B.8}$$

where $y_i = \begin{bmatrix} \beta_{i,2} \\ \beta_{i,3} \\ \vdots \\ \beta_{i,T} \end{bmatrix}$, $X_i = \begin{bmatrix} 1 & \beta_{i,1} \\ 1 & \beta_{i,2} \\ \vdots & \vdots \\ 1 & \beta_{i,T-1} \end{bmatrix}$, $\beta_i = \begin{bmatrix} \gamma_{i,0} \\ \gamma_{i,1} \end{bmatrix}$, and $\varepsilon = \begin{bmatrix} \varepsilon_{i,2} \\ \varepsilon_{i,3} \\ \vdots \\ \varepsilon_{i,T} \end{bmatrix}$.

Let $n = T - 1$.Then, Y and ε are vectors of size $3n \times 1$; for an autoregression of order one, X is of size $3n \times 6$, and β is of sized 6×1. The OLS estimator is given by $\beta = (X^\top X)^{-1} X^\top Y$. From this step, we compute the residuals and estimate the error covariance matrix Σ, whose elements are defined as: $\sigma_{ij} = \frac{1}{n}\varepsilon_i^\top \varepsilon_j$.

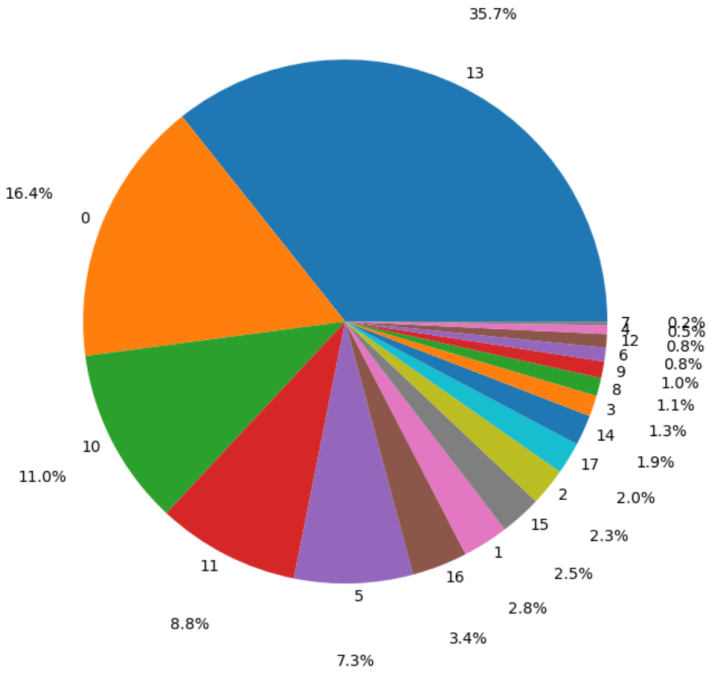
In the second step, we estimate the GLS model using the variance matrix $\hat{\Omega} = \hat{\Sigma} \otimes \mathbb{I}_n$. The FGLS estimator is then: $\beta = (X^\top(\hat{\Sigma}^{-1} \otimes \mathbb{I}_n)X)^{-1} X^\top(\hat{\Sigma}^{-1} \otimes \mathbb{I}_n)Y$.

B.5 ClusteringApproach

B.5.1 Distribution of Clusters Among High-Frequency Issuers

The proportion of high-frequency issuers by cluster is presented in the pie chart below. A total of 18 clusters have been retained based on the BIC criterion. Cluster membership was estimated using the Gaussian Mixture Model (GMM). A disproportionate representation across clusters has been identified: 5 clusters account for approximately 80% of the data, while the remaining clusters each represent less than 3%.

Figure B.7. Proportion of High-Frequency Issuers by Cluster



Clusters were estimated using a Gaussian Mixture Model. The number of clusters was determined using the Bayesian Information Criterion (BIC).

B.5.2 Confusion Matrix

The table below presents the confusion matrix obtained from the high-frequency issuers. The confusion matrix displays the posterior probabilities of cluster membership, conditional on the triplet of sector, seniority, and credit rating.

Table B.1. Confusion Matrix of the High-frequency Issuers Based on GMM Clusters

Sen.	Sec.	Rating	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Other Secured	Energy	Substantial risks	-	-	-	-	100.00	-	-	-	-	-	-	-	-	-	-	-	-	-
		NR	23.91	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	76.09	-
	Financials	NIGS	-	-	-	-	-	-	-	-	-	-	-	-	100.00	-	-	-	-	-
		UMG	-	-	33.62	-	-	50.29	-	-	-	2.30	2.01	-	-	5.46	-	6.32	-	-
	Utilities	LMG	-	-	0.46	-	-	89.50	9.59	-	-	-	-	-	-	-	-	0.46	-	-
		UMG	-	-	-	-	-	100.00	-	-	-	-	-	-	-	-	-	-	-	-
Other Unsecured		LMG	15.31	-	-	-	-	80.61	-	-	-	-	4.08	-	-	-	-	-	-	-
	Financials	NR	-	5.66	2.83	64.15	-	7.08	-	-	-	0.47	-	-	-	-	-	9.91	-	9.91
		UMG	17.92	2.63	-	9.59	-	15.07	-	-	2.51	1.71	4.45	38.58	-	-	-	-	2.97	4.57
	Telecom.	LMG	-	-	-	-	-	-	-	-	-	-	-	-	-	100.00	-	-	-	-
Senior Secured	Energy	LMG	-	-	-	-	-	-	80.99	-	-	-	-	-	-	-	19.01	-	-	-
		High Grade	-	-	-	-	-	-	-	-	-	-	52.07	-	-	47.93	-	-	-	-
	Financials	NR	3.56	-	-	0.32	-	-	-	-	-	-	1.29	3.88	-	90.29	-	-	0.65	-
		NIGS	6.67	-	-	-	-	-	-	-	-	-	-	93.33	-	-	-	-	-	-
		UMG	1.86	-	1.24	1.24	-	31.06	-	-	-	-	7.45	-	-	52.80	-	3.73	0.62	-
	Health Care	Highly Speculative	-	67.57	-	2.70	10.81	-	-	-	5.41	-	-	-	-	-	13.51	-	-	-
		NIGS	45.10	-	-	-	-	-	-	-	-	-	-	39.22	-	-	-	-	15.69	-
	Utilities	Highly Speculative	-	56.67	-	-	-	-	-	-	-	-	-	-	43.33	-	-	-	-	-
Senior Unsecured	Consumer Discr.	High Grade	8.27	-	-	-	-	-	-	-	-	-	-	-	-	91.73	-	-	-	-
		Highly Speculative	-	55.81	-	-	-	-	-	-	39.53	4.65	-	-	-	-	-	-	-	-
		LMG	24.65	0.17	-	-	-	2.32	-	-	-	1.24	-	13.65	-	54.18	-	-	0.08	3.72
		NR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100.00
		NIGS	23.82	8.66	-	4.33	-	-	-	-	18.50	6.50	-	2.17	-	2.36	-	-	33.66	-
		UMG	3.11	-	-	-	-	-	-	-	-	-	1.40	5.28	-	90.06	-	-	0.16	-
	Consumer Staples	High Grade	-	-	-	-	-	-	-	-	-	-	-	-	-	100.00	-	-	-	-
		Highly Speculative	-	6.06	-	-	-	-	-	-	4.55	-	-	-	-	-	-	-	89.39	-
		LMG	20.89	0.29	-	0.29	-	0.33	-	-	0.33	0.58	2.22	12.86	-	56.98	0.54	-	3.21	1.48
		NR	41.55	-	-	-	-	-	-	-	-	-	-	12.08	-	36.71	-	-	9.66	-
		UMG	0.46	-	-	-	-	-	-	-	-	-	0.93	1.55	-	96.90	-	-	0.15	-
	Energy	High Grade	2.71	-	20.62	-	-	3.75	-	-	-	-	13.12	3.12	-	56.67	-	-	-	-
		Highly Speculative	-	45.00	-	-	10.00	-	-	-	45.00	-	-	-	-	-	-	-	-	-
		LMG	42.51	0.31	0.79	0.66	-	9.52	0.60	-	0.16	1.82	0.47	16.65	-	20.42	0.41	0.50	5.12	0.06
		NR	20.23	13.92	3.40	-	-	6.31	-	-	-	6.15	4.37	24.76	-	20.87	-	-	-	-
		NIGS	9.89	47.80	-	-	17.58	-	-	-	10.71	1.65	-	0.27	2.47	-	1.65	-	7.69	0.27
		UMG	17.19	-	-	-	-	3.47	-	-	-	-	18.68	17.02	-	43.64	-	-	-	-
	Financials	High Grade	3.57	0.18	9.17	0.09	-	3.12	-	-	-	0.18	56.55	4.67	-	11.64	0.09	10.72	-	-
		LMG	28.41	3.24	1.90	3.77	-	14.56	0.07	-	1.87	1.44	12.39	5.98	0.23	6.74	5.71	0.80	7.15	5.74
		NR	19.03	1.59	3.06	0.82	-	6.23	0.21	-	0.27	2.29	13.78	10.27	1.99	23.31	3.18	7.24	1.92	4.80
		NIGS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100.00
		Prime	3.62	-	0.99	-	-	14.47	-	-	-	0.33	78.62	1.64	-	0.33	-	-	-	-
		UMG	10.17	1.28	6.66	0.72	-	13.81	0.35	-	0.17	0.89	27.70	7.22	1.59	16.00	2.44	10.34	0.39	0.27

The clusters are predicted using GMM on the high-frequecy issuer. LMG stands for Lower Medium Grade. UMG stands for Upper Medium Grade. NIGS stands for Non Investment Speculative. NR stands for Non Rated.

Table B.2. Confusion Matrix of the High-frequency Issuers Based on GMM Clusters

Sen.	Sec.	Rating	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Senior Unsecured	Government	High Grade	14.05	-	-	-	-	-	-	-	-	-	-	-	-	85.95	-	-	-	-
		NR	2.48	-	-	-	-	-	-	-	-	-	-	-	-	97.52	-	-	-	-
		Prime	0.21	-	-	-	-	-	-	-	-	-	99.79	-	-	-	-	-	-	-
		UMG	-	-	1.49	-	-	-	-	-	-	-	55.82	-	-	42.69	-	-	-	-
	Health Care	High Grade	0.84	-	-	-	-	-	-	-	-	-	-	-	-	99.16	-	-	-	-
		Highly Speculative	-	8.00	-	-	30.00	-	-	-	62.00	-	-	-	-	-	-	-	-	-
		LMG	25.74	-	-	1.12	-	1.12	-	-	0.06	0.24	0.35	11.31	-	58.36	-	-	1.24	0.47
		NR	1.66	14.08	-	-	-	-	-	-	1.24	-	5.18	-	12.63	52.80	4.14	8.28	-	-
		NIGS	2.00	-	-	-	-	-	-	-	-	2.00	-	-	-	-	-	-	96.00	-
		Prime	1.65	-	-	-	-	-	-	-	-	-	-	-	-	98.35	-	-	-	-
		UMG	8.10	-	-	-	-	-	-	-	-	-	-	3.49	-	88.40	-	-	-	-
	Industrials	High Grade	4.28	-	-	-	-	2.33	-	-	-	-	6.61	9.34	-	77.43	-	-	-	-
		Highly Speculative	-	28.68	-	-	41.18	-	-	-	23.53	-	-	-	-	-	-	-	6.62	-
		LMG	16.76	0.50	3.97	0.56	-	11.67	0.93	-	-	0.06	9.19	10.24	-	36.06	2.79	-	0.43	6.83
		NR	-	-	-	-	-	-	-	-	-	-	-	19.62	-	22.49	29.19	28.71	-	-
		NIGS	-	-	-	-	-	-	-	-	3.09	-	-	-	1.03	-	95.88	-	-	-
		UMG	6.35	-	-	-	-	2.06	-	-	-	0.09	10.90	1.55	-	79.06	-	-	-	-
	Information Tech.	High Grade	2.86	-	-	-	-	-	-	-	-	-	15.10	-	-	82.04	-	-	-	-
		Highly Speculative	-	19.61	-	-	7.84	-	-	-	-	-	-	-	-	-	72.55	-	-	-
		LMG	23.18	-	-	2.42	-	-	-	-	-	0.35	-	4.15	-	64.01	-	-	5.88	-
		NR	45.69	-	-	12.93	-	-	-	-	-	-	-	12.93	-	7.76	-	-	20.69	-
		NIGS	0.86	30.17	-	-	-	-	-	-	-	12.93	-	-	-	-	-	-	56.03	-
		Prime	-	-	-	-	-	-	-	-	-	-	-	-	-	100.00	-	-	-	-
		UMG	8.90	-	-	-	-	-	-	-	-	-	6.79	2.56	-	81.75	-	-	-	-
	Materials	Highly Speculative	-	23.86	-	-	76.14	-	-	-	-	-	-	-	-	-	-	-	-	-
		LMG	31.44	2.65	3.13	0.08	-	4.73	0.16	-	-	2.09	3.69	10.83	-	24.62	0.80	-	11.23	4.57
		NR	3.12	7.50	-	-	-	-	7.19	-	2.19	-	-	2.50	28.75	9.38	1.88	-	37.50	-
		NIGS	19.14	33.33	-	-	-	-	-	31.02	5.94	3.30	-	-	-	-	-	-	6.93	0.33
		UMG	-	-	-	-	-	-	-	-	-	-	-	-	-	100.00	-	-	-	-
	Real Estate	LMG	16.32	-	-	1.28	-	-	-	-	-	-	0.58	17.72	-	63.05	-	-	1.05	-
		NR	15.38	-	-	-	-	10.26	-	-	-	-	2.56	9.40	-	15.38	-	-	-	47.01
		NIGS	-	-	-	-	100.00	-	-	-	-	-	-	-	-	-	-	-	-	-
		UMG	11.26	-	5.98	-	-	7.59	-	-	-	0.46	20.00	0.23	-	54.48	-	-	-	-
	Telecom.	Highly Speculative	-	95.35	-	-	1.16	-	-	-	3.49	-	-	-	-	-	-	-	-	-
		LMG	14.77	-	-	-	-	-	-	-	-	0.22	-	34.45	-	50.56	-	-	-	-
		NR	17.65	31.86	-	4.90	-	5.88	-	-	5.39	0.49	-	1.96	-	-	9.80	-	21.57	0.49
		NIGS	13.60	36.00	-	-	-	-	-	-	48.80	-	-	-	-	-	-	-	1.60	-
	Utilities	LMG	23.47	-	-	0.33	-	0.17	6.61	-	-	-	0.83	12.56	-	42.64	9.92	3.31	0.17	-
		NR	3.12	-	-	-	-	-	-	-	-	-	-	4.17	-	90.62	-	-	2.08	-
		NIGS	27.03	2.70	-	34.23	-	-	-	-	6.31	1.80	-	9.91	-	-	-	-	18.02	-
		UMG	10.42	-	0.09	-	-	26.98	10.95	-	-	1.60	2.67	5.16	-	39.63	-	0.71	0.09	1.69

The clusters are predicted using GMM on the high-frequency issuer. LMG stands for Lower Medium Grade. UMG stands for Upper Medium Grade. NIGS stands for Non Investment Speculative. NR stands for Non Rated.

References

Gauthier, G., & Simonato, J.-G. (2012). Linearized nelson–siegel and svensson models for the estimation of spot interest rates. *European Journal of Operational Research*, *219*(2), 442–451.

Goldberger, A. S. (1964). *Econometric theory.* New York: John Wiley & Sons.

Chapter C

Appendices of *Enhancing Credit Spread Forecasts through Macroeconomic Uncertainty Variables and Statistical Learning Approaches*

C.1 Target Coefficients' Procedure

C.1.1 Governmental Nelson-Siegel Coefficients Target ($\beta_{0,t}^{*gov}$, $\beta_{1,t}^{*gov}$, and $\beta_{2,t}^{*gov}$)

The target coefficients are defined with respect to Gürkaynak et al. (2007)'s U.S. Treasury zero-coupon yield curve features. From Gürkaynak et al. (2007), we obtain the Svensson parameters and the estimated U.S. Treasury zero-coupon yield curves for each day in the sample. Based on these estimated curves, we derive $\beta_{l,t}^{*gov}$, $l = 0, 1, 2$, as follows:

- $\beta_{0,t}^{*gov}$ represents the average long-term yield for maturities beyond 25 years, i.e.,

$$\beta_{0,t}^{*gov} = \frac{1}{5} \sum_{u=26}^{30} y(\theta_t^{gov}, u).$$

- $\beta_{1,t}^{*gov}$ is derived from the slope of the U.S. Treasury zero-coupon yield curve, calculated as the difference between the one-year yield and $\beta_{0,t}^{*gov}$:

$$\beta_{1,t}^{*gov} = y(\theta_t^{gov}, 1) - \beta_{0,t}^{*gov}.$$

- $\beta_{2,t}^{*gov}$ captures the curvature by evaluating twice the two-year yield minus the sum of the one-year yield and the average long-term yield:

$$\beta_{2,t}^{*gov} = 2y(\theta_t^{gov}, 2) - \beta_{0,t}^{*gov} - y(\theta_t^{gov}, 1).$$

C.1.2 Firms-Specific Spread Nelson-Siegel Coefficients Target ($\beta_{0,i,t}^{*spread}$, $\beta_{1,i,t}^{*spread}$, and $\beta_{2,i,t}^{*spread}$)

Defining $\beta_{l,i,t}^{*spread}$ is more challenging than defining $\beta_{l,t}^{*gov}$ because corporate bond maturities are not uniformly distributed. To determine these target coefficients $\beta_{l,i,t}^{*spread}$ (for $l = 0, 1, 2$) for each firm and each day, we employ the following algorithm:

- If there is more than one corporate bond with a maturity exceeding 15 years ($n_{i,t}^{15^+y} > 1$), then $\beta_{0,i,t}^{*spread}$ is set to the average yield-to-maturity (YTM) of those bonds. Otherwise, it equals the YTM of the bond with the longest maturity. More precisely,

$$\beta_{0,i,t}^{*spread} = \begin{cases} \frac{1}{n_{i,t}^{15^+y}} \sum_{j=1}^{n_{i,t}^{15^+y}} YTM_{u_j}, & \text{if } n_{i,t}^{15^+y} > 1, \\ YTM_{u_{last}}, & \text{otherwise.} \end{cases}$$

  where $n_{i,t}^{15^+y}$ denotes the number of bonds with maturities longer than 15 years, u_j is the time-to-maturity of the j^{th} coupon payment, and u_{last} is the time-to-maturity of the final coupon payment.

- If there is more than one corporate bond with a maturity under 5 years ($n_{i,t}^{5^-y}$) and more than one corporate bond with a maturity exceeding 15 years, then, $\beta_{1,i,t}^{*spread}$ is defined as the difference between the average yield-to-maturity of the short-term bonds and $\beta_{0,i,t}^{*spread}$. This represents 18.06% of the credit spread curves. Otherwise, $\beta_{1,i,t}^{*spread}$ equals the difference between the yield-to-maturity of the shortest-maturity bond and that of the longest-maturity bond. More precisely,

$$\beta_{1,i,t}^{*spread} = \begin{cases} \frac{1}{n_{i,t}^{5^-y}} \sum_{j=1}^{n_{i,t}^{5^-y}} YTM_{u_j} - \beta_{0,i,t}^{*spread}, & \text{if } n_{i,t}^{5^-y} > 1 \text{ and } n_{i,t}^{15^+y} > 1. \\ YTM_{u_1} - YTM_{u_{last}}, & \text{otherwise.} \end{cases}$$

$n_{i,t}^{5^-y}$ denotes the number of obligation maturing in less than 5 years, and u_1 is the time-to-maturity for the first coupon payment.

- If there is at least one corporate bond with a maturity between 5 and 15 years ($n_{i,t}^{[5-15]y}$), then, $\beta_{2,i,t}^{*spread}$ is defined as twice the average yield-to-maturity of those bonds minus $\beta_{0,i,t}^{*spread}$ and $\beta_{1,i,t}^{*spread}$. If no such bond exists, then $\beta_{2,i,t}^{*spread}$ is calculated as twice the yield-to-maturity of the bond closest to the 25% percentile of the firm's bond distribution at time t (i.e., $YTM_{\lceil 0.25 n_{i,t} \rceil}$) minus the yield-to-maturity of the bond with the longest maturity minus the yield-to-maturity of the bond with the shortest maturity. More precisely,

$$\beta_{2,i,t}^{*spread} = \begin{cases} \frac{2}{n_{i,t}^{[5-15]y}} \sum_{j=1}^{n_{i,t}^{[5-15]y}} YTM_{u_j} - \beta_{0,i,t}^{*spread} - \beta_{1,i,t}^{*spread}, & \text{if } n_{i,t}^{[5-15]y} \geq 1. \\ YTM_{\lceil 0.25 n_{i,t} \rceil} - \beta_{0,i,t}^{*spread} - \beta_{1,i,t}^{*spread}, & \text{otherwise.} \end{cases}$$

The yield-to-maturities(YTM) are available from the dataset. If the bond is a zero-coupon, the YTM is calculated as

$$YTM_{i,t} = -\frac{\log\left(\frac{P(t,T_i)}{100}\right)}{T_i},$$

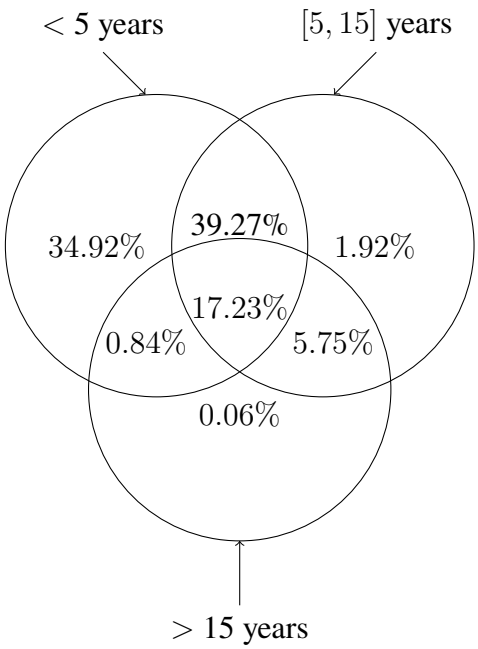
where $P\left(t, T_i\right)$ is the price at time t for company i of the zero-coupon bond maturing at T_i; Otherwise, The YTM is computed as the value y that minimizes the pricing error:

$$YTM_{i,t} = \underset{y}{\operatorname{argmin}} \sqrt{\left(\sum_{j=1}^{m_i} c_{i,j} e^{-y(u_{i,j})u_{i,j}} - B_{i,t}^{obs}\right)^2},$$

where m_i is the number of remaining coupon payment for bond i; $u_{i,j}$ is the time-to-maturity of the j^{th} coupon payment for bond i; and $y(u_{i,j})$ is the zero-coupon yield corresponding at each point $u_{i,j}$. $c_{i,j}$ is the j^{th} coupon payment for bond i.

The distribution of bonds maturities across the curve is shown in Figure C.1.

Figure C.1. Percentage of Credit Spread Curves by Bond Maturity



$52,564$ curves constitutes the sample of credit spreads from July1, 2002, to June 30, 2020.

C.2 Rank Analysis of Macroeconomic and Uncertainty Variables

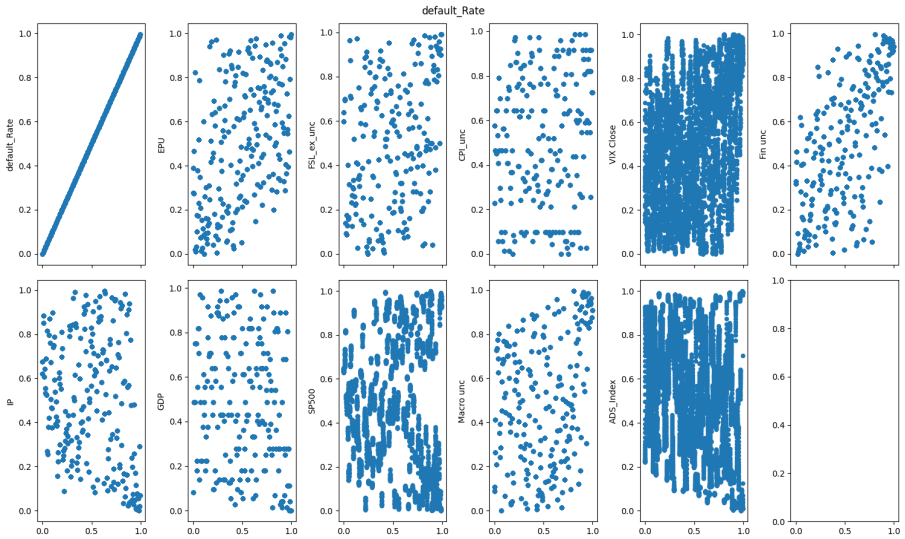
A normalized rank analysis was performed to explore relationships and dependencies among features by converting raw observations into their respective ranks. This approach facilitates comparisons across variables on different scales and can uncover nonlinear associations that might otherwise be obscured. For each element in feature s, the algorithm counts how many values in feature s' are strictly less than that element, then divides by $(n\_obs + 1)$ to normalize the rank.

Figure C.2 illustrates rank plots for all features, with each panel corresponding to a specific variable. A clear linear relationship emerges between the S&P 500 and Industrial Production (IP), while IP and the financial-uncertainty indicator exhibit a nonlinear association. Beyond these patterns, no other strong dependencies are observed.

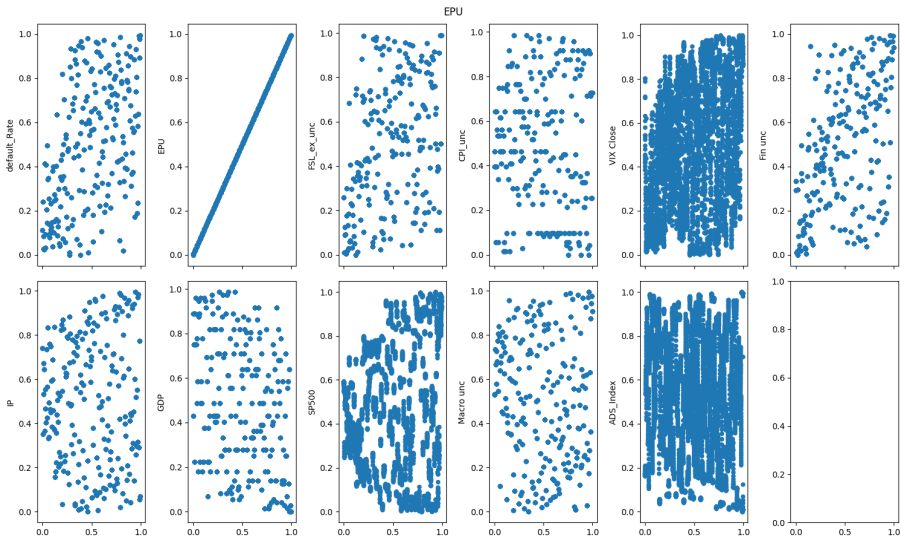
Combining the rank analysis with the correlation results in Table 4.3 motivated the exclusion of the S&P 500 and financial-uncertainty measures, given their high overlap with other explanatory variables. Consequently, no substantial linear or nonlinear correlation remains among the retained features.

Figure C.2. Rank Analysis of the Macroeconomic and Uncertainty Variables

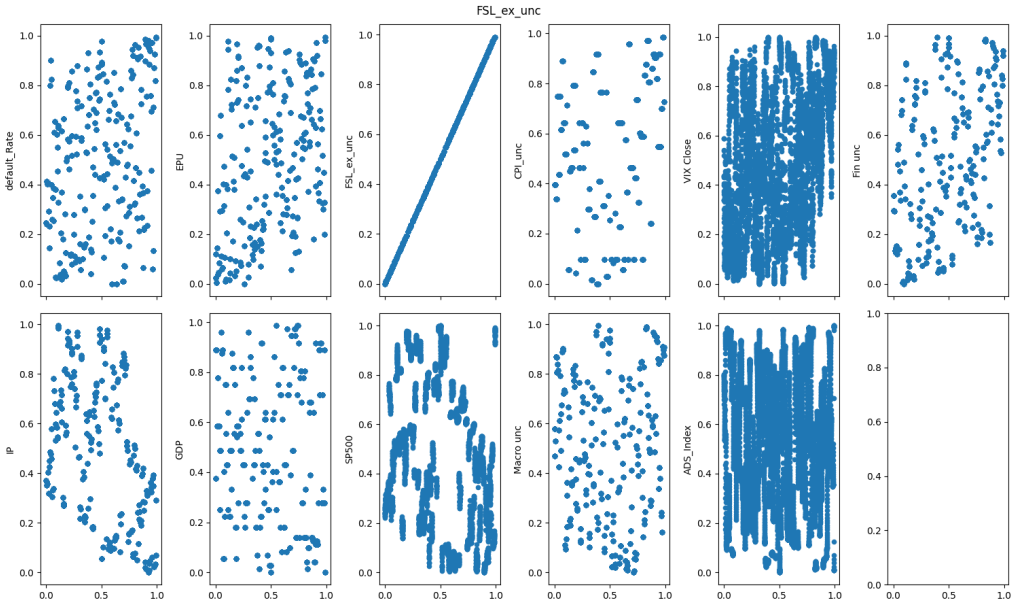
Default Rate



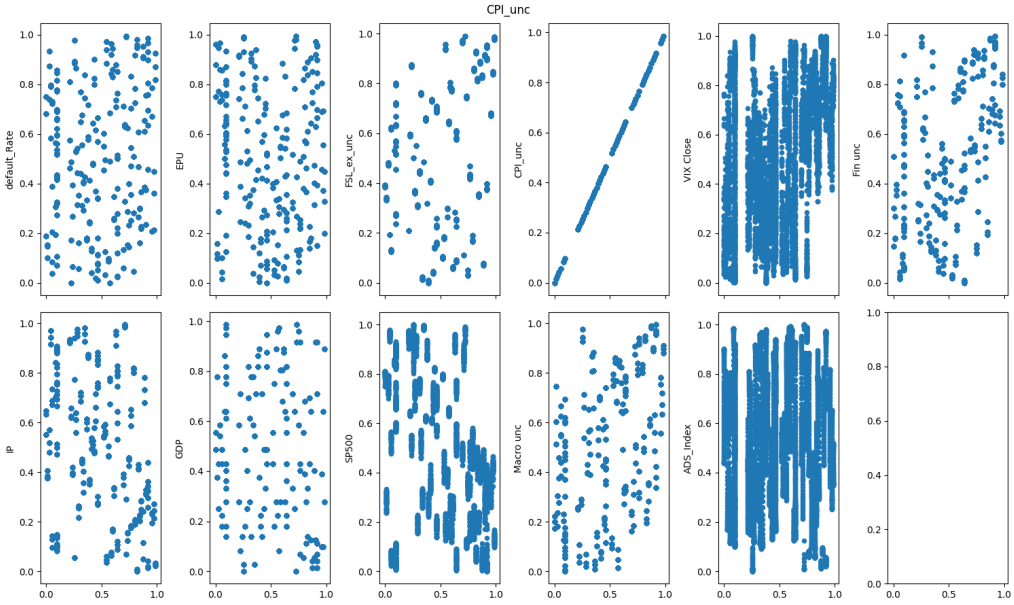
EPU



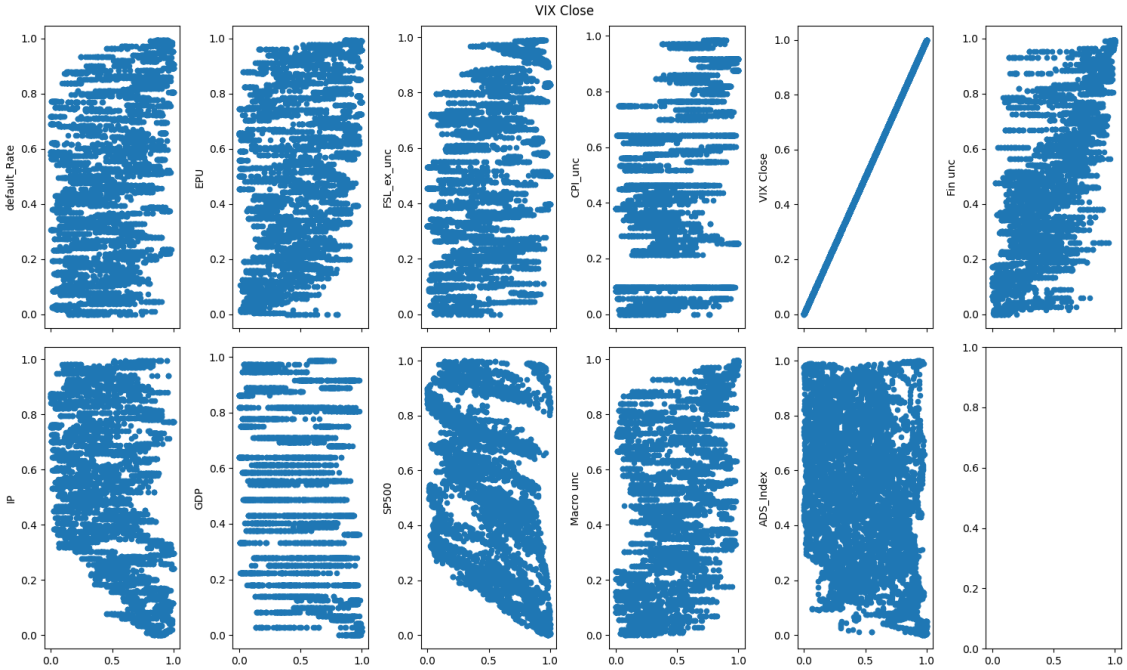
FSL Uncertainty



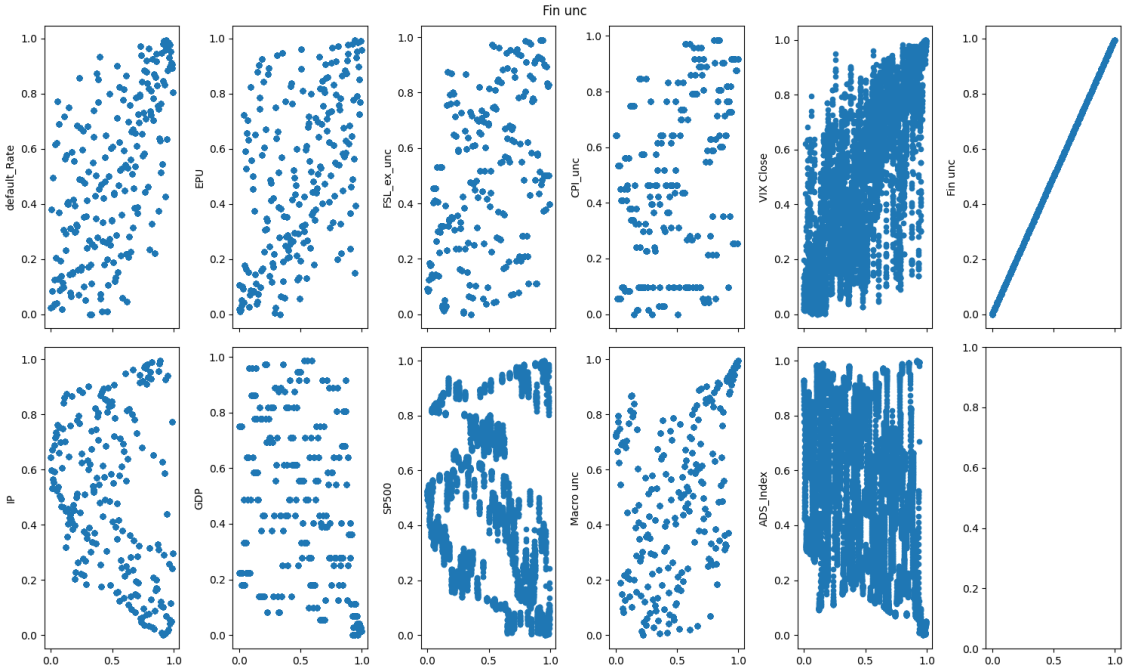
CPI Uncertainty



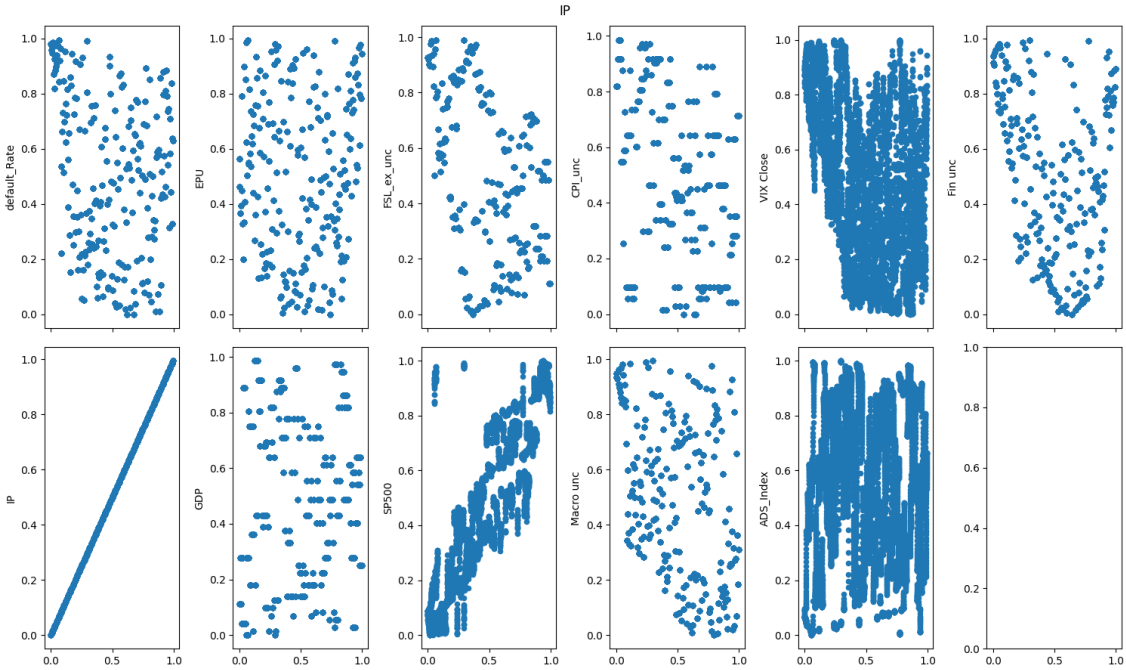
VIX Close



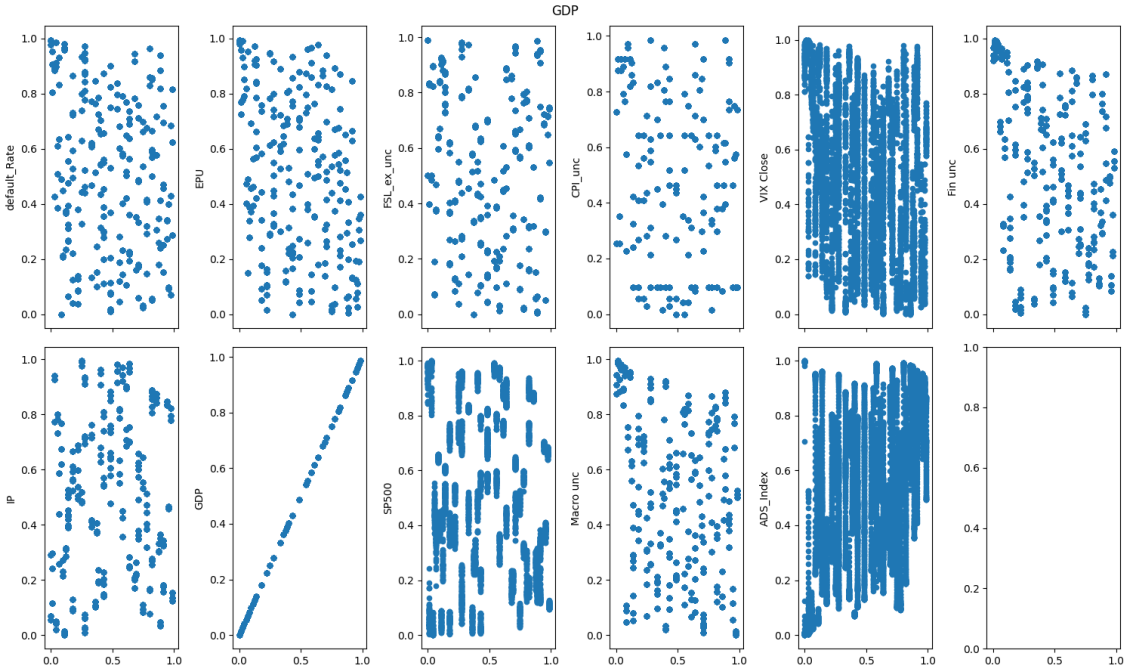
Financial Uncertainty



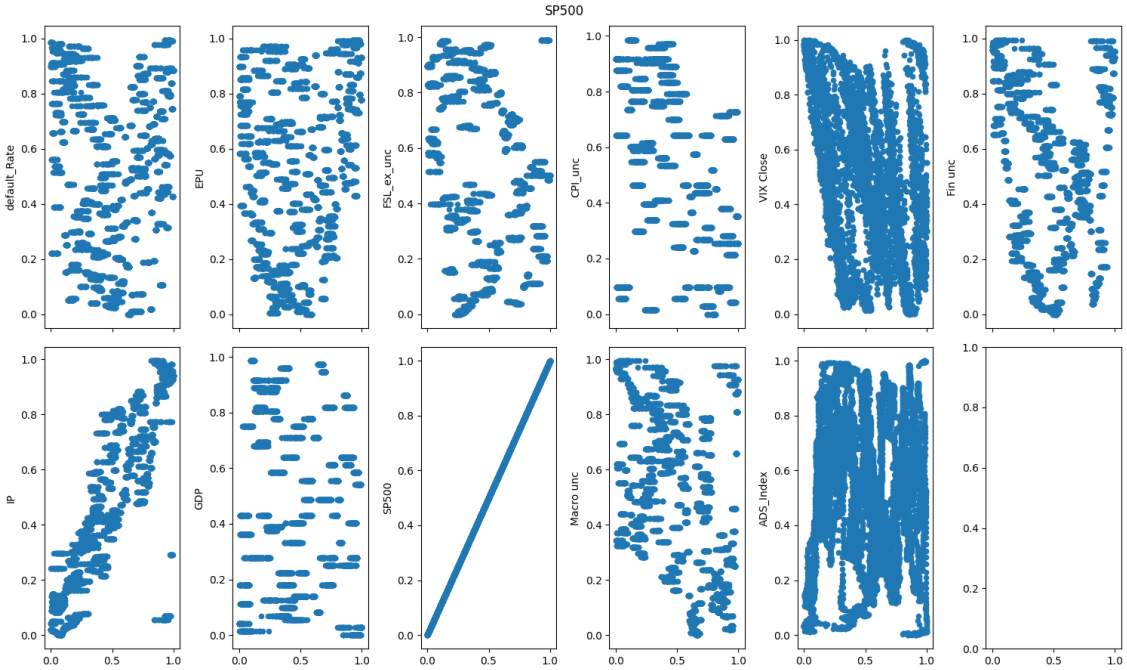
Industrial Production



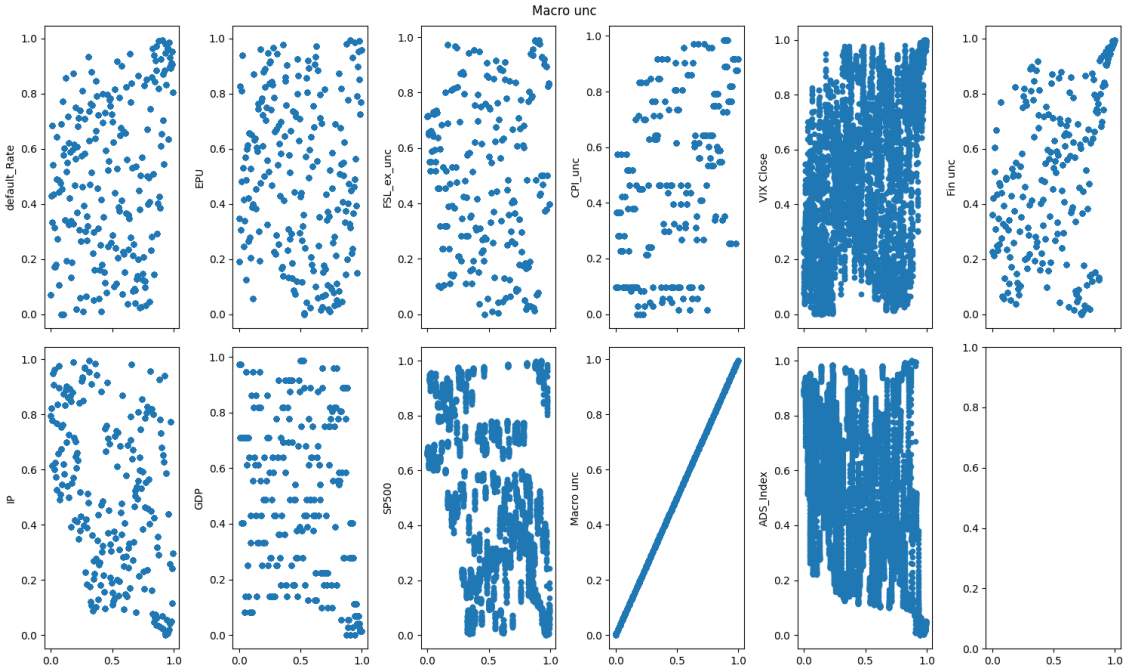
GDP



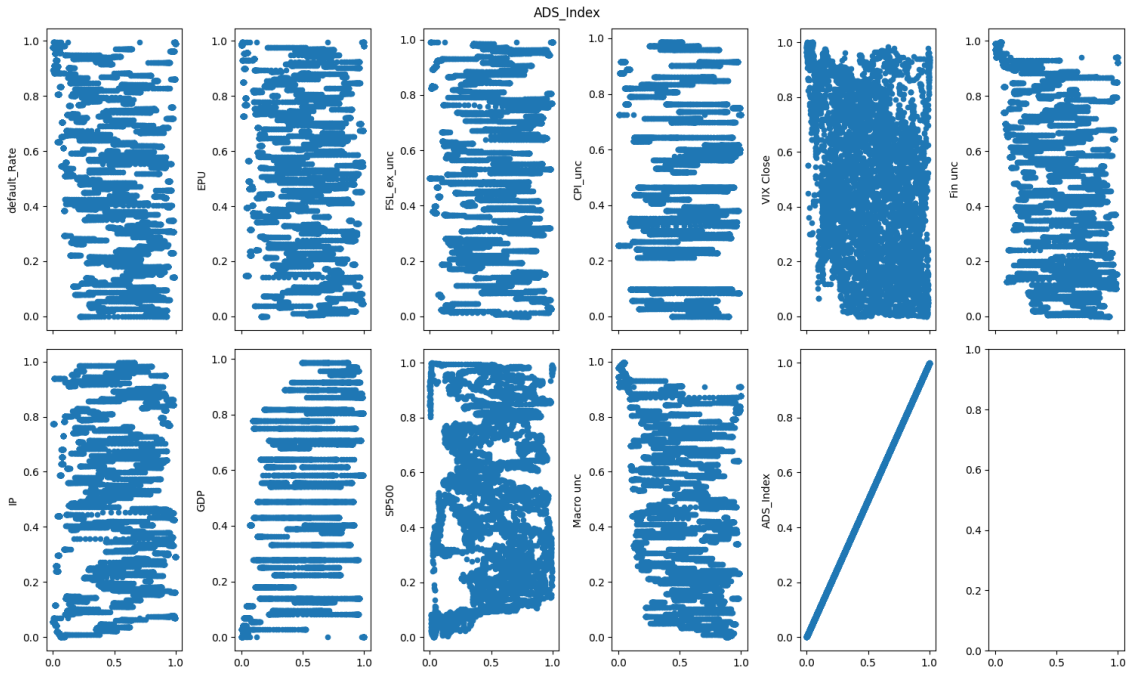
S&P500



Macro Uncertainty



ADS Index



EPU stands for the Economic Policy Uncertainty measure of Baker et al. (2016). FSL unc. stands for the uncertainty relative to both federal and state/local purchases of Baker et al. (2016). CPI unc. stands for the inflation uncertainty measure of Baker et al. (2016). The VIX index measures stock market volatility. The ADS index reflects the economic state condition and is defined by Aruoba et al. (2009). Fin unc. refers to the financial uncertainty measure, and macro unc. denotes the macroeconomic uncertainty measure; both are derived from the work of Jurado et al. (2015) and Ludvigson et al. (2021). IP stands for U.S. Industrial Production, and GDP for U.S. Gross Domestic Product.

C.3 Hyper-parameters of Ensemble Models

Model hyper-parameters are tuned using 10-fold time-series split cross-validation. The training set is divided into 10 time-series blocks. In the first iteration, the first block serves as the in-sample set and the second block as the validation set. In the second iteration, the first two blocks are in-sample and the third block is validation, and so on. We leave a one-month purging gap between each sub-training and sub-validation set to prevent data leakage and ensure an accurate performance measure. The Root Mean Squared Error (RMSE) of the credit spread curves, defined in Equation (4.13), is used as the cost function.

Table C.1 presents the hyper-parameters values of the linear models.

Table C.1. Values of the Hyper-parameters for the Linear Models

	Ridge		Lasso		Elastic Net	
	$h = 1$	$h = 5$	$h = 1$	$h = 5$	$h = 1$	$h = 5$
λ	0.00415	0.00115	1e-06	1e-06	9e-06	7.5e-07
γ	0	0	1	1	0.6	0.1

The hyper-parameters are tuned using 10 time-series blocks. The Root Mean Squared Error (RMSE) of the credit-spread curves (Equation (4.13)) is used as the cost function. h denotes the forecast horizon.

Table C.2 shows the hyper-parameter values for the ensemble models. For the bagging regressor, we use the same algorithm as the random forest but force it to consider all features. Its hyper-parameters are the number of trees (`n_estimators`), the maximum depth of each tree (`max_depth`), the minimum number of observations required to split an internal node (`min_samples_split`), and the minimum number of observations required in a leaf node (`min_samples_leaf`). The hyper-parameters of the random forest are the same as the bagging regression hyper-parameters, plus the number of randomly sampled features (`max_features`). The hyper-parameters of the boosted trees are the same as the random forest hyper-parameters, plus a learning rate that shrinks the contribution of each tree (`learning_rate`).

Table C.2. Values of the Hyper-parameters for the Ensemble Models

	Bagging Trees		Random Forest		Boosted Trees	
	$h = 1$	$h = 5$	$h = 1$	$h = 5$	$h = 1$	$h = 5$
`n_estimator`	50	50	300	100	600	600
`max_depth`	10	5	5	5	5	5
`min_samples_split`	2	4	4	4	4	4
`min_sample_leaf`	2	2	2	1	1	1
`max_features`	None	None	35	25	20	20
`learning_rate`					0.01	0.01

The hyper-parameters are tuned using 10 time-series blocks. The Root Mean Squared Error (RMSE) of the credit spread curves (Equation (4.13)) is used as the cost function. h denotes the forecast horizon.

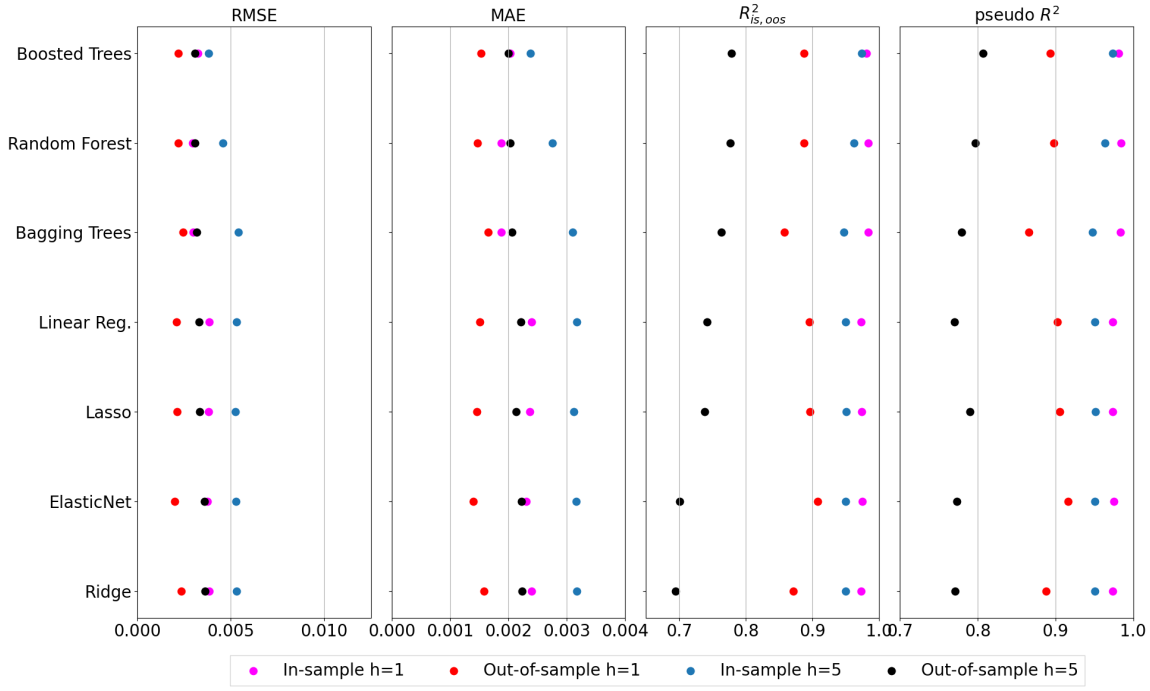
C.4 Model Results

Figure C.3 visually documents the in-sample and out-of-sample performance of the seven credit-spread models' predictions without Ford, evaluated using four key metrics: RMSE, MAE, RMSE, MAE, R^2, and R^2_p. Each model is assessed at one-day ($h = 1$) and five-day ($h = 5$) forecast horizons. The pink (blue) dots represent one-day-ahead (five-day-ahead) in-sample forecasts, while the red (black) dots represent one-day-ahead (five-day-ahead) out-of-sample forecasts.

The out-of-sample results remain unchanged, as Ford's ticker was updated in our database after March 2010. Moreover, out-of-sample performance metrics continue to exceed in-sample ones because Ford was not the only issuer displaying extremely large credit spreads during the subprime crisis. However, Ford did register the highest number of days with abnormally high spreads, so excluding it substantially narrows the gap between in-sample and out-of-sample results.

Table C.3 reports descriptive statistics of the in-sample MSE both with and without Ford's credit-spread curves. Across all models, the MSE is significantly lower once Ford is removed: linear models decrease by almost 70%, while bagging and random forests

Figure C.3. Comparison of Credit Spread Model Performances Without Ford Company Across Metrics (In Sample and Out of Sample) - One-Day Ahead ($h = 1$) and Five-Day Ahead ($h = 5$) Comparison of Credit Spread Model Performances Without Ford Across Metrics (In-Sample and Out-of-Sample) — One-Day Ahead ($h = 1$) and Five-Day Ahead ($h = 5$)



The pink (blue) dots represent one-day-ahead (five-day-ahead) in-sample forecasts, while the red (black) dots represent one-day-ahead (five-day-ahead) out-of-sample forecasts. The first 75% of observations (July 1, 2002–January 7, 2016) constitute the training set, and the remaining 25% (January 8, 2016–June 30, 2020) constitute the test set, with a one-month purging period between them. Ford's credit spread curves have been removed. Performance metrics are defined in Equations (4.13), (4.14), (4.15), and (4.16).

see around a 50% reduction. Even boosted trees, which appear less sensitive, display an approximate 11% improvement. The maximum MSE without Ford is drastically reduced, highlighting the influence of Ford's extreme observations on overall results. These findings underscore how the large increase in credit spreads during the subprime crisis disproportionately affects predictive performance.

Finally, Figure C.4 shows that the RMSE spikes during 2008's crisis have reduced

Table C.3. Statistical Descriptive of the In-Sample Credit Spread Curves Squared Errors

(a) With Ford

	Linear Reg.	Ridge	Lasso	Elastic Net	Bagging Trees	Random Forest	Boosted Trees
count				32946			
mean	0.000931	0.000930	0.000930	0.000945	0.000388	0.000327	0.000241
std	0.036407	0.036381	0.036717	0.038512	0.018106	0.013809	0.000910
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000020	0.000020	0.000019	0.000019	0.000012	0.000012	0.000014
50%	0.000064	0.000064	0.000062	0.000060	0.000041	0.000040	0.000047
75%	0.000205	0.000205	0.000199	0.000187	0.000140	0.000136	0.000156
max	3.816696	3.809219	3.859075	4.166843	3.115655	2.401139	0.042896

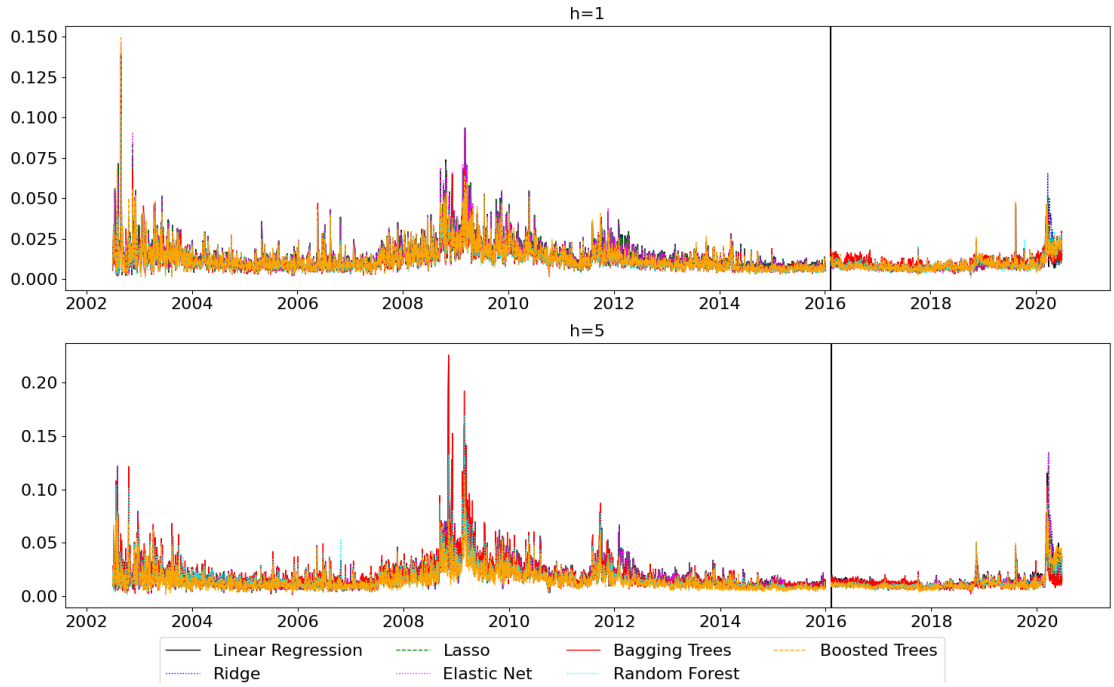
(b) Without Ford

	Linear Reg.	Ridge	Lasso	Elastic Net	Bagging Trees	Random Forest	Boosted Trees
count				31284			
mean	0.000299	0.000298	0.000294	0.000283	0.000181	0.000179	0.000213
std	0.001156	0.001158	0.001155	0.001164	0.000613	0.000589	0.000761
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000019	0.000019	0.000018	0.000018	0.000011	0.000011	0.000013
50%	0.000061	0.000061	0.000059	0.000057	0.000038	0.000038	0.000044
75%	0.000191	0.000190	0.000185	0.000172	0.000129	0.000127	0.000145
max	0.061420	0.061663	0.062494	0.065757	0.027606	0.020018	0.026823

These tables present the statistical descriptive of the squared errors between the in-sample predicted and and observed credit spread curves, with and without Ford. The results cover the daily sample period from July 1, 2002, to June 30, 2020.

without Ford company.

Figure C.4. Time Series of the RMSE Between Predicted and Observed Credit Spreads - One-Day Ahead ($h = 1$) and Five-Day Ahead ($h = 5$)



The first 75% of observations (July 1, 2002–January 7, 2016) constitute the training set; the remaining 25 % (January 8, 2016–June 30, 2020) constitute the test set, with a one-month purging period between them. Precision metrics are defined in Equations (4.13), (4.14), (4.15), and (4.16).

References

Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, *131*(4), 1593–1636.

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, *54*(8), 2291–2304.

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216.

Ludvigson, S. C., Ma, S., & Ng, S. (2021). Uncertainty and business cycles: Exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, *13*(4), 369–410.