Data Science-Based Analysis of Special Situations in Corporate Bonds

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)
von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie
genehmigte

DISSERTATION

von

M. Sc. Friedrich Baumann

Tag der mündlichen Prüfung: 7. Februar 2024
Referent: Prof. Dr. Gholamreza Nakhaeizadeh
Korreferentin: Prof. Dr. Melanie Schienle

Acknowledgments

I would like to express my sincere gratitude to my supervisors Prof. Gholamreza Nakhaeizadeh *(Karlsruhe Institute of Technology)* and Prof. Melanie Schienle *(Karlsruhe Institute of Technology)* for accepting the role of supervision and for their meticulous review of my work. I specially thank Prof. Frank J. Fabozzi *(Johns Hopkins University)* for the valuable feedback and comments on the two papers on which Chapters 3 and 5 are based on. I would like to thank Prof. Norman Schürhoff *(University of Lausanne)* and Prof. Dmitry Livdan *(University of California, Berkeley)* for the fruitful discussions and comments regarding the paper upon which Chapter 2 is based on. Furthermore, I would like to express my gratitude to Prof. Kakhbod *(University of California, Berkeley)* for his guidance and comments regarding the paper upon which Chapter 2 is based, and for the many inspiring discussions we had. I would like to especially thank Dr. Abdolreza Nazemi *(Karlsruhe Institute of Technology)* for the extensive discussions, comments, support and guidance provided throughout the process of writing this dissertation. Moreover, I would like to thank him for his patience along the way. Finally, I would like to extend my gratitude to Prof. Philipp Schuster *(University of Stuttgart)* for providing data used in Chapters 2 and 4.

Contents

List of Figures

List of Tables

List of Abbreviations

Adj.	Adjusted
AIC	Akaike information criterion
ANFCI	Adjusted National Financial Conditions Index
BBG	Bloomberg
BEA	U.S. Bureau of Economic Analysis
BIC	Bayesian information criterion
CART	Classification and regression tree
CBOE	Chicago Board Options Exchange
CDS	Credit Default Swap
CEPR	Centre for Economic Policy Research
CRISP-DM	Cross-Industry Standard Process for Data Mining
CMDI	Corporate Bond Market Distress Index
COVID-19	Coronavirus disease 2019
CUSIP	Committee on Uniform Security Identification Procedures
DB	With different biases
DRD	Moody's Default & Recovery Database
EAD	Exposure at default
EJOR	European Journal of Operational Research
EL	Expected loss
EPU	Economic Policy Uncertainty
FE	Fixed effect
FINRA	Financial Industry Regulatory Authority
FISD	Mergent Fixed Income Database
FRED	Federal Reserve Economic Data of the Federal Reserve Bank of St. Louis
GDP	Gross domestic product
GFC	Global Financial Crisis
ICE	Intercontinental Exchange
IG	Investment grade
IG Reg.	Inverse Gaussian regression
IO	Input-output
IPCA	Instrumented principal component analysis
IRB	Internal ratings based
IRS	Internal Revenue Service
IV	Instrumental variable
LASSO	Least absolute shrinkage and selection operator
Lin. Reg	Linear regression
LS-SVR	Least squares support vector regression
LGD	Loss given default

LTD	Long-term debt
M&A	Mergers and acquisitions
MAE	Mean absolute error
MC+	Minimax concave penalty and penalized linear unbiased selection
MSE	Mean squared error
NAICS	North American Industry Classification System
NBER	National Bureau of Economic Research
NFCI	National Financial Conditions Index
NIG	Non-investment grade
NIPA	National Input and Product Accounts
OLS	Ordinary least squares
OTC	Over-the-counter
PD	Probability of default
PEP	Sparse Gaussian process approximation with power expectation
PEQ	Private equipment
PPE	Property, plant and equipment
R^2	Coefficient of determination
RBF	Radial basis function
Reg. Tree	Regression tree
RF	Random forest
RMSE	Root mean squared error
RR	Recovery rate
RWA	Risk-weighted assets
S&P	Standard & Poor's
SAM	Social accounting matrix
SAS	Statistical Analysis System
SD	Standard deviation
SIC	Standard Industrial Classification
SOI	Statistics of income
SP	Semi-parametric
TRACE	Trade Reporting and Compliance Engine
VaR	Value-at-risk
VIX	CBOE Volatility Index
VXO	CBOE OEX Implied Volatility Index
WRDS	Wharton Research Data Services

Chapter 1

Introduction

Corporate bonds are a key means for firms to raise capital from buy and hold investors, such as insurance companies or pension funds, in exchange for a predictable and stable cash flow. Given the recent decade of low interest rate policies, corporate bond markets have hit record highs in new bond issuance, bringing into question whether credit risk is accurately accounted for by the financial industry, academics, and regulators.[1] In fact, as of finalizing this dissertation, evidence of increasing default rates among U.S. corporate bond issuers emerges as a result of squeezed cash flows and worsening refinancing conditions due to the recent switch to a quantitative tightening monetary policy regime.[2]

To assess a bond's credit risk across its life cycle, three parameters have to be considered. The probability of default (PD) captures the likelihood that a bond issuer will default on its issued debt. The recovery rate (RR), or 1 - loss given default (LGD), measures the proportion of a bond's value that investors are able to recoup in the case of default. Exposure at default (EAD) is the investor's investment amount at risk. Thus, a bondholder's expected loss (EL) can be quantified via $EL = PD \times (1 - RR) \times EAD$. Since the seminal work of Altman (1968), who introduces a scoring model to predict chances of a firm's bankruptcy, various studies followed on explaining the probability of default. However, the recovery rate, i.e., the fraction of an

[1] For example, Goldstein, Jiang, and Ng (2017) pose the question whether corporate bond markets become increasingly fragile with feedback externalities to the real economy. Bessembinder, Spatt, and Venkataraman (2020) raise concerns about potential financial fragility due to the increasing capital allocation in corporate bonds and thus market size growth, and a deteriorating creditworthiness of bond issuers.

[2] See, for example, a recently published article by S&P Global Ratings: *Default, Transition, and Recovery: The U.S. Speculative-Grade Corporate Default Rate Could Rise To 4.5% By June 2024* (August 17th, 2023) available via https://www.spglobal.com/ratings/en/research/articles/230817-default-transition-and-recovery-the-u-s-speculative-grade-corporate-default-rate-could-rise-to-4-5-by-jun-12825499 (Retrieved on October 25th, 2023).

investment amount that is to be repaid to investors in the event when the issuer fails to meet its contractual obligations before the bond's maturity, has received less attention.

While the recovery rate has traditionally been assumed fixed at 40% of investment (see, for example, Altman and Kishore (1996)), today's standard recovery rate estimation models explain it as a function of fundamental drivers such as bond and firm characteristics, as well as indicators of liquidity, supply and demand in the defaulted securities, and macroeconomic conditions.[3] These models, however, neglect pricing implications from the over-the-counter (OTC) bond market microstructure, and adverse contagion and cascade effects in financial markets prominently spotlighted during the Global Financial Crisis of the late 2000s and the COVID-19 pandemic of the early 2020s. Moreover, the hybrid nature of defaulted bonds which share characteristics of both stocks and bonds is not considered, and prediction models have limited real-world applicability. Nevertheless, since the introduction of the Basel II and Basel III accords, banks in the G20 countries are allowed to consider their own internal ratings based (IRB) approaches to calculate risk-weighted assets (RWA) for determining capital requirements and for stress testing. Thus, both investors and regulators are in need of more adequate credit risk assessment.

The main focus of this dissertation is to enhance our understanding of how, and to what extent, recovery rates of defaulted bonds are formed after a bond's default event, relying on data-science based approaches. As such, it aims at offering new explanations and improved recovery estimation models. This dissertation investigates the drivers of credit risk, synthesizing disparate sources of heterogeneity in recovery rates of defaulted corporate bonds. In particular, it captures economic mechanisms from different perspectives that allow to make inferences on recovery rates of the defaulted debt securities, and introduces new recovery rate models that improve the determination of corporate bond recovery rates. In the following, the dissertation outline, research topics, and the knowledge discovery approach are presented.

[3] See, for example, Varma and Cantor (2005), Altman, Brady, Resti, and Sironi (2005), Acharya, Bharath, and Srinivasan (2007), Jankowitsch, Nagler, and Subrahmanyam (2014), Mora (2015), and Nazemi and Fabozzi (2018).

1.1 Outline and Research Topics

Prior to default, a bond represents a creditor's claim on a defined fraction of the issuer's economic value (e.g., in form of a quarterly coupon payment, and repayment of the par amount at maturity). A default event alters the risk-return profile of a bond materially, as coupon payments are often halted, and ultimate repayment of the par amount becomes questionable. Thus, immediately after default, a bondholder's claim on the defaulted firm's economic value, or even the firm's economic value itself, are uncertain. The bond's claim will now be subject to a series of negotiations among equity and debt holders, contractual provisions embedded in the bond indentures, such as collateral and seniority, and the value that is still preserved within the firm's operations or its assets. Typical bond investors often sell once a bond defaults, spurred by internal or external regulations that restrict investment decisions, as is often the case for pension funds, insurance companies, or bond mutual funds. Hence, default events initiate the need for an ownership change from buy-and-hold investors to specialized vulture investors. For the bondholders who sell timely after default, the recovery rate will be equal to the price they obtain in the bond sale transaction. However, given the OTC market structure of bond markets, supply and demand do not meet on an efficient centralized exchange as for equities, and thus, arranging and pricing a transaction are likely not only a function of the seller's and buyer's views on the defaulted firm's fundamentals.

Using detailed bond trading and pricing data of defaulted U.S. corporate bonds from the Trade Reporting and Compliance Engine (TRACE), Chapter 2 examines investors' and dealers' trading patterns in defaulted bonds. While the subsequent Chapters 3, 4 and 5 investigate the drivers of recovery rates from a fundamental perspective, Chapter 2 captures the endogenously adapting microstructure of the OTC bond market, the associated price effects, and thus identifies the intermediation mechanism at work once a bond defaults. Beyond fundamental recovery drivers, it therefore highlights the importance of the dealer network and dealers' expertise for defaulted bond intermediation.

When a bond defaults, the natural holders of bonds change from buy-and-hold to specialized vulture investors, leading to elevated trading levels around the default event. As shown in Chapter 2, post-default intermediation shifts to dealers with prior expertise in the defaulted bond. It identifies primary dealers that are critical for the intermediation of bonds from investors that need to sell quickly in response to the default surprise. These primary dealers locate higher-valuation counterparties within the opaque OTC market through elongating intermediation chains and provide immediacy and liquidity through their inventory, thereby stabilizing the market functioning. While previous studies estimate aggregate recovery rates, this study captures investors' idiosyncratic response to bond default and the heterogeneity of recovery rates not only across bonds, but also across investors even in the same bond: If investors trade with primary dealers, they obtain a recovery premium of 8% over the mean recovery, lowering credit risk ex-ante. Having identified pricing implications from the OTC market microstructure on the defaulted bonds, the subsequent Chapters 3 and 4 investigate fundamental drivers of bond recoveries, and Chapter 5 targets the improvement of modeling accuracy.

As mentioned above, recoveries are determined by creditors' claims on a defaulted firm's economic value in its operations or assets. Acharya, Bharath, and Srinivasan (2007) show that industry-wide distress causes a decrease in recovery rates, given that a defaulted firm's assets disposal channel is impaired by a lower demand for its assets, in line with the phenomenon of fire-sales described by Shleifer and Vishny (1992). Chapter 3 expands the research on the asset disposal channel from a network-based perspective and provides interesting insights in the formation of recovery rates in the context of complex economic interactions between industries.

Creating a network representation of the U.S. economy, this chapter establishes a relation between recovery rates of defaulted bonds and the issuing firm's position within the U.S. economy network of inter-industry trade. Because trade relationships between industries facilitate the transfer of assets from one industry to the other, bonds in better connected or central industries recover more than bonds in peripheral industries. Inter-industry connections not only serve the transfer of firm assets, but also economic shocks. The study shows that distress within an industry propagates across industry borders and affects recoveries in adjacent industries connected via close trade ties. Finally, recovery rates in the core of the inter-industry network are more closely related to macroeconomic conditions than in the periphery.

In Chapter 4, corporate bond recovery rates are examined in relation to observable conditions in financial markets. Because a bond default event alters the bond's risk-return profile and investment characteristics from normal bond-type (low-risk, low-return) to equity-like (higher risk up to the total loss, with potentially higher returns), this chapter considers both conditions in the bond and the stock markets as drivers of recovery. Standard asset pricing models use factors that reflect the relationship between returns of individual assets and the prevailing market-wide conditions. Thus, a variety of equity and bond risk factors have been identified over the last decades.[4] This chapter examines the sensitivity of corporate bond recovery rates to equity risk factors, bond risk factors, and other observable conditions in equity and bond markets. Thereby, this chapter exposes a relationship between the prevailing financial market conditions and the pricing of defaulted corporate bonds. Moreover, it contributes to understanding the integration of equity and bond markets (see, for example, Choi and Kim (2018) and Kelly, Palhares, and Pruitt (2023)).

Chapter 5 examines recovery rates via machine learning to improve the modeling accuracy while accounting for the time-varying structure of recovery rates that is not reflected in traditional recovery rate models or even more advanced models that also employ machine learning. Thereby, it offers a new modeling approach for accurate recovery rate estimation that is more suitable for real-world applications than approaches from the literature. In fact, until today, the time variation in recovery rates has gained little attention in the literature (see, for example, Kalotay and Altman (2017)). Because one can only make predictions on recovery outcomes with information available at the time of default, out-of-sample tests are likely biased due to incorporating data of defaulted bonds from the same issuer for training the model, or from employing data of later points in time.

This chapter shows that machine learning techniques significantly outperform traditional approaches not only out-of-sample as documented in the literature but also in various out-of-time prediction setups. Among the machine learning models, the newly applied sparse power expectation propagation approach provides the most compelling out-of-time prediction results. Motivated by the association of systematic factors with the time-varying characteristic of recovery rates, we study the effect of text-based news measures to account for bond investors' expectations about the future which translate into market-based recovery rates. Especially during recessions, government-related news are associated with higher recovery rates. Although machine learning is a data-driven approach rather than considering economic intuition for ranking a group of predictors, the most informative groups of predictors for recovery rates are nevertheless

[4] See, for example, Fama and French (1993), Jensen, Kelly, and Pedersen (2023), and Dickerson, Julliard, and Mueller (2023)

economically meaningful.

The final chapter, Chapter 6, concludes the work, summarizes the research performed in this dissertation, and gives an outlook on topics that remain for future research.

1.2 Knowledge Discovery Approach

In order to obtain the research results, this dissertation relies on the Cross-Industry Standard Process for Data Mining (CRISP-DM) as a data mining process model that assists the knowledge discovery in the underlying data.[5] CRISP-DM is considered the industry standard for data science projects (see, for example, Kutzias, Dukino, Kötter, and Kett (2023)). Following Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer, and Wirth (2000), the CRISP-DM reference model comprises six phases of a data science project's life cycle: *Business understanding*, *Data understanding*, *Data preparation*, *Modeling*, *Evaluation*, and *Deployment*. Figure 1.1 illustrates the different phases and sequences within the CRISP-DM reference model. In order to provide transparency on the knowledge discovery approach, the dissertation's alignment with each of the model's phases is presented in the following.

Business understanding

As mentioned above, the recovery rate is one of the three key parameters to estimate credit risk. Given the sheer growth of the corporate bond market over the recent years with potentially elevated risks of bond market fragility and the increasing regulatory focus on the IRB approach under Basel II and III, understanding and modeling of the recovery rate has gained in importance for practitioners, academics, and regulators alike. Accurate models for recovery rates allow bond investors to understand the drivers and determinants of economic loss in case of a default event. Risk managers, e.g. in banks, need to estimate the recovery rate for determining the capital reserves that have to be held available. Finally, regulators benefit from understanding how bond markets operate under stress in order to develop policies that preserve market stability. Thus, there is a need for developing a comprehensive understanding of the emergence of corporate bond

[5] Funded by the European Commission, CRISP-DM was developed by a consortium of U.S. and European industrial and commercial organizations: NCR Systems Engineering Copenhagen (U.S. and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (U.S.) and OHRA Verzekeringen en Bank Groep B.V (The Netherlands) (Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer, and Wirth, 2000).

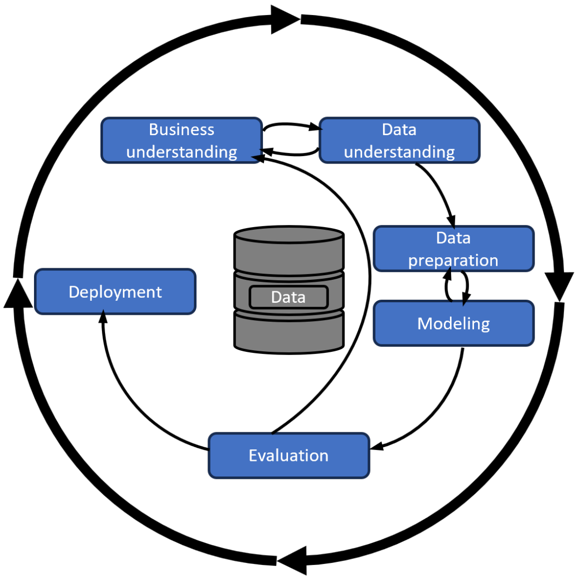


Figure 1.1: Phases of the CRISP-DM reference model (Own illustration following Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer, and Wirth (2000)).

recovery rates and how to accurately estimate them. To achieve this objective, the recovery rate has to be analyzed from various perspectives. Therefore, Chapter 3 sheds light on the OTC trading microstructure in defaulted bonds. Chapters 3 and 4 investigate new fundamental explanations of recovery rates, and Chapter 5 offers a more accurate and real-world oriented approach to recovery rate estimation.

Data understanding

In order to create rich datasets of defaulted bonds, a broad set of data sources is considered. Default events are identified by collecting data from Moody's Default & Recovery Database (DRD), Mergent Fixed Income Securities Database (FISD), S&P Capital IQ fixed-income data, and data from Thomson Reuters. Because default rates and recovery rates are sensitive to the economic cycle, the data spans from 2001–2016 (Chapters 3 and 5) and from 2004–2016 (Chapters 2 and 4), thus, capturing the entire economic cycle including periods before, during, and after the Global Financial Crisis (GFC). As bonds trade in opaque OTC markets, the pricing of bonds is not readily accessible as for equities, which typically trade on centralized exchanges where prices are published immediately. Therefore, bond prices were difficult to obtain historically. Today, bond pricing data is available from data vendors that aggregate the data from different sources, and from regulatory bodies who publish the reported transaction data with a substantial time lag. While the recovery rate in Chapters 2 and 4 is determined based on bond prices that are obtained via S&P Capital IQ from the Intercontinental Exchange (ICE) and considers dealer quotes, live trading levels and data of executed trades, data from the Trade Reporting and Compliance Engine (TRACE) is obtained for Chapters 2 and 4. Because the employed TRACE data sample is available only after 2004, the datasets' time frames differ.

The literature considers a host of explanatory variables, such as bond and firm characteristics or macroeconomic conditions, for recovery rate estimation (see, for example, Altman, Brady, Resti, and Sironi (2005), Acharya, Bharath, and Srinivasan (2007), Bruche and Gonzalez-Aguado (2010), Jankowitsch, Nagler, and Subrahmanyam (2014), and Nazemi and Fabozzi (2018)). This data is collected from various data sources and described in more detail in the respective chapters of this dissertation. In each chapter, further explanatory variables are introduced to improve our understanding and modeling accuracy of recovery rates. For a better understanding of the data, descriptive statistics illustrate the nature and trends within this newly applied data. Chapter 2 relies on detailed transaction data from the Trade Reporting and Compliance Engine (TRACE), which allows not only the identification of individual bond dealers to form a network representation of the inter-dealer U.S. corporate bond market but also to trace each bond as it flows from an investor that sells to a dealer, through the dealer network, and then to another investor. This characteristic of the data enables the observation of endogenous dynamics within the OTC microstructure in response to bond default events. Chapter 3 relies on input-output tables provided by the U.S. Bureau of Economic Analysis (BEA) to create an inter-industry network representation of the U.S. economy. Although the network structure does not exhibit a strong time variation, the data allows to expand on industry-specific recovery rate drivers due to inter-industry trade connections. Chapters 4 and 5 rely on various data from U.S. financial markets and text-based news to complement the traditional recovery rate models.

In order to match bond-specific data obtained from different data sources, CUSIP identifiers are utilized. For time-series data, the most recent observations prior to default are matched with the respective default event, and industry-specific data is matched with a bond issuer's SIC (Standard Industrial Classification) or NAICS (North American Industry Classification System) industry identifier. After constructing the datasets, these comprise cross-sectional default observations with various attributes.

Data preparation

To prepare the data for employing in recovery rate estimation, various data cleaning tasks were performed. For example, TRACE data contains erroneous transaction reports, which are removed through a filter suggested by Dick-Nielsen and Poulsen (2019). Furthermore, TRACE also contains multiple reports of the same trade, thus, duplicate entries are removed. Moreover, potentially falsely reported prices are filtered following Jankowitsch, Nagler, and Subrahmanyam (2014). Corrupted data is removed, and also data for which important accompanying data or features are not available. For example, when creating bond market liquidity measures in Chapter 4, the bond trading information from TRACE must be matched with bond characteristics from FISD. Hence, only bonds that are available in both databases are considered.

Moreover, dimension reduction techniques and feature selection are performed in order to obtain a reasonable dataset. In Chapter 2, a dealer's centrality is considered as a control variable. In order to obtain dealer centrality, however, principal component analysis (PCA) is performed to extract a meaningful centrality representation from a host of alternative centrality measures created with methodologies from network theory. As each of these measures capture different dimensions and definitions of centrality, PCA allows to reduce these measures to only one universal centrality representation. In Chapter 4, where multiple variables from bond and stock markets are considered, principal component analysis is used to reduce dimensions for highly correlated variables. This allows, for example, to obtain one liquidity proxy from twelve bond market liquidity features. Naturally, these variables are correlated with each other as they all capture market liquidity in different ways. Due to the large number of macroeconomic factors, Chapters 3 and 5 further use feature selection techniques (SparseStep, MC+, and stability selection).[6]

Modeling

The task is to estimate the recovery rate, i.e., the bond's price as a fraction of 100% of the bond's par value. In the recovery rate literature, OLS regression models are the standard con-

[6] Data preparation is implemented via Microsoft Excel, SAS, Python, MATLAB, and R.

vention to explain the influence of explanatory variables on the recovery rate (see, for example, Varma and Cantor (2005), Altman, Brady, Resti, and Sironi (2005), Acharya, Bharath, and Srinivasan (2007), Jankowitsch, Nagler, and Subrahmanyam (2014), Mora (2015), and Nazemi and Fabozzi (2018)). To improve modeling accuracy, machine learning techniques are applied (see, for example, Qi and Zhao (2011), Yao, Crook, and Andreeva (2015), and Nazemi and Fabozzi (2018)). However, improved modeling accuracy via machine learning often comes at the cost of reduced interpretability.

As Chapter 2 captures the endogenously adapting OTC microstructure of the bond market for explaining recovery rates, it relies on an OLS approach for recovery rate estimation, complemented with instrumental variable (IV) approaches to allow for consistent and unbiased estimation of the causal effect of primary dealers on bond pricing. For binary dealer classification tasks, generalized linear regressions with a Probit link are utilized. Chapters 3 and 4 also rely on OLS regression for estimating the effect of the tested explanatory variables on the recovery rate. The objective of Chapter 5 is to improve the recovery rate modeling accuracy while accounting for its time-varying structure. Therefore, it not only considers OLS regression, but also machine learning models in various out-of-time prediction settings. These settings are more close to real-world estimation problems than the explanatory studies which aim at explaining the formation of recovery rates. Hence, these out-of-time prediction settings generate more realistic predictions that can be directly implemented by practitioners.

In addition, variable importance ranking methodology is utilized in Chapters 4 and 5 to form an understanding of which of the competing groups of factors matter the most for recovery rate estimation. For machine learning models and variable importance ranking, hyperparameters have to be selected. To do so, grid search with cross-validation is performed. Depending on the objective, different model evaluation metrics are considered. For understanding the effects of newly introduced explanatory variables in Chapters 2, 3, and 4, p-values based on clustered standard errors are used. The (adjusted) coefficient of determination shows how much of the recovery rates' variation can be explained by the linear models. In addition, other metrics such as RMSE or MAE are used in Chapters 3 and 5, which employ machine learning models.[7]

[7] Modeling is implemented via Python, MATLAB, and R.

Evaluation

Model robustness is furthermore checked in various ways, e.g., by altering model specifications with various alternative control variables, or altering variable definitions. For example, Chapter 2 evaluates the results with and without additional dealer characteristics that may interfere with the primary dealer characteristic as the main variable of interest. Chapter 3 proxies industry centrality within the U.S. economy network with several different centrality measures from network theory. As a robustness test, Chapter 3 furthermore considers out-of-sample machine learning models. In Chapter 5, the out-of-time predictions also evaluate the performance of the recovery rate models on subsets of data which have not been used for model training, allowing for a more robust performance benchmark. Furthermore, Chapter 5 benchmarks the performances of various traditional and machine learning techniques to obtain the most accurate recovery rate predictions.

Deployment

While actual deployment in real-world applications will be left for practitioners and regulators, this dissertation provides interesting and important insights and guidelines into explaining and accurately modeling recovery rates of defaulted corporate bonds. A deployment plan should therefore take into account the pricing implications from the endogenously adjusting OTC market structure exposed in Chapter 2, the relationship between recovery rates and interdependences between different industries within the U.S. economy network (Chapter 3), as well as prevailing conditions in stock and bond markets (Chapter 4). Moreover, deployment will benefit from accounting for the time-varying structure of recovery rates (Chapter 5). Constant monitoring, reviewing, and maintenance will help to adapt to changing environments in order to continuously obtain accurate recovery rate estimation results.

Chapter 2

Life after Default: Dealer Intermediation and Recovery in Defaulted Corporate Bonds

This chapter is joint work with Ali Kakhbod, Dmitry Livdan, Abdolreza Nazemi, and Norman Schürhoff.[1] It is a working paper published in 2023 with the same title as Swiss Finance Institute Research Paper No. 23-85 and CEPR Discussion Paper DP18482.[2]

2.1 Introduction

Corporate bonds are traded in decentralized over-the-counter (OTC) markets through dealers who have special trading skills and expertise in searching and locating counterparties, assessing counterparties' willingness to pay, and taking bonds into inventory (Duffie, Garleanu, and Pedersen, 2005; Goldstein and Hotchkiss, 2020; Glode and Opp, 2016; Glode and Opp, 2019; Hugonnier, Lester, and Weill, 2019; Colliard, Foucault, and Hoffmann, 2021; Sambalaibat, 2022;

[1] Ali Kakhbod is with the Walter A. Haas School of Business, University of California, Berkeley, USA. Dmitry Livdan is with the Walter A. Haas School of Business, University of California, Berkeley, USA. Livdan is a Research Fellow of the CEPR. Abdolreza Nazemi is with the School of Economics and Management, Karlsruhe Institute of Technology, Germany. Norman Schürhoff is with the Faculty of Business and Economics and Swiss Finance Institute, University of Lausanne, CH-1015 Lausanne, Switzerland. Schürhoff is a Research Fellow of the CEPR.

[2] Baumann, Kakhbod, Livdan, Nazemi, and Schürhoff (2023), available via https://www.sfi.ch/en/publications/n-23-85-life-after-default-dealer-intermediation-and-recovery-in-defaulted-corporate-bonds (Retrieved on October 25th, 2023) and https://cepr.org/publications/dp18482 (Retrieved on October 25th, 2023).

Chaderina and Glode, 2023). In normal times, insurers and pension funds are the largest investors in corporate bonds with a preference to buy and hold (Koijen and Yogo, 2023). However, when corporate bonds default the natural holders of corporate bonds change due to the altered investment characteristics and risk profile of the bonds (Ivashina, Iverson, and Smith, 2016).[3] Corporate distress events therefore provide a natural setting to study which dealers accommodate the transition from one group of investors to another, how and why dealer expertise matters, and how this affects the pricing of corporate debt beyond the heightened cash-flow risk.

Primary dealers are trading firms that possess the expertise required for providing liquidity and facilitating trading activity in a particular bond on the secondary OTC market. The designation of primary dealer is most commonly associated with government bond markets that are often referred to as "primary dealer markets." Primary dealers play a crucial role in the smooth flow of trading and liquidity for investors interested in buying or selling that bond.[4] In the defaulted corporate bond setting, we consider primary dealers as those that handle most of the bond's order flow prior to default and have developed expertise in that particular bond. The primary dealer is the central dealer in the defaulted bond, so it is bond-central, and it is not necessarily the bond's underwriter or a central dealer in the entire corporate bond dealer network.

We show that a primary dealer-type system has emerged in the corporate bond market without any government intervention or regulation. Consistent with theories of dealers' endogenous trading skills and expertise (Glode and Opp, 2019; Hugonnier, Lester, and Weill, 2019; Sambalaibat, 2022; Chaderina and Glode, 2023), corporate bonds' primary dealers, defined as those dealers that handle most of the order flow after issuance during normal times, play a crucial role in the orderly trading and reallocation of defaulted bonds. Primary corporate bond dealers maintain an orderly market after a bond's default by locating counterparties willing to

[3] When the issuing company fails to meet its contractual obligations, the default event triggers a series of negotiations between bondholders and the issuer that requires specialized expertise and often leads to court enforcement. In many default cases, a creditors' committee or trustee is formed to represent the interests of bondholders during the recovery process and negotiate on their behalf to maximize the recovery for bondholders.

[4] In the government bond market, the primary dealer, or primary market maker designation signifies that this dealer was likely involved in the original issuance of the bond in the primary market and has continued to be a prominent participant in the secondary market for that bond. The U.S. Treasury market has a well-established primary dealer system. Primary dealers are financial institutions authorized by the U.S. Department of the Treasury to participate directly in the auctions of Treasury securities.

trade and pay high prices, executing trades on behalf of other market participants, and buying and selling the bond from their own inventory.

We start by showing that trading volume spikes during corporate default. The dealers' intermediation network then readjusts to handle the abnormal trading activity and investors' need to accommodate the shift in ownership. In particular, investors switch to trading with more central dealers and dealers that have built expertise in intermediating a given bond prior to default. Primary dealers offer a sales channel to small investors when selling pressure dries up liquidity, being willing to take inventory risk in the bonds that they are familiar with, and being able to locate high-valuation investors who specialize in distressed products and other opportunistic investors — potentially faster and more directly than other dealers but through longer intermediation chains. The endogenous reorganization of liquidity provision in the OTC corporate bond market causes recovery prices to increase by over 3 percentage points when routed via primary dealers, equivalent to a more than 8% premium over the mean recovery rate, which significantly stabilizes recovery rates ex-post and lowers credit risk ex-ante. Consistent with the hypothesis that primary dealers have superior trading and pricing skills, that is, bond-specific expertise, we find that post-default price rebounds are significantly attenuated when orders are routed through primary dealers.

More generally, our study highlights that corporate bond market structure evolves endogenously and by taking on a special role during times of stress, primary dealers that are more centrally located in the dealer network stabilize decentralized OTC markets. We provide this novel evidence using a comprehensive sample of corporate bond defaults between 2004 and 2016, assessing how default affects trading, how the OTC dealer network facilitates the ownership transition to specialized distressed investors, and how the intermediation process affects bond recovery. We use granular transaction data and dealer identifiers to construct the intermediation chains and client-to-dealer, dealer-to-dealer, and dealer-to-client trading networks, quantify dealers' inventory risk-taking and search efforts for trading counterparties, and determine the impact of the dealer network on the bonds' recovery rates and post-default price rebound.

Our paper is related to several strands of literature. We explore corporate bond default events as an exogenous shock to a bond's natural holders and, as such, we document event-driven responses of OTC markets and trading implications that were studied in other setups before, such as rating downgrades (May (2010), Ellul, Jotikasthira, and Lundblad (2011) and Bao, O'Hara, and Zhou (2018)), bond index exclusions (Dick-Nielsen and Rossi (2019)) and corporate bond mutual fund redemptions (Goldstein, Jiang, and Ng (2017) and Choi, Hoseinzade, Shin, and Tehranian (2020)). Only a few studies make use of trading data in defaulted bonds. Ivashina, Iverson, and Smith (2016) and Feldhütter, Hotchkiss, and Karakaş (2016) examine the link between pre-default bond trading and the concentration and value of debt claims. Demiroglu, Franks, and Lewis (2022) investigate how transparency of defaulted bond prices impacts wealth transfers between different classes of creditors in Chapter 11 processes. While these studies consider the implications of price transparency, market liquidity, and trading volumes, we highlight an important mechanism for the functioning of the bond market and market participants' trading behavior in defaulted bonds.

We also contribute to the literature exploring the role of dealer networks in OTC markets. Di Maggio, Kermani, and Song (2017), Li and Schürhoff (2019) and Hendershott, Li, Livdan, and Schürhoff (2020) create network representations of trading relations and derive network-driven explanations for the emergence of transactions as well as transaction outcomes. Colliard, Foucault, and Hoffmann (2021) reconcile dealer inventory management with network frictions and the positions of dealers within the dealer network. We highlight the role of primary dealers for defaulted bond intermediation and demonstrate the counterbalancing effects of primary dealers on depressed prices of recently defaulted bonds.

Our study is related to the broader literature on implications of OTC search and bargaining frictions, such as Duffie, Garleanu, and Pedersen (2005) and Feldhütter (2012), as well as Bessembinder, Maxwell, and Venkataraman (2006), Edwards, Harris, and Piwowar (2007), and Goldstein, Hotchkiss, and Sirri (2007). Our findings suggest that transaction outcomes differ for investors depending on their dealer selection and dealers' prior experience in a defaulted bond. We also capture the implications of general bond pricing models (see, for example, Friewald and Nagler (2019)).

We additionally complement the growing literature on OTC dealer capital commitment and liquidity provision. Bao, O'Hara, and Zhou (2018), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), Dick-Nielsen and Rossi (2019), Goldstein and Hotchkiss (2020), Goldberg and Nozawa (2021), and Colliard, Foucault, and Hoffmann (2021) study dealer inventory management in OTC intermediation from various perspectives. Our study highlights the role of dealers in liquidity provision by absorbing excess supply in defaulted bonds through their collective inventories.

Finally, we offer new explanations for defaulted corporate bonds' recovery rates. The recovery rate is usually explained by fundamental drivers such as bond and firm characteristics, as well as macroeconomic conditions (see, for example, Acharya, Bharath, and Srinivasan (2007), Bruche and Gonzalez-Aguado (2010), and Nazemi and Fabozzi (2018)). Altman, Brady, Resti, and Sironi (2005) and Jankowitsch, Nagler, and Subrahmanyam (2014) consider observable market dynamics of defaulted debt securities, such as aggregate supply and demand indicators, as well as liquidity proxies. Compared to their works, our study offers a tangible new contribution, as we implement the newly exposed trading patterns, triggered by default events, in recovery rate estimation. Thereby, we bridge the gap between OTC market mechanics and their implications on recovery rates in credit risk management. Furthermore, we implement recovery rate estimation on the transaction level for the first time in the literature. While previous studies consider average recovery rates per bond, neglecting the variation in recovery rates in the same bond for different investors, we offer a new approach that captures the heterogeneity of recovery prices across different transactions and thereby provides new insights into the idiosyncratic risks in defaulted bonds for practitioners and regulators. Finally, we estimate post-default price appreciation for investors who do not sell immediately after default, which is a novelty in the literature.

The rest of the paper is organized as follows. Section 2.2 describes the data and approach for defining the sample used in our empirical analysis. Section 2.3 documents dealer intermediation in defaulted bonds, and Section 2.4 quantifies the impact of dealer intermediation on recovery rates in defaulted bonds. Section 2.5 examines intermediation complexity in defaulted bonds, and Section 2.6 documents the counterbalancing price effects of trading with primary dealers during default. We conclude in Section 2.7.

2.2 Data

This section describes the data sources used in our empirical analysis and the sample filters used to clean the data. We first create the sample of defaulted corporate bonds by combining several data sets, define pre- and post-default trading periods, and provide descriptions of the explanatory variables used in our analysis. We then construct the dealer network and identify the defaulted bonds' primary dealers.

2.2.1 Default events

In order to create the sample of defaulted bonds, we start by identifying defaulted bonds and their default date during the period 2004–2016 based on two approaches. First, we consider Moody's Default & Recovery Database (DRD), Mergent Fixed Income Securities Database (FISD), S&P Capital IQ fixed-income data, and data from Thomson Reuters to identify and retrieve information about U.S. corporate bond default events. These data sources yield observations associated with three types of corporate default events: reorganizations (Chapter 11), liquidations (both Chapter 11 and Chapter 7), and distressed exchanges. Second, we follow Jankowitsch, Nagler, and Subrahmanyam (2014) and consider rating downgrades to the two worst possible rating categories for which we utilize comprehensive historical rating information by the rating agencies Moody's, S&P, and Fitch Ratings retrieved from FISD. The two rating-based default events are: downgrades to the second worst rating class, e.g., S&P's C rating, representing unlikely-to-pay events or situations in which formal default is considered inevitable but has not yet taken place, and downgrades to the worst rating class, e.g., S&P's D rating, representing actual formal defaults. To capture a bond's default as a single event in our analysis, we select the first default date for a bond and eliminate from our sample all consecutive default events observed within one year.[5]

We apply a number of data filters to our sample of defaulted bonds. In order to be included in

[5] Collecting default events from the various data sources can yield different default dates for some of the bonds. For example, one data source may report a rating downgrade to S&P's C rating which occurs weeks before a downgrade to S&P's D rating or a bankruptcy filing. All these observations are likely to refer to the same default event. In order to represent a bond's unique default event in our analysis as a single observation, we ignore all reported consecutive default dates of a bond that occur within one year after the first default date was observed. After the one-year time lag, a consecutive default observation will be considered a new default event, and the procedure repeats. This approach accounts for consecutive default events of a given bond when it was reinstated following the initial default event.

our analysis, a defaulted bond needs to have basic firm- and bond-specific information in FISD, such as issuer identity and bond seniority. It also needs to be in the Transaction Reporting and Compliance Engine (TRACE) for determining a recovery rate based on transaction prices within the 30-day period immediately after default and in order to match default events to pre- and post-default transactions. We match the defaulted bonds to FISD and TRACE data based on the bonds' CUSIP identifiers. Following this procedure, we are able to identify 2,636 unique U.S. corporate bond default events in the 2004-2016 time period. The default events reflect defaults of 2,425 distinct bonds issued by 498 unique firms. A total of 182, or 7.5% of the bonds defaulted more than once between 2004 and 2016.[6]

Our transaction data is from Academic Corporate Bond TRACE Data, provided by FINRA. The data allows us to track trading volume, terms of trade, and the direction of flows between dealers and investors, to whom we also refer as the dealers' clients. We match the default events with transaction data from TRACE in a 365-day window prior to default as the pre-default period during normal times, and a 30-day window subsequent to default. We define the default day as the event date and the 30 days subsequent to it as the post-default period during which we expect investors' and dealers' trading decisions to be affected by the default event, given the default's surprise character. This definition is supported by the findings of Jankowitsch, Nagler, and Subrahmanyam (2014) who show that trading prices during the 31st to 90th day after default already differ significantly from the 30 days immediately after default.[7] The 30-day period for measuring default event-driven OTC trading patterns is in line with related event studies on bond market reactions.[8] The sample for comparing the pre- and post-default time periods comprises a total of 2,271,772 transactions in 2,636 defaulted bonds. Thereof, 1,956,480 bond transactions occur within the pre-default period corresponding to an average of about 740 individual trades per defaulted bond, and a total of 315,292 bond transactions occur on

[6] Although we take a very similar approach in creating the set of defaulted bonds as Jankowitsch, Nagler, and Subrahmanyam (2014), our approach differs in that we consider only a bond's first default date as a default event and allow consecutive defaults only after a one-year time lag. With our methodology, we count about 1.1 default events per bond over the 13-year period examined in this study. Jankowitsch, Nagler, and Subrahmanyam (2014) consider several different default events for a bond even if they occur simultaneously or within a few days, and count about 2.7 default events per bond over the 8-year period that they analyze.

[7] The choice of a 30-day period is further supported by the possibility to resolve default events timely after default. E.g, emergence from bankruptcy can be achieved in as less as 45–60 days in prepackaged Chapter 11 cases. Time to completion may be also short in distressed exchange events, as exchange offers have to be kept open for a minimum of 20 days, according to Rule 14e-1 of the Securities Exchange Act of 1934.

[8] Ellul, Jotikasthira, and Lundblad (2011) use a 5-week period and Bao, O'Hara, and Zhou (2018) use a 1-month period in their empirical studies on market reactions to corporate bond rating downgrades.

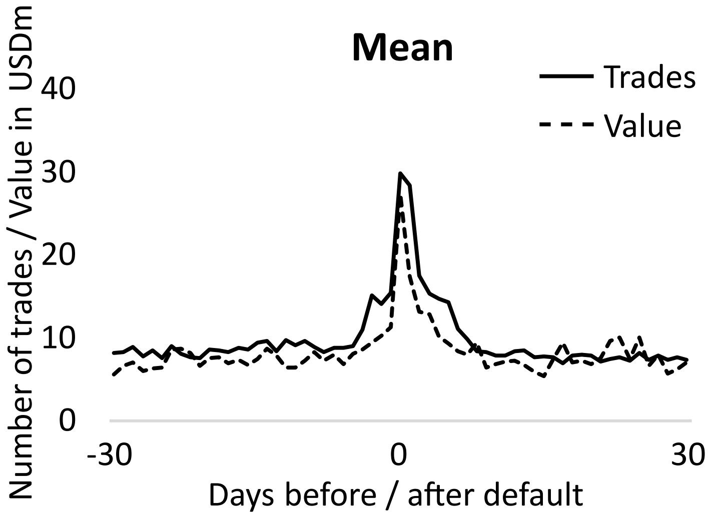


Figure 2.1: **Trading activity before and after default.** Average daily bond trading volume of 2,636 bonds that defaulted between 2004 and 2016 during 30 days before and after default. Average daily volume is shown as the number of trades and total trading value in USDm.

the default date and within the post-default period corresponding to 130 trades per defaulted bond.

Figure 2.1 illustrates trading patterns around default. Trading volume increases as default approaches. During 30 to 1 days before default, the average number of daily trades rises to an average of about 30 trades on the default day as illustrated in Figure 2.1, indicating that default is not fully anticipated by market participants.

For the empirical analysis, we consider additional explanatory variables well-established in the literature to control for alternative channels. We rely on data from FISD and S&P Capital IQ for firm-, industry-, and bond-specific data. Macroeconomic data is retrieved from the Federal Reserve Economic Database of the Federal Reserve Bank of St. Louis (FRED). We replicate several bond liquidity measures using the defaulted bonds' transaction data reported to TRACE. Finally, we utilize information on the pre-default bond ownership structure from eMaxx data to proxy a supply shock as a response to default for our instrumental variable approaches. We provide additional details on the explanatory variables in A.1.

2.2.2 Defaulted-bond dealer network

Masked dealer identifiers in Academic Corporate Bond TRACE data allow us to identify dealers' exposure to intermediation in individual bonds and create a network representation of the

U.S. corporate bond inter-dealer trade relationships. We start by identifying the primary dealer in a defaulted bond by examining client-to-dealer transactions during the year prior to default for each bond separately. Of all dealers identified in the sample, 179 dealers served as primary dealers in at least one or more bonds, that is, having intermediated the highest number of client-to-dealer trades among all dealers in a given bond prior to default. While many of the primary dealers are the primary dealers in just one or a few bonds, 34 of the 179 primary dealers are primary dealers in at least 10 different defaulted bonds. The primary dealers in our data handle on average 31% (median 25%) of the pre-default order flow in a given bond, dispersed between 16% of all trades at the first quartile and 40% at the third quartile.

Following the approach in Li and Schürhoff (2019), we then create a corporate bond inter-dealer network over the years 2004–2016 based on dealer-to-dealer transactions reported to TRACE. We describe the TRACE data, data cleaning, and preparation of the dealer network in more detail in A.1. Figure 2.2 shows a representation of the dealer network as a directed graph, based upon all inter-dealer transactions covered within the period 2004-2016. The network illustrates whether two dealers (nodes) have executed buy or sell transactions (links) with one another and represents all 3,383 dealers that maintain trade relationships with other dealers, based upon 44,065,910 inter-dealer transactions. In Figure 2.2, the majority of both primary dealers and other dealers that intermediate recently defaulted bonds are located within the periphery of the network's core. Top primary dealers are highlighted as triangles. The remaining defaulted-bond dealers are highlighted as cross marks.

In addition to the primary dealer feature, we characterize dealers' centrality for defaulted bonds as controls formally by considering eight commonly used centrality measures to reflect the centrality of dealers within the network: degree, in-degree, out-degree, eigenvector (Bonacich (1972)), betweenness (Freeman (1977)), closeness, as well as in-closeness and out-closeness (Bavelas (1950)). We compute these measures both for (i) an equal-weighted dealer network which solely indicates the existence of a transaction relationship between two dealers and for (ii) an alternative dealer network in which links are weighted by the number of transactions executed between dealers.

In the Appendix, Table A.1 provides summary statistics on dealer centrality. Irrespective of which centrality measure is considered, the observed dealer centralities are consistent with

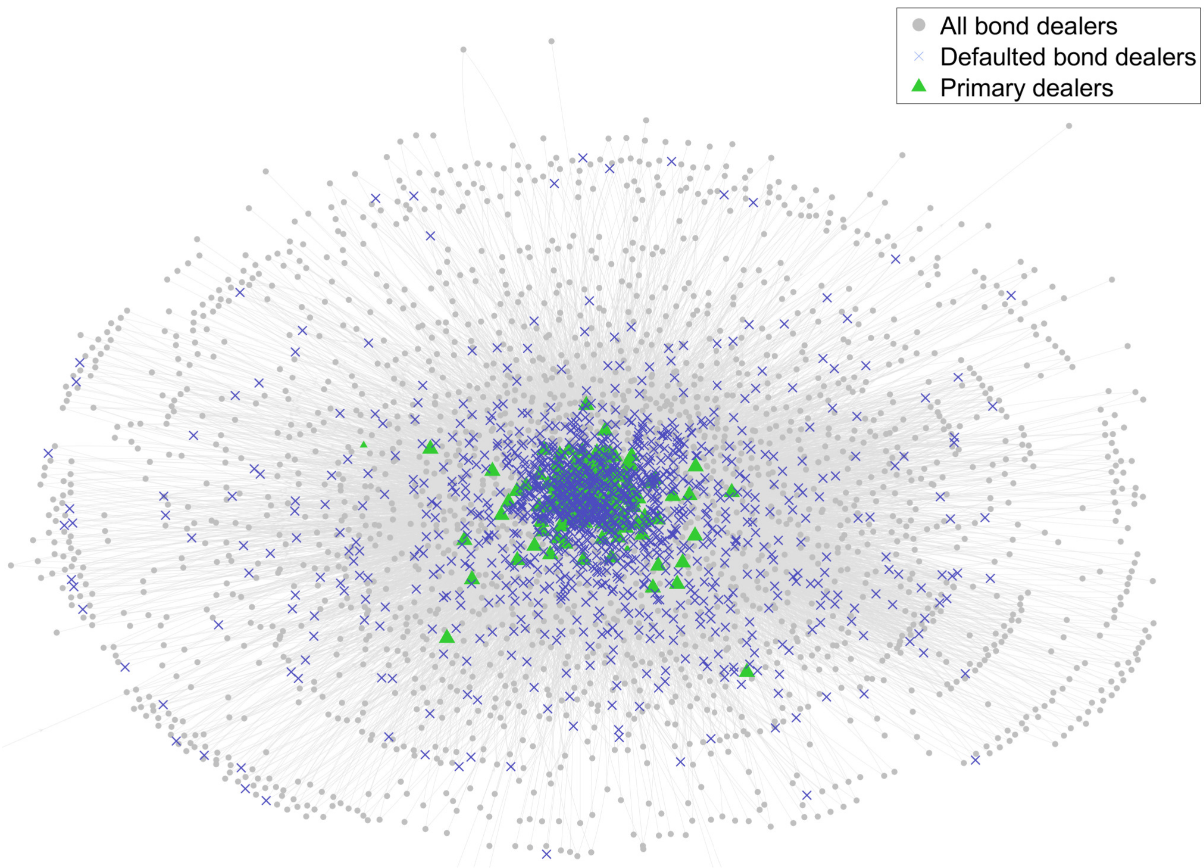


Figure 2.2: **Dealer network before and after default.** The figure illustrates the defaulted-bond dealer network, representing 44,065,910 inter-dealer transactions over the period 2004–2016. Nodes represent the 3,383 bond dealers and links indicate trade relationships between two dealers via bond transactions reported to TRACE. The visualization of the network is performed by a force-directed algorithm that creates attractive forces between neighboring nodes and repulsive forces between distant nodes. Dealers that intermediate bonds within 30 days after a bond's default event are shown as cross marks. Primary dealers that handle most of the pre-default order flow of a bond are shown as triangles. Primary dealers and other defaulted-bond dealers are located around the periphery of the network's core.

the core-periphery structure of dealer networks described in the literature (see, for example, Di Maggio, Kermani, and Song (2017), Hollifield, Neklyudov, and Spatt (2017), and Li and Schürhoff (2019)), as the majority of dealers bear low-rank centralities and are located at the periphery, while few dealers are placed at the core and have high centralities. For the empirical analysis, we implement monthly 1-year trailing dealer networks to capture the network structure timely before default and to account for time variation in the dealer network.[9]

[9] We rely on monthly 1-year trailing networks rather than the complete 2004-2016 dealer network for two reasons. First, we avoid including network information not available at the time of default. Second, although the time variation in the dealer network is limited, we thus account for network changes over time. E.g., one major bond dealer ceased to exist during the financial crisis.

2.3 Dealer Intermediation in Defaulted Bonds

This section examines bond trading before and after bond default and documents intermediation chains and how default affects them. Bondholders faced with a bond's default need to trade and locate suitable specialized distressed and other opportunistic investors willing to hold defaulted bonds. As a result of this need for bond ownership to change, investors' and dealers' trading decisions adjust after default. We will show that retail investors are more likely to sell defaulted bonds to primary dealers after default than before, that is, to dealers who are experts in handling a given bond.

We begin by examining trading behavior in corporate bonds before and after defaults and link it to dealers' identities as primary dealers. When investors decide to sell bonds, they approach dealers, facing a trade-off between execution speed and transaction costs. Dealers themselves face the challenge of locating buyers within the opaque OTC market. Li and Schürhoff (2019) show that central dealers in the municipal bond dealer network provide faster but costlier trade executions. For being able to provide liquidity in defaulted bonds, a dealer must therefore provide sales channels to specialized vulture investors. Retail investors who have less access to prime brokers that cater to large institutions (see, for example, Glode and Opp (2016)) need to locate specialized dealers within the dealer network that are capable of intermediating the distressed securities.

Figure 2.3 illustrates the empirical distribution of dealer centrality among primary and non-primary dealers. The figure shows that dealers intermediating defaulted bonds are different from the average dealer; they are more central than the average dealer. Primary dealers in particular, while not the most central core dealers, tend to be more central than the average dealer in defaulted bonds. A dealer within the network's core belongs to the group of the 19 most central dealers that cumulatively account for more than 25% of all corporate bond inter-dealer trading volume reported to TRACE.

We next test the prediction that (retail) investors are more likely to sell their defaulted bond positions to primary dealers compared to before default. We use the following Probit

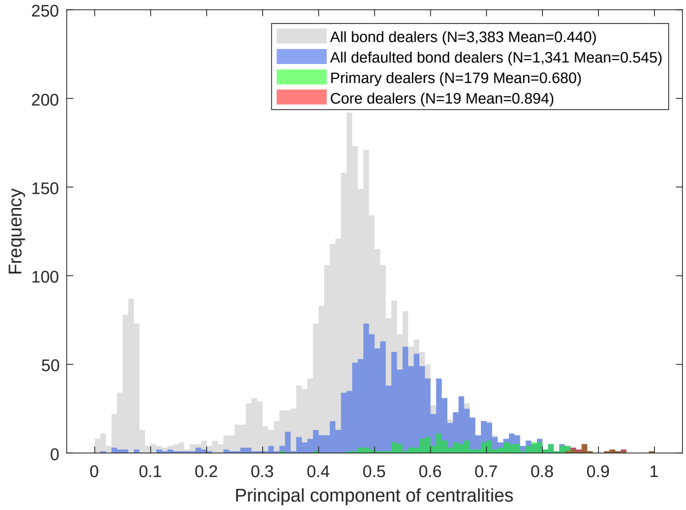


Figure 2.3: Empirical distributions of dealer centralities within the 2004-2016 equal-weighted dealer network. Centrality is scaled on the interval [0,1].

specification to investigate whether clients sell bonds after they default to primary dealers:

$$\begin{aligned} \Pr(\mathit{PrimaryDealer}_{ij} \mid \mathit{Trade}^{CD}_{ij}) = \Phi(\alpha_0 &+ \alpha_1\, \mathit{PostDefault}_{ij} + \alpha_2\, \mathit{PostDefault}_{ij} \times \text{Retail}_{ij} \\ &+ \alpha_3\, \mathit{PostDefault}_{ij} \times \text{LargeInstitutional}_{ij} \\ &+ \alpha_4\, \mathit{DefaultType}_j + \beta'\, X_{ij} + \epsilon_{ij}), \end{aligned} \tag{2.1}$$

with the standard errors adjusted for heteroskedasticity and clustered by bond issue and time. The sample comprises 625,548 client-to-dealer transactions within the year before default and 30 days thereafter. In specification (2.1), $\mathit{PrimaryDealer}_{ij}$ indicates whether the bond j in the client-to-dealer (CD) trade i, labeled as Trade^{CD}_{ij}, is sold to the primary dealer. The key variable of interest is a dummy variable $\mathit{PostDefault}_{ij}$ equal to one if trade i takes place during the post-default period and zero otherwise. To differentiate the effect of the default event on sellers' size, we interact $\mathit{PostDefault}_{ij}$ dummy with an indicator for a retail trade size, Retail_{ij}, defined as positions below \$100,000, and with an indicator for an institutional trade size, $\text{LargeInstitutional}_{ij}$, defined as positions above \$1 million par value. We control for unobserved heterogeneity by including dummies for different default event types, $\mathit{DefaultType}_j$, as well as year fixed effects and trade and bond characteristics, X_{ij}.[10]

[10] These characteristics are comparable to those used by Li and Schürhoff (2019) to predict investors' choice to sell to central dealers in the municipal bond market. However, as we examine the corporate bond market, certain bond characteristics that we employ differ from those available for examining the municipal bond market.

Table 2.1: **Trading with primary dealers before and after default.** Specification 1 is a Probit specification that estimates the probability of clients trading with primary dealers. A total of 625,548 (494,050 pre-default and 131,498 post-default) client-to-dealer trades are considered. The *PostDefault* dummy variable indicates whether a trade takes place after the default event. For comparison, specification 2 uses dealer centrality as the dependent variable in an otherwise similar OLS specification. The explanatory variables further include default event type, trade characteristics, bond characteristics, and year fixed effects. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by issue and time. Significance is denoted *** (1%), ** (5%), and * (10%).

	$PrimaryDealer \mid Trade^{CD}$	$DealerCentrality \mid Trade^{CD}$
Specification	(1)	(2)
$PostDefault$ * Retail	0.67***	0.02
$PostDefault$ * LargeInstitutional	0.02	0.09***
$PostDefault$	−0.50***	−0.07***
Retail	0.25***	−0.21***
LargeInstitutional	−0.04	0.18***
Distressed exchange	0.07	−0.05
Risk rating	0.19**	−0.04
Chapter 11 reorganization	0.23***	−0.07**
Chapter 11 liquidation	0.22**	−0.14***
Chapter 7 liquidation	0.59**	0.21*
Maturity	0.08***	0.03***
Seasoning	−0.07***	0.01
Issue size	−0.17***	0.04**
Rating	−0.01	−0.05*
Junk-rated	−0.14	0.07
Unrated	−0.01	−0.07
Enhanced	0.01	0.06**
Callable	0.38***	−0.01
Sinking fund	0.29**	0.02
Senior unsecured	−0.01	−0.03
Senior subordinate	0.05	−0.04
Subordinate junior	−0.11	−0.10
Coupon	−0.05**	−0.04**
CDS availability	0.07*	0.13***
Covenants	−0.13**	0.01
# observations	625,548	625,548

Table 2.1 reports the results. It is clear from column 1 that primary dealers are more likely to intermediate retail-size trades as the coefficient on Retail_{ij}, 0.25, is positive and statistically significant at the 1%-level. By contrast, the coefficient on $\text{LargeInstitutional}_{ij}$, -0.04, is neither economically nor statistically significant. Column 1 shows that primary dealers continue to intermediate retail-size trades after default events, as the regression coefficient on the interaction term, $PostDefault_{ij} \times \text{Retail}_{ij}$, is equal to 0.67 and it is both economically and statistically, at the 1%-level, significant. Primary dealers do not significantly increase the intermediation of large institutional-size trades for defaulted bonds after the default event since the regression coefficient on the interaction term, $PostDefault_{ij} \times \text{LargeInstitutional}_{ij}$, is equal to 0.02 and it is neither economically nor statistically significant.

Defaulted bonds of firms in Chapters 7 and 11 bankruptcies are significantly more likely to be sold to primary dealers, and in distressed exchanges, sales to primary are the least likely. Bonds with a longer time to maturity are more likely to be sold to primary dealers, and older bonds and bonds with smaller issue sizes are less likely to be sold to primary dealers. For the defaulted bonds, our analysis confirms our prediction that default events trigger a change in investors' trading decisions. Given investors' need to sell recently defaulted bonds, retail investors are more likely to approach primary dealers once a bond default occurs. Results from column 1 demonstrate that predominantly primary dealers are responsible for the intermediation of retail-size trades for defaulted bonds.

To check whether intermediation patterns also change after the default for large institutional trades, we use dealer centrality, $DealerCentrality \mid Trade^{CD}$, as the dependent variable in an OLS specification that is otherwise similar to the specification (2.1):

$$\begin{aligned} DealerCentrality_{ij} \mid Trade^{CD}_{ij} = \alpha_0 &+ \alpha_1\, PostDefault_{ij} + \alpha_2\, PostDefault_{ij} \times \mathrm{Retail}_{ij} \\ &+ \alpha_3\, PostDefault_{ij} \times \mathrm{LargeInstitutional}_{ij} \\ &+ \alpha_4\, DefaultType_j + \beta'\, X_{ij} + \epsilon_{ij}, \end{aligned} \tag{2.2}$$

The idea is that larger institutions tend to have their trades intermediated by more central dealers as shown by Li and Schürhoff (2019). Column 2 of Table 2.1 reports these results. Indeed, the regression coefficient on the interaction term, $PostDefault_{ij} \times \mathrm{LargeInstitutional}_{ij}$, is equal to 0.09 and it is both economically and statistically, at the 1%-level, significant. By contrast, the regression coefficient on the other interaction term, $PostDefault_{ij} \times \mathrm{Retail}_{ij}$, is neither economically nor statistically significant. Unlike primary dealers, central dealers are less likely to intermediate defaulted bonds of firms in Chapter 11 bankruptcy and more likely to intermediate bonds with large issue sizes primarily held by institutional investors.

Overall, these results highlight the changing patterns in the intermediation of defaulted corporate bonds. Retail-size trades of bonds in good standing are mostly intermediated by primary and periphery dealers. After the default, the periphery dealers stop intermediating defaulted bonds potentially due to risk, inventory costs, regulatory constraints, or a combination of all three of them. The primary dealers pick up the slack in intermediating retail-size trades of defaulted bonds. Primary dealers have expertise in these bonds and thus know potential clients who specialize in distressed assets such as hedge funds and other risk-taking investors. Large institutions trade mostly with more central dealers who are capable of placing defaulted bonds with hedge funds or/and taking them in their inventories. Therefore, large institutions continue to trade with the same set of central dealers.

After establishing that the client-dealer network for a bond rearranges itself after a bond default, the next important question is how these changes in the dealer-client intermediation affect the investors' recovery of losses suffered in default. We investigate this question in the next section.

2.4 Impact of Dealer Intermediation on Recovery Rates

Having identified the shift in trading patterns around corporate bonds' default events in Section 2.3, we now examine whether these trading patterns have an effect on the amount investors can recover on their defaulted bonds. Existing studies estimate recovery rates by predominantly utilizing three main groups of economic drivers: firm-specific characteristics, instrument-specific characteristics, and macroeconomic conditions.[11] Only two studies consider observable market dynamics of defaulted debt securities when estimating recovery rates. Altman, Brady, Resti, and Sironi (2005) explain aggregate annual bond recovery rates as a function of aggregate supply and demand for defaulted debt securities by evaluating quoted bond prices. Jankowitsch, Nagler, and Subrahmanyam (2014) extend this market-based approach to recovery rate modeling by implementing bond liquidity proxies derived from trading volumes and prices available from TRACE. We will show that recovery rates are not only driven by factors established in the prior literature which include bond and firm characteristics, and macroeconomic factors, but also by dealers' and investors' endogenously adjusting trading structure in the OTC bond market. Thus, recovery will differ not only across bonds but also across investors even in the same bond.

We define the trade-level recovery rate for transaction i in bond j, RR_{ij}, as a ration of the transaction price $price_{ij}$ to the bond's par value, par_j:

$$RR_{ij} = \frac{price_{ij}}{par_j}. \tag{2.3}$$

[11] See, for example, Altman and Kishore (1996), Altman, Brady, Resti, and Sironi (2005), Acharya, Bharath, and Srinivasan (2007), Bruche and Gonzalez-Aguado (2010), and Nazemi and Fabozzi (2018).

Table 2.2: **Summary statistics for per-bond mean recovery rates.** The recovery rate is calculated as the average trading price in cents per 100 cents in par value, of transactions on the default day and during the 30-day period thereafter.

	Mean recovery rate RR_j in defaulted bond j (% of par)								
	Mean	*SD*	*Min*	*q25*	*q50*	*q75*	*Max*	*N*	*% total*
All defaulted bonds	38.8	28.8	0.0	13.8	34.2	57.2	119.8	2,636	100.0%
Distressed exchange	59.2	29.8	0.4	30.3	59.6	84.8	113.7	197	7.5%
Default risk rating	57.5	27.4	5.7	34.2	58.7	80.5	119.8	306	11.6%
Chapter 11 reorganization	37.4	27.9	0.0	12.3	37.6	53.2	119.6	1,428	54.2%
Chapter 11 liquidation	38.0	28.8	0.0	13.1	29.9	57.2	103.4	92	3.5%
Default rating	26.3	20.7	0.0	13.2	15.9	35.1	106.6	542	20.6%
Chapter 7 liquidation	26.3	34.0	0.0	0.3	11.6	42.5	99.9	71	2.7%

To compute the bond's recovery rate, we utilize transaction prices reported to TRACE. We consider a 30-day post-default period which is a commonly used market convention for defining recovery rates. For bondholders who sell promptly after default or mark their investments to market, the price at default represents actual recovery on investment (Acharya, Bharath, and Srinivasan, 2007). The trade-specific rates in our sample are representative of the actual realized recovery for investors. In addition, we utilize a mean recovery rate in bond j similar to the one used in Jankowitsch, Nagler, and Subrahmanyam (2014):

$$RR_j = \frac{1}{T+1} \sum_{s=t}^{t+T} \left(\frac{1}{|K_{js}|} \sum_{i \in K_{js}} RR_{ij} \right), \tag{2.4}$$

where t is the bond's default date, T is the horizon, and K_{js} is the number of reported bond transactions in bond j on day s. To calculate RR_j we consider transactions reported to TRACE between the default date t and 30 days thereafter.

Table 2.2 reports summary statistics for mean recovery rates defined in (2.4) and split by the event type. The table shows a wide variation in recovery rates. The mean recovery rate in our sample is 38.8% with a standard deviation of 28.8% and a spike at 10-20%. The statistics are in line with the 40% recovery rate that has historically been used as a fixed recovery estimate, as noted by Altman and Kishore (1996). Consistent with the literature, distressed exchange events exhibit the highest recovery rates (see, for example, Franks and Torous (1994), Varma and Cantor (2005), and Mora (2015)). Default risk rating downgrades (e.g. S&P's C rating) yield the second-highest recoveries whereas actual default ratings (e.g., S&P's D rating) and

Chapter 7 liquidations show the worst recovery rates. Default risk rating downgrades may occur well ahead of a formal default event and creditors may then be able to impose more timely measures to preserve their investments, e.g., selling their bond holdings before the firm's situation worsens, or influencing the debtor.

The wide variation in recovery rates that we observe occurs across bonds and time. Over time, the post-global financial crisis years 2009 and 2014 and the credit market stress year 2016 yielded the lowest recoveries. Across industries, bonds from utilities (electricity and gas) recover the most; financial services and savings/loan providers have the worst recovery rates (see, for example, Jankowitsch, Nagler, and Subrahmanyam (2014) and Mora (2015) for evidence).

2.4.1 Mean recovery rates and dealer intermediation

We start by exploring mean recovery rates as defined in (2.4) capturing the expected bond-level recovery. To check the impact of the primary dealer intermediating defaulted bonds, we document whether an intermediating dealer is the primary dealer in a client-to-dealer trade. For a given bond, we define the percentage share of trades that are performed by the primary dealer as:

$$PrimaryDealer_j \equiv \frac{1}{K_j} \sum_{i=1}^{K_j} PrimaryDealer_{ij}, \tag{2.5}$$

where K_j is the number of client-to-dealer trades in bond j.

We set up the following specification to model mean recovery rates, employing variables established in the recovery rate literature (see, for example, Acharya, Bharath, and Srinivasan (2007), Jankowitsch, Nagler, and Subrahmanyam (2014), Mora (2015), and Nazemi and Fabozzi (2018)), and adding the percentage share of primary dealer intermediation as the explanatory variable:

$$RR_j = \alpha_0 + \alpha_1 PrimaryDealer_j + \alpha_2\, DefaultType_j + \beta'\, X_j + \epsilon_j, \tag{2.6}$$

with standard errors adjusted for heteroskedasticity and clustered by issue and time. Controls X_j include a host of fixed effects, default event characteristics, bond characteristics, liquidity measures, macroeconomic features, and firm-level characteristics.

A potential concern with explaining mean recovery rate by $PrimaryDealer_j$ is that common shocks can affect both a bond's recovery after the default and primary dealers' incentives to

Table 2.3: **Mean recovery rates per bond.** The *PrimaryDealer* variable represents the share of client-to-dealer transactions where a given bond is sold to a primary dealer. The mean recovery rate RR_j is the dependent variable in specifications 1 - 4. A total of 2,060 and 1,270 defaulted bonds are considered in specifications 1/3, and 2/4, respectively. Explanatory variables, other than the *PrimaryDealer*, central dealer, and binary variables, are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by firm and time. Significance is denoted *** (1%), ** (5%), and * (10%).

	Mean recovery rate RR_j			
	(1)	(2)	(3)	(4)
Specification	OLS	Bartik-IV	OLS	Bartik-IV
PrimaryDealer	3.44**	12.64***	3.24*	11.84***
Central dealer			−3.10*	−1.99
Dealer size			0.49	0.24
Dealer inventory			−0.64	−2.08***
Trade size	−5.18***	−5.80***	−5.10***	−5.61***
Distressed exchange	25.97***	21.86***	25.91***	21.39***
Risk rating	22.37***	16.10***	22.18***	15.03***
Chapter 11 reorganization	5.58	1.93	5.46	1.60
Chapter 11 liquidation	2.36	−0.47	2.51	−0.08
Chapter 7 liquidation	−19.02**	−18.16**	−19.49**	−18.05**
Bond features	Yes	Yes	Yes	Yes
Seniority FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Industry distress FE	Yes	Yes	Yes	Yes
Liquidity features	Yes	Yes	Yes	Yes
Macroeconomic features	Yes	Yes	Yes	Yes
Company features	Yes	Yes	Yes	Yes
R^2	0.6303	0.6273	0.6312	0.6297
# observations	2,060	1,270	2,060	1,270

intermediate this bond. We utilize Bartik (1991) shift-share instrumental variables (IV) approach to address this endogeneity problem. We first present OLS estimates for α_1 and then IV estimates. In the spirit of Bartik (1991), we instrument the percentage share of primary dealers $PrimaryDealer_j$ participating in the trades of bond j through the expected primary dealer share by first estimating trading with the primary dealer for each transaction separately, using a Probit model. We then aggregate the predicted trade-level primary dealer participation, $\widehat{\Pr}(PrimaryDealer_{ij})$, to an expected percentage share of primary dealer trading for each defaulted bond. Our Bartik (1991)-type instrumental variable for the primary dealer share in bond j is defined as follows:

$$\widehat{PrimaryDealer}_j = \frac{1}{K_j} \sum_{i=1}^{K_j} \widehat{\Pr}(PrimaryDealer_{ij}). \tag{2.7}$$

While the data for constructing (2.7) is on a transaction basis, the analysis is on a per-bond basis.

Using information about bond ownership from the eMaxx database, the instrument is the ownership concentration (Herfindahl–Hirschman index) of institutional investors including insurance companies, bond mutual funds, and pension funds in a given bond prior to default. It proxies for a supply shock, given that investment restrictions may force these investors to sell a bond once it defaults.[12] Therefore, a high ownership concentration of institutional investors in a bond is expected to dry up its market liquidity, such that retail investors who want to sell need to find specialized dealers that still provide liquidity. Thus, the identifying assumption here is that the concentration of ownership among institutional investors is predetermined just before default and affects bond-level recoveries through investors' trade-level decisions to trade with a specialized primary dealer.

Table 2.3 reports results for the impact of primary dealers on mean recovery rates. In specification 1, we use the share of primary dealer participation across all post-default trades, and in specification 2 we replace it with its instrumented primary dealer indicator defined in (2.7). Specifications 3 and 4 are identical to specifications 1 and 2, respectively, except we control for dealer characteristics such as central dealer as a binary indicator denoting whether a dealer is located in the network's core, dealer size, and inventory. The number of observations is lower in columns 2 and 4 because the eMaxx data is missing in some cases.

The coefficient estimates on $PrimaryDealer$ are large and significant. OLS regression specifications 1/3 suggest that recovery rates for a given bond are \$3.44/\$3.24 per \$100 invested higher when transacting with primary dealers compared to non-primary dealers. Given that the mean bond-level recovery rate in our sample is \$38.8 per \$100 invested, this implies an almost 9% premium obtained by investors who sell to primary dealers. The IV estimates are even larger at \$12.64/\$11.84. Since the coefficient on the primary dealer indicator is positive and significant in all specifications, there is a positive relationship between trading with primary dealers and recovery rates at the bond level which is robust when controlling for mean dealer

[12] Investment mandates, internal and external regulatory constraints restrict certain investors such as insurance companies, bond mutual funds, and pension funds in the composition of investment portfolios. Furthermore, except for the temporary spike in volume around the default date, defaulted debt securities may no longer trade in a liquid market. Thus, the sale decisions may not be rational from an unrestricted investor's perspective. Consistently, May (2010), Ellul, Jotikasthira, and Lundblad (2011), Bao, O'Hara, and Zhou (2018), and Reichenbacher and Schuster (2022) show that supply shocks can already be observed in rating downgrades.

size, inventory, and centrality.

The regression coefficient on central dealer is equal to -1.99 in column 4, and it is neither economically nor statistically significant. However, its negative sign provides additional support to a hypothesis that central dealers play a different role than primary dealers when intermediating defaulted corporate bonds. This could be the "need for speed" by institutions willing to trade the faster execution for inferior prices. Specification 4 also reports that the relationship between the dealer size and the mean recovery rates is insignificant and that dealers with more full inventories tend to offer inferior prices. The other regression coefficient estimates suggest that recovery rates decline with the average trade size, mainly because primary dealers intermediate retail-size trades. Recovery rates are higher in distressed exchanges and in risk rating cases than in Chapters 11 and 7. Next, we shift our focus to trade-specific recovery rates.

2.4.2 Trade-specific recovery rates

We estimate recovery rates at the trade level for investors selling to primary compared to non-primary dealers in order to better understand the marginal effect of dealer intermediation on recoveries and the heterogeneity therein across investors.

We set up the following regression specification to model trade-level recovery rates, again employing the explanatory variables established in the recovery rate literature and adding the primary dealer indicator as a determinant:

$$RR_{ij} = \alpha_0 + \alpha_1\, PrimaryDealer_{ij} + \alpha_2\, DefaultType_j + \beta'\, X_{ij} + \epsilon_{ij}, \tag{2.8}$$

where $DefaultType_j$ includes fixed effects of the different default event types. The controls X_{ij} include seniority, year, industry, and industry distress fixed effects as well as bond, liquidity, macroeconomic, and company features. We also add dealer fixed effects in a saturated specification to focus on the default-episode specific role of $PrimaryDealer_{ij}$. Again, we cluster standard errors by bond issue and time. A total of 108,476 post-default client-to-dealer trades are included in the recovery rate specification.

We estimate equation (2.8) in several ways to account for the endogeneity in dealer selection. Given that retail investors select to trade defaulted bonds with primary dealers as shown in Section 2.3, the dealer selection may not be independent from the investors' expected bond recovery. In addition, unobservables may cause endogeneity of the primary dealer. Thus, standard OLS regression potentially provides biased coefficient estimates. Our approach addresses these issues in the following ways.

In addition to specification 1 in Table 2.4 with seniority, year, industry, and industry distress fixed effects and bond, liquidity, macroeconomic, and company features, specification 2 saturates the regression with dealer fixed effects to account for any dealer-specific explanation. In specification 3, we add an instrumental variable to control for correlation between the primary dealer indicator $PrimaryDealer_{ij}$ and ϵ_{ij} in (2.8). As an instrumental variable, we once again utilize the pre-default ownership concentration (Herfindahl–Hirschman index) of institutional investors which proxies for a supply shock. Given the supply shock, investors are then expected to be more likely to sell defaulted bonds to primary dealers. The exclusion restriction here is that a defaulted bond's transaction pricing is not directly correlated with the bond's post-default supply shock at default, except through the liquidity provision that depends on the seller's dealer selection, and thus the selected dealer's identity as a primary or non-primary dealer. In order to adjust for potentially biased estimates stemming from investors' dealer selection, we employ a Heckman (1979) correction approach in an alternative specification. In specification 4, we add the Heckman correction term, the inverse Mills ratio Lambda, as an additional explanatory variable. These approaches allow us to identify the average effect of trading with a primary dealer on recovery prices at the trade level.

Table 2.4 presents the regression results for trade-level recovery rates. The first column shows the coefficient on the instrument in a first-stage Probit model for dealer choice. The instrument is significant in this specification, supporting the instrument's validity assumption. The positive sign suggests that when a bond's ownership structure is concentrated among insurance companies, pension funds, and bond mutual funds, we observe a higher propensity of investors to sell to primary dealers subsequent to a default event. This is expected if institutional investors cause a supply shock in response to default, increasing the need for specialized dealers that provide sales channels to potential investors.

In specifications 1–5 of Table 2.4, we focus on the effect of selling to a primary dealer, $PrimaryDealer_{ij}$, on the transaction-specific recovery rate. The standard OLS specification (specification 1) yields a recovery of \$4.16 per \$100 invested, both statistically at the 1% level

Table 2.4: **Trade-based recovery rates.** The left column is the Probit specification that estimates the probability of clients trading with primary dealers when selling recently defaulted bonds in order to create the instrumental *PrimaryDealer* variable. The binary *PrimaryDealer* variable indicates whether the bond is sold to a primary dealer. The recovery rate *RecoveryRate* is the dependent variable in specifications 1–5. Specification 2 controls for dealer-specific effects. Specifications 3–5 control for potential endogeneity, selection bias, and essential heterogeneity. A total of 108,476 post-default client-to-dealer trades are considered for recovery rate estimation. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by issue and time. Significance is denoted *** (1%), ** (5%), and * (10%).

	Pr(*PrimaryDealer*)	Trade-level recovery rate RR_{ij}				
		(1)	(2)	(3)	(4)	(5)
Specification	1st stage	OLS	Saturated	IV	Heckman	Essent. Het.
PrimaryDealer ($\times p$ in (5))		4.16***	1.45***	7.78***	3.94***	7.93***
Lambda					−9.76	
p						−24.63**
p^2						24.00**
Pre-default institutional HHI	0.93***					
LargeInstitutional	0.16	−0.31	−1.37**	−0.48	−0.52	−0.32
Retail	0.59***	−0.32	−0.23	0.29	−0.96	1.07
Distressed exchange	0.27	8.10	7.93	7.70	7.73	8.42
Risk rating	0.22	11.18**	11.73**	10.82*	10.93*	11.11*
Chapter 11 reorganization	0.09	−7.21	−6.82	−7.83	−7.39	−7.67
Chapter 11 liquidation	−0.02	0.62	0.93	0.44	0.78	0.24
Chapter 7 liquidation	0.61	−2.85	−3.07	−3.08	−3.47	−2.29
Seniority FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry distress FE	Yes	Yes	Yes	Yes	Yes	Yes
Bond features	Yes	Yes	Yes	Yes	Yes	Yes
Liquidity features	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic features	Yes	Yes	Yes	Yes	Yes	Yes
Company features	Yes	Yes	Yes	Yes	Yes	Yes
Dealer FE	No	No	Yes	No	No	No
R^2		0.6004	0.6223	0.5989	0.6007	0.6061
# observations	108,476	108,476	108,476	108,476	108,476	108,476

and economically significant. The estimate is also similar to the estimate of the mean recovery rate from column 1 of Table 2.3. Specification 2 that is saturated with dealer fixed effects yields a consistent sign and significance of trading with a primary dealer as specification 1, however, at a smaller magnitude with \$1.45 per \$100 invested, indicating that some fraction of the pricing benefit captured by the primary dealer is likely captured by other dealer-specific characteristics. The IV specification 3 of the transaction-level recovery rate yields a \$7.78 per \$100 higher recovery at primary dealers, both statistically at the 1% level and economically significant, but somewhat lower than the Bartik (1991) estimate for the mean recovery rate from column 2 of Table 2.3. When we apply the Heckman (1979) correction in specification 4 to correct for selection bias, we obtain a recovery of \$3.94 per \$100 invested, both statistically at the 1% level and economically significant, which is close to the estimate from column 1 of Table 2.3. Overall, the results from specifications 1–4 imply that clients transacting with primary dealers expect to recover between \$4 and \$8 more per each \$100 invested on each transaction as compared to transacting with other dealers, and about \$1.45 when primary dealers' other unobserved characteristics are accounted for via dealer fixed effects.

To account for a heterogeneous response on recovery prices across investors to transacting with primary dealers, we augment specification (2.8) with a model of essential heterogeneity for treatment effects (Heckman, Urzua, and Vytlacil, 2006) derived in A.2 to yield:

$$RR_{ij} = \alpha_0 + \alpha_1\, PrimaryDealer_{ij} \times p_{ij} + K(p_{ij}) + \alpha_2\, DefaultType_j + \beta'\, X_{ij} + \epsilon_{ij}, \tag{2.9}$$

where p_{ij} is the propensity of trading with a primary dealer. Specification (2.9) includes the interaction term between $PrimaryDealer_{ij}$ and the propensity p_{ij}, which we derive from the first-stage Probit model, as well as a linear-quadratic function $K(p_{ij})$ capturing the non-linear dependence of observed recoveries on p_{ij}.

The right column of Table 2.4 reports results from specification (2.9). The negative and statistically significant coefficient on p_{ij} and positive and statistically significant coefficient on p_{ij}^2 indicate a strong essential heterogeneity in the propensity to transact with primary dealers. Moreover, we find that the regression coefficient on the interaction term $PrimaryDealer_{ij} \times p_{ij}$ is positive and statistically significant at the 1% level. Thus, investors decide to sell their defaulted bonds to primary dealers while having at least partial knowledge of the idiosyncratic recovery outcomes of such dealer selection.[13]

The significant positive coefficient on trading with primary dealers in all specifications supports the interpretation that $PrimaryDealer$ positively impacts recovery rates. These results are consistent with theoretical frameworks in which dealers have special skills in intermediating bonds (Glode and Opp, 2019; Hugonnier, Lester, and Weill, 2019; Sambalaibat, 2022; Chaderina and Glode, 2023).

All results reported in Table 2.4 are robust to including and excluding controls for dealer

[13] If we consider third or fourth-order polynomials, signs and significance of the interaction $PrimaryDealer_{ij} \times p_{ij}$ remains stable, and the other coefficient estimates do not change materially.

centrality, dealer size, and dealer inventory, robust to replacing default type fixed effects with firm-level default event fixed effects, and robust to relaxing the single primary dealer definition to the group of the most active dealers, ranked by decreasing number of intermediated trades within the pre-default period, that make up 5% of total trading volume.

2.5 Complexity of Trading and the Role of Primary Dealers

In this section, we explore different dimensions along which the intermediation of bonds changes when they become distressed and how this can explain the higher recovery rates. One possibility is that primary dealers search for higher-valuation buyers potentially longer and by taking bonds into overnight inventory thus making intermediation chains longer and reducing the chance of direct intraday client-to-client matches. This expertise channel can explain why switching to trading through primary dealers causes bond recovery to be higher. More generally, this allows us to investigate how dealers intermediating defaulted bonds match with clients and other dealers when markets are thin. Specifically, we investigate intermediation chain lengths, dealers' role as brokers as opposed to principal traders, whether dealers are more likely to prearrange trades or place defaulted bonds in their inventories, and whether the extensive margins of intra-day trading, i.e., the probabilities of matching with the counterparty within a day, are lower for defaulted bonds than for regular bonds.

2.5.1 Intermediation chains in defaulted bonds

Here we examine whether intermediation chains change for bonds affected by default. Specifically, we are interested in how dealers match sellers and buyers of recently defaulted bonds. We analyze intra-day round-trip intermediation chain length to infer predictions on primary and non-primary dealers' ability to successfully and timely match the supply and demand in defaulted bonds.

To study the length of intermediation chains, we focus our analysis on a sub-sample of successful *intra-day* round trips.[14] Dealers may either complete the bond intermediation chain

[14] We consider intra-day round trips instead of round trips over several days for two reasons. First, only intra-day round trips can be distinctively allocated to either the pre- or the post-default period, as they do not overlap both periods. Second, the economics behind intra-day observations are less likely to be affected by interfering market dynamics or news events related to the defaulted bond that may alter the bond's trading characteristics during the intermediation process, as only a short time frame from the start of the intermediation

by selling to a client, or dealers may sell to another dealer, who then locates the next buyer. Ultimately, bond intermediation is completed through a chain of trades starting with a client-to-dealer trade and ending with a dealer-to-client trade, hence called a round-trip.[15] In between, there may occur one or several consecutive dealer-to-dealer trades. We denote a complete round-trip with N dealers between the seller-client and the buyer-client as a *C(N)DC* intermediation chain. The head dealer within the chain either sells immediately to the next client (*CDC* round-trip), to the next dealer (*CDD* trade chain), or keeps the bond in inventory until the next buyer is located.

We estimate the length of intra-day completed *C(N)DC* intermediation chains before and after default events using the following specification:

$$\begin{aligned} log(N_{ij}) \mid RoundTrip_{ij}^{C(N)DC} &= \alpha_0 + \alpha_1\, PostDefault_{ij} + \alpha_2\, PrimaryDealer_{ij} \\ &+ \alpha_3\, PrimaryDealer_{ij} \times PostDefault_{ij} + \alpha_4\, DefaultType_j + \beta'\, X_{ij} + \epsilon_{ij}, \quad (2.10) \end{aligned}$$

with the standard errors adjusted for heteroskedasticity and clustered by bond issue and time. The sample consists of a total of 143,787 (124,438 pre-default and 19,349 post-default) intra-day *C(N)DC* round-trips. In specification (2.10), N_{ij} is the number of dealers within an intra-day completed *C(N)DC* intermediation chain i in bond j, denoted $RoundTrip_{ij}^{C(N)DC}$, and $PrimaryDealer_{ij}$ indicates whether the head dealer that offsets the initial client-to-dealer trade and initiates a successful intra-day round-trip is the primary dealer in that bond. Controls X_{ij} are the same as used in specification (2.1), but we add dealer characteristics such as size, centrality, inventory, and a dummy for a dealer acting as a broker. To control for unobserved dealer-specific characteristics, we add a saturated specification with dealer fixed effects in column (5).

Table 2.5, reports our results for specification (2.10) with, column 1, and without, columns 2–5, bond dummies. Column 1 shows that intermediation chains are 9% shorter for primary dealers (the regression coefficient is equal to −0.09 and statistically significant at the 1%-level), 12% longer for retail-size trades (the regression coefficient is equal to 0.12 and statistically significant at the 1%-level), and 22% shorter for large institution-size trades (the regression

to its completion is considered.

[15] We allow for trade splits to offset a dealer's position in a dealer-to-client sale.

Table 2.5: **Length of intra-day *C(N)DC* round-trip chains before and after default.** The table provides results of OLS regression that estimates the length of intra-day *C(N)DC* round-trip chains, under the condition that the initial client-to-dealer trade results in a complete intra-day *C(N)DC* round-trip. The dependent variable is the logarithm of the number of dealers within the intermediation chain between two clients. A total of 143,787 (124,438 pre-default and 19,349 post-default) intra-day *C(N)DC* round-trips are considered. The *PostDefault* dummy variable indicates whether a trade takes place after the default event. *PrimaryDealer* indicates whether the bond is sold to the primary dealer. The explanatory variables further include default event type, dealer characteristics, trade characteristics, bond characteristics, and year fixed effects. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by issue and time. Significance is denoted *** (1%), ** (5%), and * (10%).

	$IntermediationChainLength_{ij}$				
Specification	(1)	(2)	(3)	(4)	(5)
$PostDefault$	−0.02*	0.01	0.00	−0.01	−0.01
$PrimaryDealer$	−0.09***	−0.13***	−0.13***	−0.14***	−0.07***
$PrimaryDealer$ * $PostDefault$				0.12***	0.11***
Distressed exchange		−0.05**	−0.06***	−0.06***	−0.02
Risk rating		−0.08***	−0.08***	−0.08***	−0.04***
Chapter 11 reorganization		−0.01	−0.02	−0.02	0.00
Chapter 11 liquidation		−0.06**	−0.09***	−0.09***	−0.04*
Chapter 7 liquidation		−0.13***	−0.11**	−0.11**	−0.10**
Dealer size			0.11***	0.11***	0.07***
Dealer centrality			−0.16***	−0.16***	0.00
Broker role			−0.09***	−0.09***	−0.11***
Dealer inventory			0.01***	0.01***	0.01***
Retail	0.12***	0.20***	0.16***	0.15***	0.04***
LargeInstitutional	−0.22***	−0.28***	−0.27***	−0.27***	−0.19***
Maturity		−0.03***	−0.03***	−0.03***	−0.02***
Seasoning		0.00	0.00	0.00	0.00
Issue size		−0.01	−0.01	−0.01	0.00
Rating		−0.05***	−0.04***	−0.04***	−0.04***
Junk rated		0.15***	0.14***	0.14***	0.15***
Unrated		−0.07*	−0.04	−0.04	0.00
Enhanced		0.00	0.00	0.00	0.02**
Callable		−0.05***	−0.05***	−0.05***	−0.03***
Sinking fund		−0.06	−0.06	−0.06	−0.05
Senior unsecured		0.04**	0.03**	0.03**	0.02**
Senior subordinate		0.02	0.02	0.02	0.02
Subordinate junior		0.04	0.06*	0.06*	0.06**
Coupon		0.02***	0.02***	0.02***	0.01***
CDS availability		0.03**	0.02*	0.02*	0.01
Covenants		0.01	0.00	0.00	0.02*
Bond FE	Yes	No	No	No	No
Dealer FE	No	No	No	No	Yes
# observations	143,787	143,787	143,787	143,787	143,787

coefficient is equal to −0.22 and statistically significant at the 1%-level). While the regression coefficient on the post-default dummy is negative and equal to −0.02, it is not economically significant and statistically significant only at the 10%-level.

When we use bond characteristics instead of bond dummies, column 2 of Table 2.5, and then add dealer size, centrality, inventory, and the broker role dummy, column 3 of Table 2.5, the regression coefficient on the post-default dummy loses its statistical and economic significance. Columns 2/3 show that intermediation chains are 13% shorter for primary dealers, 20%/16%

longer for retail-size trades, and 28%/27% shorter for large institution-size trades. The length of intermediation chains increases with the dealer size and inventory, and they are longer for coupon-paying and riskier bonds, i.e., for high-yield bonds, senior unsecured and junior subordinate bonds issues, and bonds with CDS contracts. The length of intermediation chains declines with dealer centrality (except for specification 5), and it is shorter in dealer-brokered trades, for bonds with a longer maturity, callable, and lower-rated bonds, and bonds traded in risk downgrade default events. Finally, intermediation chains are shorter for bonds of firms in Chapters 7 and 11 bankruptcy liquidations.

We add the interaction term between the primary dealer and post-default dummies to the specification in column 3 of Table 2.5 and report the results in column 4 without, and in column 5 with dealer fixed effects. The regression coefficient on the interaction term is equal to 0.12 and 0.11, respectively, and both are statistically significant at the 1%-level. This implies that intermediation chains for primary dealers are 11%/12% longer post-default than for other dealers. Thus, primary dealers' intra-day matching capability is more affected by default events. By contrast and in line with Hollifield, Neklyudov, and Spatt (2017), intermediation chains initiated by central dealers are generally shorter than those initiated by non-central dealers, and similarly, dealers that act as agencies without taking inventory risk initiate shorter intermediation chains.

2.5.2 Broker vs. dealer role

During normal times, dealers absorb excess supply in corporate bonds through their balance sheets (Goldberg and Nozawa, 2021). We examine primary dealers' tendency to take bonds into inventory as opposed to prearranging trades in defaulted bonds in the role of a broker. We consider trades denoted as agency trades in TRACE and principal trades that are offset within one minute as agency trades.[16]

We estimate the effect of default on dealers' role as brokers versus principals in a Probit specification that controls for a variety of alternative factors, employing 625,548 client-to-dealer trades during the year before a bond's default event until 30 days thereafter. The dependent variable indicates whether the dealer acts as a broker ($BrokerRole_{ij} = 1$) or principal

[16] Our definition of agency trades follows the standard convention in the literature and is in line with Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), Bao, O'Hara, and Zhou (2018), and Li and Schürhoff (2019). In A.1, we provide more information on the prevalence of agency trades.

Table 2.6: **Broker vs. dealer role before and after default.** Probit regression for the probability of dealers to trade as brokers when buying bonds from clients. The dependent variable indicates 1 when the dealer takes the role of a broker (agency) and 0 otherwise (principal). A total of 625,548 (494,050 pre-default and 131,498 post-default) client-to-dealer trades are considered. The *PostDefault* dummy variable indicates whether a trade takes place after the default event. *PrimaryDealer* indicates whether the bond is sold to the primary dealer. The explanatory variables further include dealer characteristics, default event type, trade characteristics, bond characteristics, and year fixed effects. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by issue and time. Significance is denoted *** (1%), ** (5%), and * (10%).

	Pr($BrokerRole_{ij}$)				
Specification	(1)	(2)	(3)	(4)	(5)
$PostDefault$	−0.11***	−0.18***	−0.22***	−0.01	0.01
$PrimaryDealer$	−0.31***	−0.32***	−0.26***	−0.07	0.21***
$PrimaryDealer$ * $PostDefault$				−1.05***	−1.07***
Distressed exchange		−0.10*	−0.13*	−0.09	−0.09
Risk rating		−0.07	−0.10	−0.08	−0.06
Chapter 11 reorganization		−0.10**	−0.15***	−0.11**	−0.11**
Chapter 11 liquidation		−0.15*	−0.24***	−0.21***	−0.22***
Chapter 7 liquidation		−0.19	−0.16	−0.06	−0.17
Dealer size			0.31***	0.31***	0.38***
Dealer centrality			−0.71***	−0.71***	−0.18***
Dealer inventory			−0.01*	0.00	0.00
Retail	0.56***	0.53***	0.39***	0.40***	0.25***
LargeInstitutional	−0.18***	−0.22***	−0.11***	−0.10***	0.07***
Maturity		−0.04***	−0.04***	−0.03**	−0.01
Seasoning		−0.03**	−0.03*	−0.02	−0.01
Issue size		−0.01	0.00	−0.01	0.02
Rating		−0.01	0.00	−0.02	−0.04
Junk rated		0.00	−0.02	0.04	0.09
Unrated		−0.10	−0.07	−0.11	−0.04
Enhanced		0.02	0.05	0.03	0.06
Callable		−0.18***	−0.18***	−0.14***	−0.07*
Sinking fund		0.10	0.12*	0.11	0.25***
Senior unsecured		0.01	0.00	0.00	0.05
Senior subordinate		−0.04	−0.05	−0.06	0.02
Subordinate junior		0.07	0.04	0.04	0.15
Coupon		0.01	−0.01	−0.01	−0.02
CDS availability		−0.14***	−0.11***	−0.11***	−0.08**
Covenants		0.04	0.05	0.04	−0.02
Bond FE	Yes	No	No	No	No
Dealer FE	No	No	No	No	Yes
# observations	625,548	625,548	625,548	625,548	625,548

($BrokerRole_{ij} = 0$). We estimate the following relationship:

$$\Pr(BrokerRole_{ij} \mid Trade_{ij}^{CD}) = \Phi(\alpha_0 + \alpha_1\, PostDefault_{ij} + \alpha_2\, PrimaryDealer_{ij} + \alpha_3\, PrimaryDealer_{ij} \times PostDefault_{ij} + \alpha_4\, DefaultType_j + \beta'\, X_{ij} + \epsilon_{ij}), \quad (2.11)$$

with standard errors adjusted for heteroskedasticity and clustered by bond issue and time. In this specification, controls X_{ij} include trade and bond characteristics, and year fixed effects.

Table 2.6 provides results for the Probit estimates for the dealer's role. In specifications

1–3, we find that dealers are significantly less likely to act as brokers once a bond defaults. This highly significant effect suggests that dealers provide immediacy to sellers of recently defaulted bonds who must sell defaulted bonds quickly. Dealers take recently defaulted bonds and the associated risks on their own balance sheets rather than searching for a willing buyer first. We further find a strong negative effect of primary dealers in specifications 1–3 where we do not add primary dealer interactions which demonstrates that primary dealers more readily risk their own capital for intermediating defaulted bonds for which they had handled most of the order flow prior to default. Moreover, the negative interaction effects in specifications 4 and 5 demonstrates that primary dealers are even more likely to act as principals once a bond defaults. As one would expect, primary dealers are more likely to take bonds into inventory, likely because they are familiar with the bond, and the potential investor universe, and are thereby able to better manage the risk of ownership than other dealers. Complementing the findings of Goldstein and Hotchkiss (2020) who show that central dealers are more likely than peripheral dealers to provide inventory capacity, we find that primary dealers are more likely to take defaulted bonds into their inventory than other dealers. We furthermore find a significant positive effect of dealer size on the probability of trading as a broker.

2.5.3 Complexity of matching

We estimate whether default events have an impact on the probability of dealers selling bonds on the same day as acquired, instead of keeping them in inventory overnight. Here, we consider all 625,548 client-to-dealer and consecutive offsetting trades observed during the year before default and until 30 days thereafter. More formally, we apply a Probit model that estimates whether a dealer sells a bond that they recently acquired from one of their clients through a consecutive dealer-to-client trade (*CDC* round-trip) or dealer-to-dealer trade (*CDD* trade chain) on the same day ($\mathit{IntraDayMatch}_{ij} = 1$), or whether the dealer keeps the bond in inventory overnight ($\mathit{IntraDayMatch}_{ij} = 0$):

$$\begin{aligned}\Pr(\mathit{IntraDayMatch}_{ij} \mid \mathit{Trade}_{ij}^{CD}) = \Phi(&\alpha_0 + \alpha_1\, \mathit{PostDefault}_{ij} + \alpha_2\, \mathit{PrimaryDealer}_{ij} \\ &+ \alpha_3\, \mathit{PrimaryDealer}_{ij} \times \mathit{PostDefault}_{ij} + \alpha_4\, \mathit{DefaultType}_j + \beta'\, X_{ij} + \epsilon_{ij}). \quad (2.12)\end{aligned}$$

Variable definitions are similar to those used in specification (2.10). Additionally, we include an interaction term between $PostDefault_{ij}$ and $PrimaryDealer_{ij}$ in the baseline specification. We adjust standard errors for heteroskedasticity and cluster by bond issue and time.

Table 2.7 shows the regression results. We find in specifications 1–3 that dealers are significantly less likely to sell a recently defaulted bond on the same day as acquired, compared to bonds that have not yet defaulted. This finding demonstrates that dealers are indeed more likely to keep bonds that recently defaulted in overnight inventory at the end of the day on which clients offload their bonds to the dealers. Thus, dealers commit their capital by taking these defaulted bonds in overnight inventory. The results show that primary dealers are more likely to take bonds in overnight inventory than non-primary dealers, and the significant and negative interaction between the primary dealer indicator and the post-default dummy variable in specifications 4 and 5 indicates that the likelihood of primary dealers to utilize overnight inventory increases significantly more for primary dealers once a bond defaults.

Dealers absorb client sell orders depending on the severity of the default event type, with distressed exchanges and risk rating downgrades showing the smallest, and Chapter 7 liquidations showing the largest significant effects. Thus, dealers are more likely to keep bonds of the most severe default event type in inventory overnight. When dealers act as brokers, the likelihood of them offsetting trades on the same day is significantly higher, which corresponds to the broker role of dealers without utilizing their own inventory. This is reasonable, given that dealers prearrange trades when they act as brokers. Moreover, when dealers have accumulated bond inventories over the recent month, indicated by dealer inventory, the likelihood of putting additional bonds into inventory is significantly lower, highlighting constraints in dealer inventory that impede additional risk-taking. In general, our analysis shows that dealers, particularly primary dealers, are more likely to keep recently defaulted bonds in inventory overnight rather than sell them to another counterparty on the same day.

2.5.4 Inventory risk taking

We now explore dealers' role in facilitating transactions of defaulted bonds by providing inventory capacity, and potentially conducting proprietary trading in defaulted bonds. In Section 2.5.3 we have demonstrated that dealers are more likely to take a bond in overnight inventory

Table 2.7: **Intra-day matches before and after default.** Probit regression for the probability of dealers matching a client-to-dealer trade with a consecutive buyer on the same day. The dependent variable indicates 1 if the dealer sells the bond to the next buyer on the same day as acquired from the client, and 0 if the bond remains in the dealer's inventory at the end of the day. Buyers may be clients or other dealers. A total of 625,548 (494,050 pre-default and 131,498 post-default) client-to-dealer trades are considered, of which 331,315 (276,193 pre-default and 55,122 post-default) are matched with either dealer-to-client trades (*CDC* round-trip) or dealer-to-dealer trades (*CDD* trade chain) on the same day. The *PostDefault* dummy variable indicates whether a trade takes place after the default event. *PrimaryDealer* indicates whether the bond is sold to the primary dealer. The explanatory variables further include dealer characteristics, default event type, trade characteristics, bond characteristics, and year fixed effects. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by issue and time. Significance is denoted *** (1%), ** (5%), and * (10%).

	Pr($IntraDayMatch_{ij}$)				
Specification	(1)	(2)	(3)	(4)	(5)
$PostDefault$	−0.21***	−0.25***	−0.24***	−0.09**	−0.07*
$PrimaryDealer$	−0.61***	−0.68***	−0.63***	−0.49***	−0.14***
$PrimaryDealer$ * $PostDefault$				−0.76***	−0.93***
Distressed exchange		−0.19***	−0.20***	−0.17**	−0.14**
Risk rating		−0.17***	−0.19***	−0.18***	−0.12**
Chapter 11 reorganization		−0.20***	−0.24***	−0.21***	−0.17***
Chapter 11 liquidation		−0.24***	−0.29***	−0.27***	−0.22***
Chapter 7 liquidation		−0.41***	−0.38**	−0.31**	−0.39**
Dealer size			0.12***	0.13***	0.16***
Dealer centrality			−0.38***	−0.40***	−0.08***
Broker role			1.59***	1.56***	1.48***
Dealer inventory			−0.02***	−0.02***	−0.01**
Retail	0.08***	0.11***	−0.24***	−0.23***	−0.31***
LargeInstitutional	0.26***	0.19***	0.34***	0.34***	0.54***
Maturity		−0.03**	−0.01	0.00	0.01
Seasoning		−0.01	0.01	0.01	0.01
Issue size		−0.06***	−0.06***	−0.07***	−0.04**
Rating		−0.04	−0.04	−0.06**	−0.05*
Junk rated		−0.09	−0.12*	−0.07	−0.04
Unrated		−0.26***	−0.26***	−0.28***	−0.21***
Enhanced		−0.02	−0.02	−0.04	0.02
Callable		−0.18***	−0.13***	−0.10***	−0.01
Sinking fund		−0.06	−0.11	−0.11	−0.10
Senior unsecured		0.00	−0.01	0.00	0.01
Senior subordinate		−0.09*	−0.09**	−0.10**	−0.10**
Subordinate junior		−0.06	−0.12**	−0.12**	−0.14**
Coupon		0.05***	0.04***	0.05***	0.03**
CDS availability		−0.16***	−0.11***	−0.11***	−0.02
Covenants		−0.08**	−0.13***	−0.14***	−0.08**
Bond FE	Yes	No	No	No	No
Dealer FE	No	No	No	No	Yes
# observations	625,548	625,548	625,548	625,548	625,548

after its default event than before. Dealers thus commit their own capital in trades, facilitating the timely execution of bondholders' sale orders. After committing capital for placing bonds in overnight inventory, a dealer may sell the bond to another client or dealer on the next day, or several days later, provided that a counterparty is found. Thereby, the dealer eliminates their idiosyncratic exposure to the defaulted bond, but in cases when the bond is sold to another dealer, the collective dealers' commitment to a defaulted bond's par value will remain constant.

Table 2.8: **Dealers' aggregate inventory in bonds one day before their default and 30 days thereafter.** The dealer inventory is denoted in the percentage of a bond's par value that is held on dealers' balance sheets.

	N	*Mean*	*SD*	*q5*	*q25*	*q50*	*q75*	*q95*
Dealer inventory before default	2,474	1.7%	8.5%	-8.1%	-0.6%	0.8%	3.7%	14.8%
Dealer inventory after 30 days	2,474	2.6%	9.8%	-8.4%	-0.6%	1.2%	5.1%	18.2%
Difference	2,474	0.9%***						

As we are interested in knowing whether, and to which degree, dealers collectively commit capital in order to compensate for a mismatch in market supply and demand triggered by a bond's default event and to bridge the gap between bondholders' sale and the time high-valuation buyers are found, we analyze dealers' net inventory positions in defaulted bonds. That is, we examine whether dealers collectively absorb investors' selling pressure induced by corporate bond default events.

As bond dealers do not disclose inventory levels or the amount of capital they put at risk in individual bonds, we use a relative measure of inventory that tracks changes in inventory from a fixed reference date, following the methodology of Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018). The details for determining inventories are provided in A.1.

Table 2.8 reports summary statistics of dealer inventories. Here, the reference date is one year prior to default and the inventory is reported relative to this reference date. Compared to the day before default, dealers additionally accumulate an average of 0.9 percentage points of a defaulted bond's par value during the 30-day period after default, which is significant at the 1% level. While the table shows that dealers even reduce inventories in some cases, dealers accumulate more than 5.1% of par value for one-quarter of all defaulted bonds during the 30-day period after default. Hence, dealers' collective capital commitment has an important effect on the liquidity provision of recently defaulted bonds. As the inventory buildup prior to default further demonstrates, dealers absorb selling pressure even when the bond has not yet defaulted. This evidence is consistent with dealers working harder to intermediate defaulted bonds.

2.6 Dealer Expertise and Post-Default Price Efficiency

The defaulted-bond setting is special in that dealers have to both counterbalance selling pressure and at the same time find high-valuation buyers, e.g., specialized vulture investors that can reap high recoveries in post-default negotiations. To check whether primary dealers counterbalance the negative price impact from selling pressure during default and trade with higher-valuation buyers at more information-efficient prices, we investigate how trading with primary dealers affects price reversal subsequent to default.

We measure the bond price rebound between observed recovery prices in transactions of investors who sell immediately after default (i.e., within the 30-day post-default period) and subsequent prices observed when the surprise element of default has already faded. As we intend to capture only prices that relate to investors' recovery, we again consider only client-to-dealer sale transactions. As such, we do not include prices paid between dealers or by investors. Furthermore, we focus on the short-term price appreciation rather than long-term effects, as we intend to identify the primary dealers' stabilizing effect related to the default surprise which is likely to vanish shortly after the initial supply shock. Because prices may still fluctuate even after the default surprise has vanished, we consider a relatively short 10-day window at the beginning of the second month after default for measuring the pricing of client-to-dealer trades likely unaffected by the initial price pressure.[17]

We define price appreciation $PriceAppreciation_{ij}$ as bond j's price difference in a client-to-dealer transaction i during the 30 days after default, and the mean daily prices paid in client-to-dealer transactions from 31 to 40 days after default:

$$PriceAppreciation_{ij} = \frac{1}{T+1} \sum_{s=t}^{t+T} \left(\frac{1}{|K_{js}|} \sum_{k \in K_{js}} RR_{kj} \right) - RR_{ij}, \tag{2.13}$$

where K_{js} is the number of trades in bond j on day s, starting 31 days after default, day t, until 40 days after default. $PriceAppreciation_{ij}$ thus captures the percentage points of a bond's par value that an investor who sells immediately after default forgoes, rather than

[17] Trading volume is still high in the second month after default, and most bankruptcies are unlikely to be resolved by then, allowing us to employ a comparable sample size as for estimating primary dealers' effects on recovery rate. Furthermore, as claim holders typically enter negotiations shortly after default, prices in later time periods may already reflect measures taken by the firm to resolve distress or new expectations about ultimate recovery.

holding the bond until the second month after default. We estimate how investors' decision to trade with a primary dealer affects the observed price differences between the two time periods. More specifically, we employ the following specification for estimating post-default bond price appreciation:

$$PriceAppreciation_{ij} = \alpha_0 + \alpha_1\, PrimaryDealer_{ij} + \alpha_2\, DefaultType_j + \beta'\, X_{ij} + \epsilon_{ij}, \qquad (2.14)$$

where $PriceAppreciation_{ij}$ is the price difference as defined in (2.13). The control variables are similar to those used in (2.8). Here, a total of 106,961 post-default client-to-dealer trades are considered. To account for a heterogeneous response, we also consider the model in A.2, now with $PriceAppreciation_{ij}$ as dependent variable, and otherwise the same as specification (2.9):

$$\begin{aligned} PriceAppreciation_{ij} = \alpha_0 + \alpha_1\, PrimaryDealer_{ij} \times p_{ij} + \alpha_2\, K(p_{ij}) \\ + \alpha_3\, DefaultType_j + \beta'\, X_{ij} + \epsilon_{ij}. \end{aligned} \qquad (2.15)$$

We expect more efficient prices and hence less price rebound due to the primary dealer's expertise. Given that bonds that are sold to primary dealers immediately following a default event achieve higher recoveries, the subsequent price reversal should be less pronounced for these transactions.

Table 2.9 reports our results for five variants of specifications (2.14)/(2.15). The two columns to the left use the actual primary dealer indicator as explanatory variable, without (specification 1) and with (specification 2) dealer fixed effects. Specification 3 considers the instrumented primary dealer indicator, specification 4 follows the self-selection correction approach of Heckman (1979) and specification 5 employs the model of essential heterogeneity. The instrument is created as in Table 2.4. As all five columns show, selling to primary dealers immediately after default is equivalent to counterbalancing temporary price pressure, given that the following price rebound is less pronounced for those trades routed via primary dealers. In specification 2, where we add dealer fixed effects, the primary dealers' effect remains negative, although at a smaller magnitude than the other specifications. This is in line with the presence of unobserved dealer-specific characteristics that are correlated with the primary dealer indicator. The Lambda in column 4 is significant, indicating the presence of selection bias in specifications 1–3. Finally,

Table 2.9: **Post-default price appreciation.** The binary *PrimaryDealer* variable indicates whether the bond is sold to a primary dealer. The price appreciation *PriceAppreciation* is the dependent variable in specifications 1–5. Specifications 3–5 control for potential endogeneity, selection bias, and essential heterogeneity. A total of 106,961 post-default client-to-dealer trades are considered for price appreciation estimation. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by issue and time. Significance is denoted *** (1%), ** (5%), and * (10%)..

	PriceAppreciation				
	(1)	(2)	(3)	(4)	(5)
Specification	OLS	Saturated	IV	Heckman	Essent. Het.
PrimaryDealer ($\times p$ in (5))	−3.48***	−0.99**	−7.63***	−2.95***	−6.45***
Lambda				23.72***	
p					3.79
p^2					−11.34*
LargeInstitutional	−0.67	0.05	−0.49	−0.21	−0.29
Retail	0.34	−0.13	0.33	1.93***	0.64
Distressed exchange	−27.45***	−27.05***	−27.01***	−26.53***	−26.92***
Risk rating	−22.45***	−22.21***	−22.07***	−21.87***	−21.95***
Chapter 11 reorganization	−25.99***	−25.47***	−25.38***	−25.53***	−25.35***
Chapter 11 liquidation	−29.69***	−28.70***	−29.61***	−30.10***	−29.73***
Chapter 7 liquidation	−22.85***	−21.24***	−21.94***	−21.40***	−21.73***
Dealer FE	No	Yes	No	No	No
Seniority FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Industry distress FE	Yes	Yes	Yes	Yes	Yes
Bond features	Yes	Yes	Yes	Yes	Yes
Liquidity features	Yes	Yes	Yes	Yes	Yes
Macroeconomic features	Yes	Yes	Yes	Yes	Yes
Company features	Yes	Yes	Yes	Yes	Yes
R^2	0.4705	0.5038	0.4705	0.4747	0.4834
# observations	106,961	106,961	106,961	106,961	106,961

the specification that accounts for heterogeneity captures a similar effect of *PrimaryDealer* on *PriceAppreciation* as the other specifications, and it includes a significant non-linear term. Although slightly smaller, the estimated coefficients of *PrimaryDealer* in Table 2.9 are of a similar magnitude as in the corresponding specifications in Table 2.4.

Overall, the primary dealer coefficients in Tables 2.4 and 2.9 indicate that trading with a primary dealer leads to higher and more stable recovery prices immediately after default vis-à-vis prices observed once the initial default surprise has vanished. These findings suggest that a major share of the pricing benefits provided by primary dealers during the default-induced times of stress results from their superior expertise, and is not due to fire sale discounts or price pressures.

2.7 Conclusion

While there exists an extensive literature on the intermediation of corporate bonds in good standing, little is known about the intermediation of defaulted corporate bonds. When a bond becomes distressed, a need arises to change bond ownership since specialized vulture investors are better able to recoup high recovery values and avoid aggregate losses. We present a comprehensive body of evidence on the intermediation of defaulted corporate bonds.

Our empirical analysis reveals that trading spikes and intermediation patterns undergo significant changes following a bond's default, in that not all dealers transact in defaulted bonds, intermediation chains prolong, the average centrality of transacting dealers rises, and primary dealers intermediate much of the post-default order flow. Similar to the primary dealer system observed in government bond markets, investors direct their order flow to the bond's primary dealer(s) who have developed specialized intermediation expertise in that particular bond prior to its default. This preference stems from primary dealers' ability to offer more direct access to specialized, higher-valuation investors and their superior capability in managing inventory risks associated with defaulted bonds during times of market stress.

The advantages for investors of transacting with primary dealers are both higher recovery rates and these recovery rates are more informationally efficient in that they are closer to the bond's long-term value. Despite the general drop in value for bonds after default, investors who sell to primary dealers realize recoveries that exceed 8% of the average recovery compared to other dealers. The higher recoveries are accompanied by more stable post-recovery bond prices and less price rebound. Primary dealers thus contribute to recouping higher recovery values and stabilizing a bond's market functioning which lowers credit risk ex-ante for all investors.

Chapter 3

Inter-Industry Network and Corporate Bond Recovery Rates

This chapter is joint work with Abdolreza Nazemi and Frank J. Fabozzi.[1] At the time of completing this dissertation, this study has received a revise and resubmit decision at Journal of Banking and Finance.[2]

3.1 Introduction

Over the past decades, financial economists have identified a great number of risk drivers that explain the probability of default. In contrast, far fewer studies are devoted to the recovery rate as another key credit risk parameter. With the experience of the recent Global Financial Crisis (GFC), it is critical to enhance the understanding of common risk factors that determine recovery rates. Bankruptcies of major financial institutions caused by distress spillovers through economic interconnections and subsequent government bailouts have shown that the financial industry had not been prepared to cope with distress propagation through the economy's inherent interconnections. Credit risk models turned out to be inadequate, as they do not sufficiently account for linkage-based interdependencies between agents within the economy. By creating a network of inter-industry customer-supplier ties, we investigate the relationship between the recovery rates of U.S. corporate bonds and network-derived characteristics. This paper shows

[1] Abdolreza Nazemi is with the School of Economics and Management, Karlsruhe Institute of Technology, Germany. Frank J. Fabozzi is with the Carey Business School, Johns Hopkins University, Baltimore, USA.

[2] Nazemi, Baumann, and Fabozzi (2023)

that network-derived characteristics indeed play a key role in determining recovery rates in the corporate bond market.

Interdependencies of behavior, information and monetary flows within economic systems are particularly evident today, and as both the GFC and the COVID-19 pandemic have shown, the world scale of interconnected economies and quick interactions of market participants can lead to adverse contagion and cascade effects if not appropriately managed. On the other hand, as not only the globalized economy demonstrates, connectedness has the potential to serve as a catalyst for economic success. Hence, it is essential to understand both the merits and risks of connectedness within the economy. In this paper, we introduce the tool of financial networks for investigating corporate bond recovery rates. We find significant evidence that the inter-industry trade network structure of the U.S. economy and a firm's position within the network explain a large fraction of U.S. corporate bond recovery rates, as trade relations serve as a channel for the inter-industry transfer of assets and distress. Our work is closely connected to a number of recent studies in finance based on inter-industry network approaches.[3]

The disposal of assets is a major tool to recover economic value in a bankruptcy proceeding. As Ahern and Harford (2014) illustrate, strong inter-industry trade relationships serve as predictors of inter-industry mergers and acquisitions (M&A) activity between two industries. Building on these findings, we examine the role of inter-industry trade linkages in facilitating inter-industry asset disposals as a determinant of recovery rates of defaulted corporate bonds. For studying the relationships between recovery rates and inter-industry trade ties, we collect a set of U.S. corporate bonds which defaulted during the period 2001-2016. We construct annual networks of inter-industry trade from input-output tables provided by the U.S. Bureau of Economic Analysis (BEA). We then derive industry-specific network characteristics, which are more informative than the commonly used industry fixed-effects. First, we consider centrality measures (e.g., eigenvector centrality) to reflect an industry's degree of connectedness within the U.S. economy. Second, we develop a neighbor industry distress measure, a new measure that captures distress propagation between adjacent industries. In order to capture the inter-

[3] For example, Herskovic (2018) employs inter-industry trade network structures in order to explain stock returns. Gofman, Segal, and Wu (2020) link a firm's stock returns and exposure to aggregate productivity with their vertical position within the inter-industry supply chain. Evgeniou, Peress, Vermaelen, and Yue (2021) study the returns following share buybacks and insider trading as a function of firm centrality within the inter-industry network of trade.

industry asset transfer channel, we further consider firm- and industry-specific characteristics which are expected to co-determine the ability to transfer assets between industries together with centrality. Lastly, we observe actual inter-industry asset transfers in bankruptcy that validate the stylized facts established in our empirical analysis.

This paper contributes to three strands of the literature. The first and main contribution of the paper is investigating and depicting the importance of connections between industries and the position of a firm within the inter-industry trade network in explaining the recovery rate as one of the key parameters in credit risk. Unveiling the channel for asset disposals through inter-industry trade connections, our approach provides a new network-based explanation of recovery driven by asset transfer frictions between industries. Related to prior studies which explore the effects of economic linkages in asset pricing (see, for example, Herskovic (2018)), we study the effects of U.S. economy linkages in credit risk. Our results help to improve understanding the role of financial interconnectedness.

The paper's second contribution is that it bridges the inter-industry network and the business cycle, as well as industry distress for explaining the variation in recovery rates. This allows us to study, for the first time, the economic cycle's effect on the recovery rate of corporate bonds for central and non-central industries within the U.S. economy. Acharya, Bharath, and Srinivasan (2007) report that industry-wide distress causes a decrease in recovery rates. Unveiling the effect of industry distress on recovery rates in adjacent industries, this paper expands their work.

Third, in contrast to prior studies, we analyze recovery rate drivers more broadly by employing a comprehensive set of explanatory variables that includes information about inter-industry linkages, macroeconomic variables, as well as firm and bond-specific variables.

Through our analysis, we find that the centrality, or connectivity of a defaulted firm within the inter-industry network, is significantly driving the recovery rates of its bond issues. We show that the centrality's effect on recovery is operating jointly with frictions and drivers of inter-industry asset transfers, and that the connectedness between industries supports recovery due to facilitating inter-industry asset sales. By observing actual inter-industry asset transfers in bankruptcy, we validate these stylized facts. Moreover, we find that industry distress spills across industry borders through trade connections and has significant adverse effects on recoveries in closely connected neighbor industries.

We show that selected macroeconomic variables improve recovery rate prediction specifically in central industries. Our comprehensive analysis confirms the robustness of our findings to a variety of alternative explanations. Notably, we employ machine learning techniques that account for non-linear relationships between recovery rates and the extensive set of explanatory variables, and as our out-of-sample results show, network-derived characteristics can explain a similar order of magnitude of the variation in recovery rates as the drivers that are well-established in the literature, such as firm- and bond-specific variables and macroeconomic variables.

The remainder of this paper is organized as follows. Section 3.2 reviews related research on determinants of recovery rates and draws from research on financial networks in order to develop the hypotheses of this paper. In Section 3.3, we outline how we create the inter-industry trade network and introduce the centrality and neighbor industry distress measures. Section 3.4 provides details of the data used in our analysis, specifies our regression models, describes the main results of the network-related approach to recovery rate modeling, and conducts various robustness tests. Section 3.5 concludes the work.

3.2 Literature Review and Hypotheses

The empirical literature on corporate bond recovery rates has mostly omitted to study inter-industry relations, although industry fixed effects are commonly used to explain the variation in recovery rates. We borrow from the literature on inter-industry linkages from which we infer new hypotheses on recovery rate estimation. In this section, we review the related literature, discuss the research questions addressed, and explain the hypotheses tested in this study.

3.2.1 Dynamics of recovery rates

Historically, corporate bonds' recovery rates were commonly estimated at about 40% of bond par value. In the pioneering study of Altman and Kishore (1996), the foundations for a more differentiated approach to recovery rate estimation were set. Besides finding a positive effect of bond seniority on recovery, Altman and Kishore (1996) document for the first time in the literature corporate bond recovery rates' heterogeneity across industries. Controlling for the dependency on seniority, they show that recovery rates are driven by firms' industry affiliation. Business activities are industry-dependent, and hence determine firms' competitive environments, compositions of assets and their liquidity, consequently influencing debt recovery.

In subsequent studies, various accompanying determinants were uncovered in the growing breadth of research on recovery rates. Frye (2000) shows that first-generation credit models ignore the inverse relationship between the probability of default (PD) and recovery rate. Consistently, Varma and Cantor (2005) find a positive relationship between recovery rates and economic growth, in addition to demonstrating that the seniority and security class of a defaulted bond are among the key recovery rate drivers. Altman, Brady, Resti, and Sironi (2005) argue that aggregate recovery rates are a function of supply and demand for the defaulted securities. Their analysis on a data set of defaulted bonds over the period 1982-2002 shows that the supply of defaulted bonds and the size of the high-yield bond market explain a substantial fraction of recovery rates, regardless of seniority or collateral level.

Bruche and Gonzalez-Aguado (2010) analyze the effects of seniority classes, default events, credit cycles and GDP growth on recovery rates of corporate bonds. Their findings confirm that recovery rates of defaulted bonds decrease during recessions. By combining the approaches of macroeconomic and industry-specific explanations of recovery rates, Mora (2015) provides evidence that macroeconomic downturns operate differently at the industry level. Economy-wide distress affects certain industries' recovery rates more than others because economy-wide events potentially propagate to the industry level and induce fire-sales effects. Accordingly, if the overall economy declines, recovery rates drop much more in industries where sales growth is highly correlated to GDP. Nazemi and Fabozzi (2018) report that recovery rate models which include macroeconomic variables selected by the least absolute shrinkage and selection operator (LASSO) significantly outperform the models that incorporate only few macroeconomic variables.

Including bond and industry characteristics, firm-specific, and macroeconomic variables for modeling recovery rates, Chava, Stefanescu, and Turnbull (2011) report that industry factors and regime dynamics have more influence on the PD than on the recovery rate. Jankowitsch, Nagler, and Subrahmanyam (2014) provide a comprehensive study on the bond trading microstructure around default events, and additionally consider firm- and bond-specific characteristics, e.g. bond covenants and information from bond issuers' financial statements, as explanatory variables for their recovery rate models. Their study demonstrates significant explanatory power of a bond's trading liquidity for modeling its recovery rate. Furthermore, their findings confirm the importance of bond seniority and macroeconomic conditions, as well the industry in which a firm operates, for recovery rate estimation. Gambetti, Gauthier, and Vrins (2019) consider an economic uncertainty measure to explain recovery rates' systematic time variation.

Another group of variables for recovery rate prediction is employed by Donovan, Frankel, and Martin (2015), who use five different accounting conservatism measures to assess defaulted firms from Moody's Ultimate Recovery Rate Database (DRD) of the period 1994-2011. They discover that firms with less conservative financial accounting regimes prior to the default date tend to have lower recovery rates.

Complemented by an increasing number of factors for explaining corporate bond recovery rates, a firm's industry affiliation is commonly considered one of the key determinants. For example, Acharya, Bharath, and Srinivasan (2007) and Kim and Kung (2016) are in the spirit of Altman and Kishore (1996) who argue that industry affiliation matters as industries with more tangible and liquid assets will produce higher debt recoveries. Acharya, Bharath, and Srinivasan (2007) find that industry-specific economic factors have more explanatory power for recovery rates than the macroeconomic conditions have. In their study on defaulted loans and bonds over the period 1982-1999, they analyze the effects of industry-wide downturns on recovery rates. They report that recovery rates are not only depressed by poor macroeconomic conditions or reduced economic prospects as a result from industry-wide distress. Disposals of assets are a major tool to recover liquidity in both reorganizations (Chapter 11 bankruptcy) and liquidations (Chapter 7 bankruptcy). A firm which is forced to sell assets recovers less if its industry is illiquid and if its assets are industry-specific, such that those assets are not easily deployable in other industries. This is explained by the fire-sales effect described by

Shleifer and Vishny (1992): Industry-level distress reduces economic prospects of a company and its peers, and a sudden rise of supply of highly industry-specific assets combined with fewer potential would-be buyers results in low proceeds from fire-sales. Hence, industry-level distress and fire-sales in industries that are characterized by a high asset specificity lead to significantly lower recovery rates.

Kim and Kung (2016) measure the redeployability of assets for a broad cross-section of industries. They find that a high asset redeployability (i.e., the extent to which assets have alternative uses) implies higher liquidation values and thus higher recovery rates. Kermani and Ma (2023) study industry-wide liquidation values of assets in Chapter 11 filings and show that industry-specific asset characteristics determine the receipts in asset sales. However, they do not apply their findings for explaining the recovery of defaulted debt securities. While those industry-dependent variables proxy for heterogeneity of assets across industries and capture more industry-related information than what is already included in industry fixed-effects, the interconnections between industries are omitted. We expand the research on the asset disposal channel as a driver of recovery rates building on the tool of inter-industry linkages.

3.2.2 Inter-industry trade network and recovery rate

In recent years, a growing number of financial networks have been applied in academic research.[4] Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) provide evidence that microeconomic idiosyncratic shocks can aggregate into macroeconomic fluctuations through intersectoral input-output linkages. Using input-output data of 471 industries provided by the BEA, Ahern and Harford (2014) create a network model of the U.S. economy by linking industries through supplier and customer relationships. With information about the intensity of trade relationships between different industries, they study cross-industry acquisitions and discover evidence that stronger inter-industry trade relationships with customers or suppliers increase the likelihood of acquisitions across industries. They show that M&A activity propagates in wave-like patterns through the customer-supplier industry network. By modeling customer-supplier relationships between industries, Ahern (2013) finds that central industries earn higher stock returns as

[4] Besides the inter-industry network, other financial networks were studied, for example, by Bajo, Chemmanur, Simonyan, and Tehranian (2016), Di Maggio, Kermani, and Song (2017), and Rossi, Blake, Timmermann, Tonks, and Wermers (2018).

compensation for higher exposure to sectoral shock events.[5] Aobdia, Caskey, and Ozel (2014) provide evidence that firms' returns in more centrally located industries are more dependent on aggregate risk than those in peripheral industries and that not only the connections between two industries, but also the position of an industry within the network affects the transfer of information and economic shocks within the network. Until today, no study examines the impact of inter-industry linkages in the context of post-default recovery, although such a relationship seems reasonable given the evidence that inter-industry trade ties determine the transfer of information, and assets through M&A across industry borders.

Hotchkiss and Mooradian (1998) examine asset redeployment in bankruptcy and observe that asymmetric information prevents industry outsiders from bidding for bankrupt firms. Bidders usually operate in the same industry as the defaulted firm, or in related industries and so possess information and expertise about the best use of the disposal assets. Shleifer and Vishny (1992) argue that when the industry peers of bankrupt firms are financially constrained, assets need to be sold to industry outsiders. Correspondingly, Bernstein, Colonnelli, and Iverson (2018) find that asset allocation efficiency in bankruptcy suffers when search frictions are high or when only few alternative users for the assets are available. They report that assets are typically disposed to firms within the same industry after default, and that the transfer of plants to other industries is more likely in liquidation cases than in reorganization cases. The approaches of Hotchkiss and Mooradian (1998), Shleifer and Vishny (1992) and Bernstein, Colonnelli, and Iverson (2018) to defining industry insiders and outsiders rely on the classifications of economic activities, rather than accounting for existing relationships between industries.[6] Relatedly, Strömberg (2000) demonstrates that asset sales to industry outsiders yield lower valuations, an effect which magnifies with asset specificity. Gavazza (2011) argues that capital equipment is typically specialized by industry and so has greater value within the defaulted firm's industry.

Whereas these studies consider industries' similarity of economic activities, captured by industry classifications, they neglect the imminent conjecture that also the position of a firm within the supply chain, inferred from actual inter-industry trade relationships, may determine

[5] Relatedly, Herskovic (2018) employs a production network structure in order to successfully explore asset pricing implications in stock returns.

[6] Hotchkiss and Mooradian (1998) consider the first three Standard Industrial Classification (SIC) code digits and Bernstein, Colonnelli, and Iverson (2018) consider the first three North American Industry Classification System (NAICS) digits for defining industry affiliation.

the economics underlying asset redeployment. Kim and Kung (2016) show that when assets are customized to an industry's needs, recovery rates will be lower, as they cannot be easily transferred to other industries. Relatedly, Acharya, Bharath, and Srinivasan (2007) document that coincidence of asset specificity and industry-wide distress has a negative directional effect on recoveries due to the fire-sales effect within the defaulted firm's industry. However, it remains unanswered under which circumstances asset transfers to other industries do actually take place. By intuition, when assets are not highly specified to a single industry's needs, also industry outsiders should be capable to utilize these assets.

Ahern and Harford (2014) present evidence that strong inter-industry trade connections increase the likelihood of M&A activity between two industries. They further find that asset complementarity between industries, another contributing factor, cannot be fully explained by industry classifications. Because trade ties between industries were identified in the literature to serve as channels for M&A activity, besides the transmission of information and economic shocks, we hypothesize a yet undocumented causal role of the inter-industry network of trade in facilitating the transfer of assets across industry borders in default events. This network-based approach goes beyond what is already captured by established industry-specific variables for the formation of recovery rates, as it shifts the focus from intra-industry asset redeployment in bankruptcy based on similar economic activities to the role of inter-industry asset transfers. Hereby, we expand the asset liquidity explanation of Altman and Kishore (1996) as one of the main sources of heterogeneity in recovery rates across industries. Ultimately, we desire to capture the effects of the position of a firm within the inter-industry trade network on recovery rates.

As industries are characterized by different production technologies and perform a variety of distinct value-generating activities, asset transfers between industries are impeded by asset specificity and asymmetric information about these assets. Consequently, we expect that issuers of defaulted bonds positioned in well-connected, centrally located industries may be less exposed to these asset transfer frictions thanks to the capability of inter-industry trade relations to facilitate cross-industry acquisitions.[7] In particular, the extent to which buyers from other

[7] Instead of firm-to-firm trade relationships, we consider industry-to-industry trade relationships. Using firm-to-firm relationships may introduce selection bias, as only realized outcomes, but not potential outcomes of a similar likelihood, would be considered. For that reason, we use inter-industry connections analogous to Ahern

industries require compensating asset price discounts is therefore expected to decrease for the more centrally located and better connected industries. Hence, central firms may be able to draw on a greater number and more diverse universe of potential buyers for disposal assets, implying higher recoveries. If central industries benefit from a greater ability to transfer assets across industry borders, we further expect that such an effect would be magnified in industries which use assets that are more standardized and may find application in a variety of industries compared to assets that are highly specialized. To test this hypothesis, we examine the effects of several asset characteristics, such as asset specificity and asset redeployability, on recovery rates and their interaction with centrality. We explore the emergence of recovery rates along with these aspects from various perspectives.

We take up the notion of industry distress which has proven the dependence of recovery rates from industries' business cycles. We hypothesize that if inter-industry trade ties serve as a channel for assets and economic shocks, the closely connected industries' business cycles should have directional effects that spill through industry trade relationships. Therefore, defaulted bonds of issuers whose directly connected industry neighbors experience economic distress at the time of default are expected to recover less than those of issuers whose industry neighbors don't experience economic distress. We test this hypothesis by developing a new neighbor industry distress measure and including it in our analysis.

Based upon the findings of Aobdia, Caskey, and Ozel (2014), we expect that bond's recoveries of firms in central industries are more related to macroeconomic conditions due to strong connections to the overall economy.[8] Our data allows us to test whether recoveries of defaulted bonds of central issuers depend to a greater degree on macroeconomic conditions than those of non-central issuers. This paper is the first to analyze how the specifications of economic linkages and firms' positions within the inter-industry trade network determine corporate bond recovery rates.

and Harford (2014), who argue that inter-industry relationships, besides avoiding selection bias, also account for (higher order) inter-industry dependencies. Moreover, Hotchkiss and Mooradian (1998) note that acquirers from outsider industries do not need to have any direct trade relationships with a target firm, as also relationships with other firms within the target's industry enable to obtain the knowledge required to properly utilize the acquired assets.

[8] Aobdia, Caskey, and Ozel (2014) show that the performances of firms in more centrally located industries are more closely related to macroeconomic conditions as macroeconomic shocks are more likely to originate in central industries and propagate to closely connected industries.

3.3 Empirical Implementation of the Inter-Industry Network

In this section, we describe our methodologies for creating network-derived variables which characterize the U.S. economy network. We first construct annual networks of inter-industry U.S. dollar flows which describe trade relations between industries using data provided by the BEA, yielding a complete network representation of the U.S. economy. To do so, we closely follow the methodologies of Ahern (2013) and Ahern and Harford (2014). Similar to Evgeniou, Peress, Vermaelen, and Yue (2021), we exclude sectors that are not related to the objective of our study, such as households, government, capital, and foreign sectors.[9] As the data provided by the BEA are available with different level of granularity, we consider both networks based on 67 and 471 unique private sector industries.[10] A detailed description of how the network is created can be found in Appendix B.1. We calculate a number of centrality measures characterizing the previously defined network structures, and further develop a measure which captures economic distress spillover between adjacent industries. The proposed centrality measures include eigenvector centrality, closeness centrality, degree centrality, and betweenness centrality. A discussion of the underlying centrality concepts from network theory can be found in Appendix B.2. Due to skewness, we consider the logarithm of centrality and then we match all the created measures of year $t-1$ to the related bonds defaulted in year t to ensure that the most recent network structure prior to default is used in the predictions.

A list of the 15 most central and 15 least central industries based on eigenvector centrality for the year 2001 can be found in Table B.2 in the Appendix.[11] In general, the centrality ranking order of industries remains stable over time, with few exceptions. After the GFC, industries' centralities slightly declined in 2009 and 2010, meaning that connectivity between

[9] Although excluding households, government, capital, and foreign sectors is economically meaningful, our main findings remain significant when including these sectors.

[10] While we use detailed industry definitions involving 471 unique private sector industries for robustness checks in Section 3.4.3, we utilize 67 unique three-digit NAICS industries modified by the BEA in our main analysis, as the chosen level of detail must be consistent with other industry-specific explanatory variables in our main analysis, in order to evaluate the co-determination of recovery rates by these variables together with network-derived characteristics via interaction effects. For example, the data for replicating the asset redeployability measures of Kim and Kung (2016) and asset specificity measures of Kermani and Ma (2023) are not available at a more detailed level.

[11] We discuss the choice of eigenvector centrality as the appropriate measure of industry centrality for estimating corporate bond recovery rates in Section 3.4.2

industries generally decreased within the U.S. economy. This might be due to reduced business opportunities in the U.S. inter-industry trade network as a whole, leading to fewer or weaker trade relationships between industries. In subsequent years, as the economy stabilizes, the effect reverses and centralities gradually return to higher levels after 2010, however at a slow pace.

Taking up the notion that asset transfers between well-connected industries are more likely than between unrelated industries, we extend the concept of industry distress to a network-based approach by deriving a newly developed neighbor industry distress measure from the U.S. economy network representation, and apply it to recovery rate modeling for the first time in the literature. Therefore, we consider each industry's ten most important neighbor industries by trade volume on an annual basis, taking into account both trade volumes through customer and supplier relationships. We define the trade volume with these most important neighbor industries that are in distress as the neighbor industry distress measure, calculated as a percentage of total trade volume in a given year. Here, an industry is considered in distress when it has experienced negative sales growth in the preceding year. Thus, the newly developed neighbor industry distress measure captures both the magnitude and the concentration of an industry's trade volume with distressed neighbor industries.[12] It is designed to capture distress spillover effects between adjacent industries in the economy network structure.

We also include the labor's fraction of industries' inputs (i.e. the degree to which an industry relies on human labor) to account for effects of the dependency on human labor on the ability of companies to recover from default. We derive labor's fraction of inputs from the inter-industry network before removing the households sector by dividing labor inputs (i.e., dollar flows from industries to households) by the total sum of inputs that an industry consumes. In the Table B.3 of the Appendix, we provide descriptive statistics of the network-derived variables.

For robustness checks, we introduce alternative networks that are based on inter-industry M&A transactions instead of trade relationships. Although Ahern and Harford (2014) document a high similarity between the inter-industry trade network and the inter-industry merger network, considering only realized mergers and acquisitions to form the network may introduce

[12] Due to skewness, we consider the logarithm of neighbor industry distress. Alternative definitions of the neighbor industry distress measure, such as the Herfindahl index of distressed neighbor industries or the total percentage of trade volume with distressed industries yields similar significant results.

selection bias, whereas Ahern and Harford (2014) show that inter-industry trade connections serve as predictors of inter-industry acquisitions. Nevertheless, we check the robustness of industry centrality in ex-post network structures created from realized acquisitions of assets and firms across industry borders. We retrieve two M&A transaction datasets from S&P Capital IQ containing data on U.S. M&A share and asset deals over the period 2001–2016, corresponding to our data of defaulted bonds. The transactions in both datasets contain disposals of firms, business units, sites, properties, and other firm assets. The first dataset contains 11,320 closed M&A transactions with reported transaction values of at least $100 million per transaction, reflecting a total cumulative transaction volume in excess of $12 trillion. The second dataset contains 4,659 closed M&A transactions denoted as bankruptcy sales of any size. Transaction values are reported for about two thirds of the bankruptcy M&A transactions, which add up to a cumulative bankruptcy transaction volume of more than $170 billion.

With the information on inter-industry bankruptcy and non-bankruptcy M&A transactions, we now form two alternative network representations of the U.S. economy. While we use the same industry definitions as for the trade network, we do not calculate annual networks. Instead, we create networks that contain the full period 2001-2016 due to sparsity of data.[13] We weight inter-industry links by the number of transactions instead of transaction values for two reasons. First, in the case of the bankruptcy M&A network, only about two thirds of the transaction values are reported. Second, there exist large outlier transactions in terms of transaction value. After creating the networks, we calculate eigenvector centrality for each industry in a similar fashion as for the industry network of trade. Descriptive statistics of centrality are reported in Table B.3 of the Appendix.

3.4 Empirical Study of Recovery Rate Modeling

In this section, we describe our dataset of defaulted bonds and the explanatory variables such as macroeconomic variables and firm- and bond-specific variables, as well as the control variables for checking robustness. We then create linear recovery rate models incorporating these vari-

[13] Constructing annual networks would yield sparse networks that involve merely about 700 transactions per year in the case of the non-bankruptcy M&A network, and even less so for the bankruptcy M&A network. Building networks that involve the whole time period allows us to meaningfully populate the networks of 67 distinct industries and yield informative network representations of inter-industry M&A transactions.

ables together with network-derived characteristics to shed light on their economic mechanisms from various perspectives. Subsequently, we apply various robustness tests which allow us to generalize our findings.

3.4.1 Data and descriptive statistics

We collect a universe of bonds which defaulted over the period 2001 to 2016 from S&P Capital IQ. After removing all bonds for which no price data are available, the dataset contains a total of 2,127 bonds from 603 issuers that defaulted under Chapter 7 (liquidations) or Chapter 11 (both reorganizations and liquidations) bankruptcies during the period 2001-2016. We remove four bonds from the dataset because of corrupted data. We further exclude from our sample 644 bonds of companies which had a lot of subsidiaries with different collateral, such as Lehman Brothers. Thus, the final dataset used comprises 1,479 defaulted bonds. The bonds' recovery rates are measured in observed bond prices 30 days after default as percentage of par value. The bond prices are obtained via S&P Capital IQ from the Intercontinental Exchange (ICE) and represent dealer quotes, live trading levels and data of executed trades from the Trade Reporting and Compliance Engine (TRACE), when available. The bond price 30 days after default has been established as the standard measure of recovery rate in the literature, as it represents actual recovery for bond investors who sell their bond holdings subsequent to a default event (see, for example, Mora (2015)).[14] Recovery rates in the dataset spike between 60% and 70%, and for the vast majority of bonds only 30% or less is recovered. The average recovery rate within the final dataset is 43.19% with a standard deviation of 32.50%. The empirical distribution of the recovery rates can be found in Figure B.1 in the Appendix.

We consider various explanatory variables in our recovery rate models that we retrieve from different sources. We ensure that all variables represent the last available observation before each bond's default date in order to only incorporate information explaining a state or condition prior to the recovery rate formation. We also scale all explanatory variables on the interval [0,1] for the regressions. The dataset includes senior secured, senior unsecured, senior subordinated, subordinated, and junior subordinated bonds. We introduce dummy variables

[14] Other recovery rate databases, such as Moody's Default and Recovery Database (DRD), also consider 30-day post-default bond prices as the representation of recovery rate.

Table 3.1: Descriptive statistics of recovery rates (RR) by seniority.

Seniority	**Mean RR**	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	σ	**# of bonds**	**% of bonds**
Subordinated	11.4%	2.8%	4.3%	18.0%	12.5%	15	1.0%
Senior Subordinated	24.1%	2.3%	15.5%	36.0%	28.2%	158	10.7%
Senior Unsecured	43.1%	12.0%	42.0%	68.1%	30.9%	1,112	75.2%
Senior Secured	61.5%	30.1%	70.0%	94.5%	34.8%	194	13.1%
Total	43.2%	11.5%	40.5%	68.1%	32.5%	1,479	100%

for these seniority classes, and we combine subordinated and junior subordinated bonds in the subordinated seniority class to avoid seniority classes that are too small. As one would expect, more senior bonds recover the most after default. The average recovery rates per seniority class drops monotonically in decreasing seniority. Senior secured bonds have an average recovery rate of 61.5%, followed by senior unsecured bonds with an average recovery rate of 43.1%. Bonds within the senior subordinated class have an average recovery rate of 24.1%, and bonds in the subordinated class have the lowest of 11.4%. Table 3.1 provides the summary statistics of recovery rates across seniorities in our dataset.

For relevant firm- and bond-specific variables, we follow the recent literature which has previously included such information for recovery rate modeling, e.g., Acharya, Bharath, and Srinivasan (2007), Chava, Stefanescu, and Turnbull (2011), Khieu, Mullineaux, and Yi (2012), Jankowitsch, Nagler, and Subrahmanyam (2014) and Donovan, Frankel, and Martin (2015). The data includes information from profit and loss statements, as well as balance sheet information from financial statements which were filed prior to default. Ahern (2013) reports a positive relationship between industry centrality and firm size. In order to control for this, we use the value of total assets of a firm as a proxy. We collect asset values and other pre-default information from financial statements from S&P Capital IQ. We also replicate accounting ratios to capture structural credit risk, such as long-term debt (LTD) issuance and default barrier, in a similar fashion as Jankowitsch, Nagler, and Subrahmanyam (2014).

We add macroeconomic variables in order to account for macroeconomic effects on recovery rates. Therefore, we collect a comprehensive set of 179 macroeconomic variables, including variables that were used in previous research such as Varma and Cantor (2005), Acharya, Bharath, and Srinivasan (2007), Acharya, Bharath, and Srinivasan (2007), Jankowitsch, Nagler, and Subrahmanyam (2014), Mora (2015), and Nazemi and Fabozzi (2018). We obtain macroeconomic variables from the Federal Reserve Bank of St. Louis (FRED, Federal Reserve Economic Data), except for the high-yield market size and high-yield default rate, which are obtained from Fitch Ratings. As LASSO-selected macroeconomic variables can be highly unstable due to small perturbations of the data, we use a stability selection technique in the macroeconomic variable selection for recovery rate prediction as described in Appendix B.3. In the Appendix, we also provide an overview (Table B.4) and descriptive statistics (Table B.5) of the macroeconomic variables.

Finally, we match NAICS industry codes to each of the issuers of defaulted bonds. We do this by retrieving industry classifications from S&P Capital IQ, or otherwise carefully matching by hand those where no industry classification is available.

Furthermore, we include the number of firms for each NAICS industry in order to control for industry size. The number of firms in an industry may affect recovery rates in two ways: (i) an industry with a high number of firms with similar economic activities and (ii) dependence on similar assets is expected to facilitate assets sales after the default event and also create more competition for these assets yielding higher liquidation proceeds. However, an adverse effect may take place if the whole industry is in distress: When several competitors try to liquidate similar assets, the supply of disposal assets may exceed demand and disposal proceeds diminish. As an alternative proxy of industry size, we consider industries' total amounts of assets, which we retrieve from the BEA's fixed assets tables. We retrieve the number of firms for each NAICS industry from the U.S. Census. After matching bonds to NAICS industries, we are able to also assign centrality measures and other network-derived variables described in Section 3.3 to the defaulted bonds.

We also include industry dummy variables and industry distress measures at the sector level to incorporate basic industry-related characteristics as employed by Acharya, Bharath, and Srinivasan (2007). Industry distress is measured by two dummy variables capturing (i) negative return shocks of -30% or less to industry-specific stock indices and (ii) industry-wide negative sales growth during the last 12 months prior to default. We rely on Bloomberg for industry-specific sales and stock index data.

In order to check for the robustness of our empirical results to potential alternative explanations, we determine which Chapter 11 bankruptcies eventually result in liquidation based on data from S&P Capital IQ. We further collect a variety of alternative measures that account for asset heterogeneity across industries to control for the alternative explanation that industry centrality just represents the different mix of assets used as collateral for defaulted debt and to examine the co-determination effect of asset characteristics together with industry centrality.

We collect annual industry aggregate data from the Statistics of Income (SOI) provided by the U.S. Internal Revenue Service (IRS) and fixed assets tables provided by the BEA. With the data, we create the inverse of an industry-wide quick ratio to measure the financial illiquidity within a given industry, and an asset specificity measure similar to Acharya, Bharath, and Srinivasan (2007). We account for the non-redeployability of assets by taking the inverse of the measures described by Kim and Kung (2016). We replicate their redeployability measures with data collected from capital flow tables provided by the BEA. After inverting, these measures are expected to have a similar directional effect on recovery rates as asset specificity.

We account for additional dimensions of industry-specific asset heterogeneity measured by asset mobility, asset durability and asset customization as described by Kermani and Ma (2023). For creating these measures, which are designed to explain aggregate industry asset liquidation values, we use annual data from fixed assets tables, input-output tables, and private fixed investment in equipment (PEQ) bridge tables provided by the BEA. Similar to the industry definitions used for creating the inter-industry network, we translate adjusted industry definitions to NAICS codes via concordance tables provided by the BEA.

3.4.2 Regression models explaining recovery rates

In this subsection, we provide the results of several regression models involving the factors introduced in Section 3.4.1 for explaining corporate bond recovery rates through inter-industry linkages within the economy. In particular, we examine the hypothesized asset transfer channel across industry borders through trade relationships as a driver for corporate bond recovery.

Recovery rate and network centrality

To examine the effects of industry centrality and neighbor industry distress on the recovery rates of defaulted corporate bonds, we utilize the following linear regression specification to estimate recovery rates over the period 2001-2016:

$$\begin{aligned}\text{Recovery rate} = \alpha + \beta(&\text{bond seniority and industry characteristics})\\ &+\mu(\text{firm- and bond-specific variables}) + \delta(\text{network-derived variables}) + \epsilon,\end{aligned} \tag{3.1}$$

The basic specification of our linear regression includes dummy variables for bond seniority classes, industry affiliation, industry distress, and other commonly used firm- and bond-specific variables. As shown in Model (1) in Table 3.2, this specification can explain more than 44% of the variation in creditor recoveries, and we consider it as the benchmark. In all of our OLS regressions, we adjust standard errors for heteroscedasticity as proposed by White (1980) and firm-level clustering as described by Wooldridge (2002). This allows us to avoid biased estimates of standard errors from within-cluster correlation, which is most likely to occur when multiple bonds of a single firm default. We find that the amount issued, days to maturity, coupon rate, default barrier, long-term-debt (LTD) issuance and intangibility are significantly negatively, and profitability is significantly positively related to recovery rate. While our observations are generally in line with Jankowitsch, Nagler, and Subrahmanyam (2014), we find an adverse effect of the coupon rate, contradicting the assumption of Jankowitsch, Nagler, and Subrahmanyam (2014) that bonds with a higher agreed coupon may still be likely to pay out more after default. Firm leverage has the expected and significant negative effect on recoveries, consistent to models employed by Mora (2015). When adding centrality measures in Models (2)–(6), most effects of the firm- and bond-specific variables remain intact. Only the amount of equity, default barrier, and intangibility cease to be significant in several of the specifications.

Among the network-derived variables, centrality represents the position of the defaulted firm within the network of inter-industry trade and we measure it either by eigenvector centrality, closeness centrality, degree centrality, and betweenness centrality. We further consider the first principal component of the significant centrality measures. Comparable to the studies of Ahern and Harford (2014) and Evgeniou, Peress, Vermaelen, and Yue (2021), eigenvector and degree centrality are most likely to capture the hypothesized asset transfer channel, as inter-industry assets transfers are expected to be path-independent, similar to economic shocks. Moreover, eigenvector and degree centrality more strongly account for the connectedness to the immediate periphery of an industry. This corresponds intuitively to that if asset transfers to industry outsiders are affected by asset transfer frictions between industries as outlined by Hotchkiss and Mooradian (1998), these frictions should rise for acquirers located in distant industries. While degree centrality does not account for any other connections than those to direct neighbors, eigenvector centrality further considers the importance of neighbor industries and hence has the advantage that also close connections beyond direct neighbors are reflected. As closeness centrality and betweenness centrality are path-dependent, based upon shortest paths between all industries, we do not expect these measures to offer an explanation to recovery rates. The other network-derived variables include neighbor industry distress, labor's fraction of inputs, and log(number of firms).

Models (2) – (6) in Table 3.2 present the results of the regression specifications, adding the network-derived variables to the basic specification and alternating the different types of centrality measures. The centralities of the defaulted firms' industries in the network of inter-industry trade are measured by eigenvector centrality in Model (2), closeness centrality in Model (3), degree centrality in Model (4), betweenness centrality in Model (5), and the first principal component of the significant centrality measures in Model (6).

We find that eigenvector centrality, degree centrality, and also their principal component contribute positively to the recovery rates with significance at the 1% level. The positive coefficients of the significant centrality measures provide interesting implications in general, indicating that, in line with our expectation, bonds issued by firms in central industries recover more than bonds issued by firms in non-central industries. Moreover, the R-squareds of our recovery rate models improve by about 5 percentage points when the suggested new variables are added to the basic model. Since eigenvector centrality also reflects close connections to industries that are not direct neighbors in the network, we base our further analysis on this measure.[15]

The newly developed neighbor industry distress measure is significant at the 1% level in all models and it shows that industry-wide distress has a negative effect on recovery rates across industry borders. Finding that both centrality and the neighbor industry distress measures have sizeable coefficients strongly supports the notion that recovery is considerably dependent

[15] Our empirical study yields similar results if we consider degree centrality instead of eigenvector centrality.

Table 3.2: Recovery rate and centrality, OLS regression results. The recovery rate is the dependent variable. All regressions include seniority, industry, industry distress dummy variables, and firm- and bond-specific variables. Model (1) is the basic specification including bond- and firm-specific information. Models (2)–(5) add network-derived variables and iterate different centrality measures. Model (6) includes the principal component of the significant centrality measures. Standard errors are adjusted for heteroscedasticity and clusters at the firm level. Statistical significance at the 1%, 5%, and 10% level is indicated with ***, **, and *.

Model	**(1)**		**(2)**		**(3)**		**(4)**		**(5)**		**(6)**	
Centrality measure	**None**		**Eigenvector**		**Closeness**		**Degree**		**Betweenness**		**Principal Component**	
Intercept	0.3483		0.5569	**	0.6470	**	0.5824	**	0.5288	**	0.8708	***
Amount issued	-0.3250	***	-0.2921	**	-0.3098	**	-0.2811	**	-0.3176	***	-0.2859	**
Days to maturity	-0.5200	***	-0.4513	***	-0.4616	***	-0.4450	***	-0.4521	***	-0.4483	***
Coupon rate	-0.3127	***	-0.2665	***	-0.2830	***	-0.2683	***	-0.2896	***	-0.2668	***
Total equity	0.2172	**	0.0198		0.2231		-0.0207		0.2387		-0.0056	
Default barrier	-0.2253	**	-0.1709	*	-0.1152		-0.1717	*	-0.1383		-0.1721	*
LTD issuance	-0.1617	***	-0.1322	***	-0.1250	***	-0.1326	***	-0.1362	***	-0.1322	***
Profitability	0.2364	***	0.2895	***	0.2648	***	0.2894	***	0.2284	***	0.2909	***
Intangibility	-0.1413	**	-0.1076		-0.1100	*	-0.1025		-0.1041		-0.1051	
Receivables	-0.1882		-0.0871		-0.1225		-0.0757		-0.1211		-0.0804	
Total assets	-0.0243		0.0701		0.0791		0.0540		0.0888		0.0615	
Leverage	-0.3838	***	-0.3203	***	-0.3833	***	-0.3162	***	-0.3918	***	-0.3166	***
Log(number of firms)			-0.4795	***	-0.2689	**	-0.5076	***	-0.2489	*	-0.4994	***
Labor's fraction of inputs			-0.1487		0.0260		-0.1880		0.0024		-0.1719	
Log(neighbor industry distress)			-0.2789	***	-0.2853	***	-0.2788	***	-0.2933	***	-0.2784	***
Log(eigenvector centrality)			0.4034	***								
Log(closeness centrality)					-0.2347							
Log(degree centrality)							0.4437	***				
Log(betweenness centrality)									0.0775			
Principal component of centralities											0.3058	***
R^2	0.4465		0.4948		0.4845		0.4976		0.4824		0.4965	
Adj. R^2	0.4365		0.4843		0.4738		0.4872		0.4716		0.4861	
RMSE	0.2441		0.2335		0.2359		0.2329		0.2364		0.2331	
Observations	1,479		1,479		1,479		1,479		1,479		1,479	
Industry dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Seniority dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Basic industry distress dummies	Yes		Yes		Yes		Yes		Yes		Yes	

on the structure of the inter-industry trade network. We further find that the number of firms within an industry, a variable to control for industry size, is significant with a negative effect on recovery. Labor's fraction of inputs measures industries' relative dependence on human labor and controls for the case that productive assets and asset sales for recovering liquidity play no pivotal role in labor-intensive industries. However, it is insignificant in all models.

Recovery rate, network centrality and the asset transfer channel

To rule out that the significant effects from network-derived variables and the centrality measure can be explained with asset characteristics or proxies for industry-specific asset heterogeneity, we further investigate such alternative explanations. Acharya, Bharath, and Srinivasan (2007) find that the asset specificity, i.e. the inability to redeploy assets of a defaulted firm in other industries, negatively impacts recovery during times of industry-wide distress. Following Berger, Ofek, and Swary (1996) and Strömberg (2000), they define asset specificity as book value of machinery and equipment divided by the book value of total assets. This definition is distinct from the total amount of assets as it excludes non-specific and fungible assets such as cash or property, which can easily find application in other industries.

Kim and Kung (2016) develop two measures which account for industry-specific heterogeneity of assets and the redeployability of assets to other industries, reflecting the number of industries that use specific asset types. They show that when assets can be easily redeployed in other industries, recovery rates are higher.[16]

Kermani and Ma (2023) define asset mobility, asset durability, and asset customization to infer pricing implications on assets liquidations. Their measures capture the unique physical attributes of assets that are related to industry-specific business activities, and hence reflect the productivity of assets for alternative users, as well as the ability to transfer these assets to alternative users. Asset mobility is measured as one minus transportation costs to total production costs of the typical PPE utilized within an industry. Assets with lower transportation costs (e.g., vehicles) are easier to deploy to new physical locations. Asset durability, measured as one minus the industry-wide average asset depreciation rate, reflects that faster depreciated assets may be less valuable to alternative users (e.g., computers or office equipment). Asset customization reflects the degree to which the assets that are used in a given industry are customized to the industry's requirements. Customized assets may have only limited use in other industries (e.g., optical lenses for industrial production). For our analysis, we rebuild these measures and include them in linear regression to account for asset heterogeneity that could be captured by our network-derived measures.[17]

We emulate asset specificity, the two non-redeployability variations, as well as asset mobility, asset durability, and asset customization, and investigate their main effects in our linear recovery

[16] Kim and Kung (2016) provide two interchangeable asset redeployability measures, the first of which is related to specificity of assets and the second is related to market thickness. The former variation is scaled by the number of industries to capture asset specificity, and the other is scaled by the number of firms in each industry to capture the liquidity dimension. See Kim and Kung (2016) for a detailed explanation of the construction of asset redeployability measures.

[17] Kermani and Ma (2023) use asset mobility, asset durability and asset customization to measure aggregate recoveries for two-digit SIC industries. The recovery in their study represents the liquidation receipts as a percentage of book value, from which no direct inference on the recoveries of defaulted debt instruments on the firm- and bond-specific level are made. See Kermani and Ma (2023) for a detailed explanation of the construction of the asset mobility, asset durability and asset customization measures.

rate models in combination with the new network-derived variables. We further add a variable that controls for industry-wide illiquidity, as Acharya, Bharath, and Srinivasan (2007) show that industry-wide illiquidity negatively impacts recovery rates when asset specificity is high. In addition, we consider the total amount of assets per industry as an alternative measure of industry size compared to the number of firms, and a dummy variable to capture liquidation-type defaults.

In all specifications in Table 3.3, the centrality measure and neighbor industry distress remain significant, and most notably, the coefficients maintain similar signs and a comparable order of magnitude as what we report in the previous section. Model (1) shows that the specificity of assets has a significant effect on recovery with the expected negative sign, given the limited use of highly specific assets in other industries.[18] We find a significant negative effect at the 5% level of one of the non-redeployability measures in Model (2), which is in line with what Kim and Kung (2016) find. However, the second variant of non-redeployability is not significant in Model (3).[19] Furthermore, among the measures that account for industry-specific physical asset attributes as described by Kermani and Ma (2023), asset customization (Model (6)) is significantly related to recovery rates at the 5% level. Kermani and Ma (2023) report lower liquidation values for customized assets, which corresponds to the negative effect on recovery rates that we find. We further find in Model (7) that the average illiquidity within an industry has a significant negative effect on recovery rates, which is compatible with the findings of Acharya, Bharath, and Srinivasan (2007), although they do not report the main effect of illiquidity separately. The amount of assets within an industry (Model (8)) as a proxy for industry size has a negative coefficient, but is not significant.

We additionally examine whether forced firm liquidations may coincide with centrality and distort our findings. An important difference between reorganization (Chapter 11 bankruptcies) and liquidation (Chapter 7 bankruptcies) is that under Chapter 11 the liquidation of assets is negotiated, but not mandatory as compared to Chapter 7. Nevertheless, Chapter 11 cases may eventually be converted to Chapter 7 cases, resulting in the liquidation of all assets. Bris,

[18] Consistent with Acharya, Bharath, and Srinivasan (2007), but unrelated to our research objective and hence not reported, we find that the interaction between asset specificity and industry distress has a significant negative effect on recovery rates.

[19] If we exclude industry fixed effects, similar to Kermani and Ma (2023), both non-redeployability measure are significant, with the expected negative effect on recovery.

Table 3.3: Recovery rate and centrality – Alternative control variables, OLS regression results. The recovery rate is the dependent variable. All regressions include seniority, industry, industry distress dummy variables, and firm- and bond-specific variables. Models (1)–(9) iterate alternative control variables (asset-related firm- and industry-specific characteristics) together with network-derived variables. Standard errors are adjusted for heteroscedasticity and clusters at the firm level. Statistical significance at the 1%, 5%, and 10% level is indicated with ***, **, and *.

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Intercept	1.0176 ***	0.5432 **	0.8056 ***	0.6862 **	0.8124 ***	0.8279 ***	0.8909 ***	0.8104 ***	0.8730 ***
Log(number of firms)	-0.5162 ***	-0.4006 ***	-0.4791 ***	-0.4163 ***	-0.4804 ***	-0.2944 *	-0.3991 **	-0.4722 ***	-0.4135 ***
Labor's fraction of inputs	-0.1969	-0.1159	-0.1501	-0.1210	-0.1482	0.0575	-0.2822	-0.1438	-0.0050
Log(neighbor industry distress)	-0.2527 ***	-0.2655 ***	-0.2791 ***	-0.2793 ***	-0.2787 ***	-0.2662 ***	-0.2762 ***	-0.2771 ***	-0.2892 ***
Log(eigenvector centrality)	0.3661 **	0.3132 **	0.4056 ***	0.3300 **	0.4041 **	0.2790 *	0.3267 **	0.3984 **	0.3638 **
Asset specificity	-0.2864 *								
Asset non-redeployability 1		-0.2209 **							
Asset non-redeployability 2			0.0147						
Asset mobility				0.2226					
Asset durability					-0.0042				
Asset customization						-0.4744 **			
Industry illiquidity							-0.5417 *		
Total amount of industry assets								-0.0207	
Liquidation dummy variable									-0.1900 ***
R^2	0.5051	0.4997	0.4948	0.4965	0.4948	0.5055	0.5009	0.4948	0.5153
Adj. R^2	0.4945	0.4889	0.4840	0.4857	0.4839	0.4949	0.4902	0.4840	0.5049
RMSE	0.2312	0.2325	0.2336	0.2332	0.2336	0.2311	0.2322	0.2336	0.2288
Observations	1,479	1,479	1,479	1,479	1,479	1,479	1,479	1,479	1,479
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seniority dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm- and bond-specific variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Welch, and Zhu (2006) find that recovery rates are higher in Chapter 11 cases due to better preservation of the firm's assets. Comparably, Bernstein, Colonnelli, and Iverson (2018) report inefficient asset allocations for liquidation cases. We employ a liquidation dummy variable which captures whether liquidation has eventually been implemented regardless of how the bankruptcy case had initially been filed. Consistent with the findings reported in previous studies, we find a significant negative effect of liquidation on recovery in Model (9) in Table 3.3. While the liquidation measure serves well for the verification of our regression results, it is worth noting that it cannot be included into ex-ante recovery rate prediction due to a substantial time lag between default and conversion of reorganization cases to liquidation. In all specifications, the centrality measure and neighbor industry distress remain stable and significant drivers of recovery, showing that these measures provide a new explanation to recovery rates rather than just replicating the asset heterogeneity that is captured in already known variables.

We now test the hypothesis that centrality influences corporate bond recovery rates through the asset transfer channel. Therefore, we first explore how centrality within the inter-industry network affects recovery rates conditional to industry-specific asset characteristics. If centrality has a positive effect on recovery rates because it facilitates the transfer of assets across industry borders, this effect should magnify when assets are less specified to an industry's needs and can potentially be employed by users in other industries. We start by focusing on asset specificity, the two non-redeployability measures, and asset customization, for which we have analyzed their main effects on recovery in Table 3.3. Now we test their interaction effects with centrality on recovery rates. We consider only these measures that capture the degree to which assets can be employed by alternative users, which is not necessarily the case for the asset mobility and asset durability measures that account for the transferability of assets from a cost and time perspective.[20]

The results of linear regressions including interaction effects are reported in Table 3.4. In Models (1) – (4), we find negative interaction effects of centrality with asset characteristics. In all cases, centrality has a large positive effect on recovery rates at the 1% significance level when assets are standardized and have more alternative uses, that is, when the asset specificity, non-redeployability, or customization measures are fixed at zero, since the variables are scaled on the interval [0,1]. In addition, we find highly significant interaction terms that bear negative signs of centrality with the asset characteristics. It shows that while recoveries of defaulted bonds benefit when assets are standardized, the positive effect of centrality shrinks with increasing specificity, non-redeployability, or customization. The findings for all four models are consistent, implying that the positive effect of industry centrality on recovery rates magnifies when industry assets have a higher possibility to be employed by alternative users. As all four employed explanatory variables, which reflect the alternative use of assets from different perspectives, form significant interaction effects with centrality, we find evidence that supports our hypothesis that the centrality measure operates through the asset transfer channel between industries. Although our analysis does not rule out other additional explanations of the inter-industry network's effect on recovery rates, the observed effects from asset characteristics together with centrality should be insignificant in the absence of the asset transfer channel. The analysis shows that bonds have higher recovery rates if their industry employs assets that can be used in alternative industries and is well connected within the inter-industry network of trade.

[20] Asset mobility measures the costs of physically transporting assets from one user to another, which is not likely to be related to an industry's trade relationships within the inter-industry network. Likewise, any effect of asset durability is more likely to be dependent on the duration of the legal proceedings to resolve default rather than industry centrality. As expected, we do not observe any significant interaction effects of these variables with centrality.

Table 3.4: Recovery rate and centrality – Interaction effects, OLS regression results. The recovery rate is the dependent variable. All regressions include seniority, industry, industry distress dummy variables, and firm- and bond-specific variables. Models (1) – (4) iterate the interactions of network centrality with asset characteristics. Model (5) shows the interaction of network centrality with a dummy variable indicating liquidation-type default events. Model (6) shows the interaction of network centrality with a dummy variable indicating industry-wide illiquidity. Standard errors are adjusted for heteroscedasticity and clusters at the firm level. Statistical significance at the 1%, 5%, and 10% level is indicated with ***, **, and *.

Model	(1)		(2)		(3)		(4)		(5)		(6)	
Intercept	0.0161		0.2613		0.3528		-0.0990		0.6030	**	0.4305	
Log(number of firms)	-0.5273	***	-0.3768	**	-0.5107	***	-0.2002		-0.3833	**	-0.3874	**
Labor's fraction of inputs	-0.1478		-0.0729		-0.0597		0.1142		0.0147		-0.2129	
Log(neighbor industry distress)	-0.2549	***	-0.2692	***	-0.2909	***	-0.2584	***	-0.2901	***	-0.2792	***
Log(eigenvector centrality)	1.2842	***	0.7880	***	0.7135	***	1.3289	***	0.3980	***	0.5087	***
Asset specificity	0.5508											
Log(eigenvector centrality) * Asset specificity	-0.9907	**										
Asset non-redeployability 1			0.6507	**								
Log(eigenvector centrality) * Asset non-redeployability 1			-1.4915	***								
Asset non-redeployability 2					0.9016	***						
Log(eigenvector centrality) * Asset non-redeployability 2					-1.7664	***						
Asset customization							0.4479					
Log(eigenvector centrality) * Asset customization							-1.7846	***				
Liquidation dummy									0.0293			
Log(eigenvector centrality) * Liquidation dummy									-0.3427	*		
Industry illiquidity											-6.4258	**
Log(eigenvector centrality) * Industry illiquidity											6.7233	**
R^2	0.5101		0.5172		0.5172		0.5193		0.5187		0.5064	
Adj. R^2	0.4992		0.5065		0.5065		0.5086		0.5080		0.4954	
RMSE	0.2301		0.2284		0.2284		0.2280		0.2281		0.2310	
Observations	1,479		1,479		1,479		1,479		1,479		1,479	
Industry dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Seniority dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Basic industry distress dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Firm- and bond-specific variables	Yes		Yes		Yes		Yes		Yes		Yes	

In the next step, we consider an interaction term of the liquidation dummy variable with centrality. The interaction of the liquidation indicator and industry centrality corresponds to the finding of Bernstein, Colonnelli, and Iverson (2018) that liquidated plants in bankruptcy cases are more likely to be sold to another industry than reorganized plants. At the same time, liquidation cases result in the sale of all assets, thus, the liquidation dummy variable captures the economic effects of asset sales under stress. Hence, if the industry centrality's positive effect on recovery is through the asset transfer channel, the combination of centrality and liquidation should have a significant effect on recovery rates. As shown in Model (5) in Table 3.4, the interaction effect is negative and significant at the 10% level, whereas centrality remains positive and significant at the 1% level. This shows that the effect of centrality has a positive effect both in liquidation cases (when the liquidation dummy variable is fixed at one) and reorganization cases (when the liquidation dummy variable is fixed at zero). The negative sign of the interaction term shows that the positive effect of centrality on recovery is larger for reorganization cases than for liquidation cases. As asset sales are not mandatory but also performed frequently in reorganization cases, this finding suggests that there exists a second economic mechanism of centrality beyond the asset transfer channel, possibly due to diversification aspects in well-connected industries. After all, however, this analysis confirms the positive effect of centrality on recovery for liquidation-type defaults that lead to asset sales.

Finally, we consider the interaction of centrality with illiquidity within the defaulted firm's industry at the time of default. If defaulted firms intend to sell assets, they need to find willing buyers. During times when potential buyers within the same industry are illiquid and thus constrained in their ability to absorb assets, our previous findings indicate a negative impact on recovery rates. As shown in Model (6), the interaction effect of centrality with illiquidity is significant at the 5% level. The large negative coefficient of illiquidity shows that its effect on recovery rate is highly negative when centrality is low. When centrality rises, however, the positive coefficients of centrality and the interaction term offset the negative effect of illiquidity. From an economic perspective, this means that industry-wide illiquidity is more influential on recovery rates in non-central industries, and less influential in central industries. As we hereby show that recoveries of firms in non-central industries are more dependent on the liquidity of their industry peers, this finding is another supportive evidence to our hypothesis that central industries, which are less affected when industry peers are illiquid, can more easily rely on the transfer of assets to industry outsiders. This is consistent with Shleifer and Vishny (1992), who argue that when industry peers are financially constrained, assets of bankrupt firms need to be sold to industry outsiders. As central industries have more trade connections to industry outsiders, our results provide an extension to their finding.

In summary, we collected various evidence in our analysis on the effects of centrality in combination with variables that capture asset characteristics, liquidation-type defaults and illiquidity. Our findings underpin the existence of an asset transfer channel through the inter-industry network of trade. Overall, our findings highlight how industries' positions within the U.S. inter-industry trade network and the economy's network structure affect the recovery rates of defaulted U.S. corporate bonds. We find that network-derived characteristics, particularly eigenvector centrality and neighbor industry distress, are key factors in driving recovery rates. The more centrally a firm is located within the inter-industry network network, the higher the bond recovery rate, despite controlling for alternative explanations such as established measures of industry-dependent asset heterogeneity. We additionally find that the positive effect of centrality on recovery rates is more pronounced when assets can be utilized by industry outsiders, in liquidation cases that lead to asset sales, and when the defaulted firm's industry peers are financially constrained. Thereby, we are able to establish stylized facts of the hypothesized inter-industry asset transfer channel.

Recovery rate, network centrality and the macroeconomy

Aobdia, Caskey, and Ozel (2014) highlight that returns and performance of firms in central industries within the inter-industry network of trade are more closely related to macroeconomic conditions than of those in non-central industries. They argue that central industries's better connectivity to other industries provide diversification of idiosyncratic shocks. At the same time, central industries are more prone to macroeconomic conditions than non-central industries, since macroeconomic shocks are more likely to originate in central industries and then propagate their closely connected trading partner industries. We test whether macroeconomic conditions also operate differently in central and non-central industries as drivers of recovery rates.

Only few macroeconomic variables were used in most prior studies seeking to predict recovery rates for corporate debt. In a recent study, Nazemi and Fabozzi (2018) report that recovery rate models with macroeconomic variables selected by LASSO outperform models with just a few macroeconomic variables. The main advantage of selection models compared to alternative data reduction models such as principal components is that they are much easier to interpret and link to the economic literature. The LASSO-selected variables are not robust to a small perturbation of the data (see, for example, Chinco, Clark-Joseph, and Ye (2019)). Feng, Giglio, and Xiu (2020) argue that using LASSO to select from a large number of risk factors is not a reliable way to find the best risk factors in asset pricing. We identify macroeconomic variables from a large set of U.S. macroeconomic variables that are related to recovery rates by using a stability selection technique. The stability selection generates many bootstrapped samples and counts how often each variable is selected. Lastly, the stability selection chooses the set of variables that are selected more often than a specific threshold. Therefore, the stability selection robustly identifies macroeconomic variables that are economically meaningful but not easy to realize otherwise.

Descriptive statistics for the nine stability selection technique selected macroeconomic variables can be found in Table B.5 of the Appendix. The following regression model adds the selected macroeconomic variables to the basic specification:

$$\begin{aligned}\text{Recovery rate} = \alpha + \beta(&\text{bond seniority and industry characteristics}) \\ &+\mu(\text{firm- and bond-specific variables}) + \delta(\text{network-derived variables}) \\ &+\gamma(\text{selected macroeconomic variables}) + \epsilon,\end{aligned} \tag{3.2}$$

Model (A2) in Table 3.5 shows the results of linear regression incorporating the selected macroeconomic variables (unemployment of less than five weeks, the spread of Aaa rated bonds to the U.S. federal funds rate, the high yield default rate, inventories of U.S. firms, inflation expectations, personal saving rate, new housing starts, the USD/GBP exchange rate, and volatility of the S&P 500) in addition to the basic variables employed in Model (A1). When we add network-derived variables in Model (A3), centrality and neighbor industry distress keep their individual directional relationships and order of magnitude with centrality that we observed previously. We observe significant negative effects on recovery from increasing inventories of U.S. firms, new housing starts, unemployment of less than five weeks, and the spread of Aaa rated bonds to the U.S. federal funds rate, and we find a significant positive effect of the USD/GBP exchange rate. The negative relationship of recovery with unemployment suggests that unemployment serves as an indicator of macroeconomic distress, consistent with Nazemi and Fabozzi (2018).

Our observation of a significant negative relationship between bond recovery and the spread of Aaa rated bonds to the U.S. federal funds rate supports the findings of Mora (2015) who shows that spreads negatively affect recovery rates in debt-dependent industries. While PD is generally positively associated with credit spreads, the negative relationship between spreads and recovery rate is consistent with Chava, Stefanescu, and Turnbull (2011) and the findings of Frye (2000) that PD and recovery rate are inversely correlated. In contrast to other studies such as Altman, Brady, Resti, and Sironi (2005), who find a negative relationship of recovery rates with market-wide default rates, we find evidence that the high yield default rate is positively related to corporate bond recovery rates. This observation suggests that the relationship

Table 3.5: Recovery rate and centrality – Macroeconomic effects, OLS regression results. The recovery rate is the dependent variable. All regressions include seniority, industry, industry distress dummy variables, and firm- and bond-specific variables. Panel A employs the full sample, whereas Panel B and C employ samples containing only the 50% most central, and the 50% least central observations. In each panel, the first column represents the basic specification. In the second column, macroeconomic variables are added. In the third column, network-derived variables are added. Standard errors are adjusted for heteroscedasticity and clusters at the firm level. Statistical significance at the 1%, 5%, and 10% level is indicated with ***, **, and *.

	Panel A: All data			**Panel B: Central (50% most central)**			**Panel C: Non-central (50% least central)**		
Model	**(A1)**	**(A2)**	**(A3)**	**(B1)**	**(B2)**	**(B3)**	**(C1)**	**(C2)**	**(C3)**
Intercept	0.3117	0.0432	0.0837	0.6826 ***	0.2559	0.1929	0.3726	0.4965 **	0.3409
Log(number of firms)			-0.3400 **			0.3688 **			-0.0843
Labor's fraction of inputs			0.0640			0.4381 **			0.1599
Log(neighbor industry distress)			-0.2755 ***			-0.3059 *			-0.2348 ***
Log(eigenvector centrality)			0.3064 **						
Invent change		-0.0005 **	-0.0006 ***		-0.0001	0.0003		-0.0004	-0.0005 *
Inflation expect		-0.0579 *	-0.0278		-0.1621 **	-0.1688 ***		0.0241	0.0435
Saving rate		-0.0223	-0.0366 ***		-0.0210	-0.0222		-0.0495 ***	-0.0548 ***
S&P 500 vol		-0.0013	-0.0017		-0.0032	-0.0015		-0.0010	-0.0022
Ex rate: UK		0.4921 ***	0.4091 ***		0.3916 *	0.4196 **		0.3051 **	0.2615 **
Starts: NE		-0.0016 ***	-0.0016 ***		-0.0014 *	-0.0015 *		-0.0012 ***	-0.0012 ***
U <5 wks		-0.0002 *	-0.0002 **		0.0000	0.0000		-0.0001	-0.0001
Aaa-FF spread		-0.0575 ***	-0.0303 *		-0.1055 ***	-0.0731 **		-0.0107	0.0138
High yield DR		0.0065 *	0.0117 ***		0.0035	0.0138		0.0035	0.0088 **
R^2	0.4456	0.5316	0.5562	0.5390	0.6835	0.7007	0.3101	0.3818	0.4053
Adj. R^2	0.4360	0.5202	0.5442	0.5255	0.6701	0.6867	0.2918	0.3574	0.3792
RMSE	0.2442	0.2252	0.2195	0.2163	0.1804	0.1758	0.2472	0.2355	0.2315
Observations	1,479	1,479	1,479	739	739	739	739	739	739
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seniority dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm- and bond-specific variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

between the default rate (or PD) and recovery rate may not be linear across bond yields and credit ratings.

Next, we evaluate whether the findings of Aobdia, Caskey, and Ozel (2014), that the economic performance of firms in central industries is more related to macroeconomic conditions than in peripheral industries, also hold for recovery rates. We divide the defaulted corporate bonds into two subsets involving the 50% most and the 50% least central observations. We then apply separate linear regressions on these distinct bond subsets. We are interested in determining the additional portion of variation in recovery rates that can be explained by adding macroeconomic variables. As the R-squared measures the extent to which the overall variation in recovery rates is accounted for by the explanatory variables employed, the difference in R-squareds between the central and the non-central samples allows to draw implications on the relative degree to which the recovery rate is explained by the macroeconomic conditions.[21]

[21] Note that the comparison of R-squareds only allows for a relative assessment of the dependency on the explanatory variables employed. Our approach is somewhat related to that of Aobdia, Caskey, and Ozel (2014) but differs in that they compare predicted R-squareds to measure the effect of macroeconomic conditions on rolling windows of daily and monthly stock returns in central and non-central industries. Due to the sparsity of default observations compared to stock return data, such approach is not applicable in the context of recovery rate estimation.

Panel B (central industries) and Panel C (non-central industries) in Table 3.5 present the results of adding macroeconomic variables to linear regression for explaining recovery rates of bonds in central and in non-central industries. As we use network centrality to rank and separate the data sample into two subsamples, we remove it from the universe of explanatory variables. In general, we observe that the variables employed have better explanatory power for defaults in central industries since all models in Panel B have considerably higher R-squareds relative to the corresponding models in Panel C. For the central subsample, as shown in Model (B2), adding macroeconomic variables adds more than 14 percentage points to the R-squared compared to the baseline specification. In the non-central industries, the corresponding specification with macroeconomic variables in Model (C2) adds around 7 percentage points. Furthermore, using network-derived variables in Models (B3) and (C3) shows that the neighbor industry distress measure remains significantly negatively related to recovery in both specifications, although the significance is less strong in central industries than in non-central industries.

Interestingly, the significance of macroeconomic variables differs for central and non-central industries. However, the significant macroeconomic variables keep the same signs as when considering the whole dataset in Panel A. The varying significance of macroeconomic variables by centrality is possibly due to the heterogeneity of economic activities and thus dependence on different macroeconomic environments across industries. Overall, and in line with our expectation, we find that adding macroeconomic variables to linear regression leads to a much greater amount of explained variation in the recovery rate in central industries than in non-central industries. That allows us to draw the conclusion that recovery rates in central industries are indeed relatively more related to macroeconomic conditions than in non-central industries. This interesting insight also enhances the findings of Mora (2015). While she reports that the macroeconomic variables used in modeling recovery rates do not have the same effects on each industry, we show that macroeconomic effects in fact operate differently in central and non-central industries.

3.4.3 Robustness checks

In this section, we examine the robustness of our findings in various settings. We first consider alternative data subsamples, and start by testing whether our linear regression results still hold for a subsample that excludes all bonds of financial firms. We also account for periods of recessions and non-recession times, in order to assess whether our findings hold across the economic cycle. To further test the plausibility of our findings, we then substitute the inter-industry trade network for creating the industry centrality measure with alternative networks derived from actual inter-industry asset acquisitions. Finally, we evaluate the out-of-sample validity and real-world applicability of our models through advanced machine learning methods.

Robustness to alternative data subsamples and alternative networks

We examine the robustness of our findings to symptoms that are linked to the financial industry, or to different regimes of the economic cycle. In times of economic distress, established economic mechanisms may shift or cease to operate in the way they do during periods of economic stability. As an example, the GFC had caused particular difficulties for the financial industry, leading to bankruptcies of large financial institutions. In order to rule out that the results of our analysis are driven by defaults and recoveries within the financial industry, we consider a subsample of our data excluding all bonds of issuers from the financial industry. The regression involving all 1,050 bonds from non-financial firms verifies the significance of the newly introduced centrality and neighbor industry distress measures, as shown in Model (A2) in Table 3.6, where we add them to the basic specification (Model (A1)). The regression also produces the same directional effects for all network-derived variables as the overall sample.

Having demonstrated the robustness of our findings to excluding financial firms, we now test the robustness of the recovery rate models to the economic cycle. We do so by performing linear regressions for two distinct subsets of the bonds that represent non-recession (Panel B) and recession (Panel C) periods. We choose the U.S. recession periods as March 2001 until November 2001 and December 2007 until June 2009 according to the definitions of U.S. recessions of the National Bureau of Economic Research (NBER). The results show that the basic specification, which employs bond seniority and industry characteristics, as well as other firm- and bond-specific variables, has considerably higher explanatory power during the non-recession

Table 3.6: Recovery rate and centrality – Robustness check, OLS regression results. The recovery rate is the dependent variable. All regressions include seniority, industry, industry distress dummy variables, and firm- and bond-specific variables. In each Panels A – C, the first column represents the basic specification including bond- and firm-specific information. In the second column, the network-derived variables are added. Panel A considers only default events of non-financial firms. Panels B and C only consider default during non-recession and recession periods. The recession periods are March 2001 until November 2001, and December 2007 until June 2009 following the definitions of the National Bureau of Economic Research (NBER). Panel D employs alternative networks to derive the centrality measure. Model (D1) employs centrality derived from the detailed inter-industry trade network involving 471 distinct industries. In Model (D2), the network for deriving centrality is based on inter-industry non-bankruptcy mergers and acquisitions (M&A), and in Model (D3) on bankruptcy M&A. Standard errors are adjusted for heteroscedasticity and clusters at the firm level. Statistical significance at the 1%, 5%, and 10% level is indicated with ***, **, and *.

	Panel A: Non-financial subsample		Panel B: Non-recession subsample		Panel C: Recession subsample		Panel D: Alternative networks		
Model	(A1)	(A2)	(B1)	(B2)	(C1)	(C2)	(D1)	(D2)	(D3)
Intercept	0.3430	0.5289	0.2400	-0.0191	0.5107 **	0.9095 *	0.4166 *	0.3303	-0.1586
Log(number of firms)		-0.1494		-0.0713		-0.2533	-0.1151	0.5250 ***	0.6529 ***
Labor's fraction of inputs		-0.1572		-0.0595		-0.6428	-0.2840 ***	-0.1238	-0.1221
Log(neighbor industry distress)		-0.1624 **		-0.3081 ***		-0.0795	-0.2249	-0.2903 ***	-0.2584 ***
Log(eigenvector centrality)		0.2263 **		0.3217 **		0.5835 **			
Log(eigenvector centrality) Detailed							0.2053 ***		
Log(eigenvector centrality) M&A								0.5250 ***	
Log(eigenvector centrality) Bankruptcy M&A									0.6529 ***
R^2	0.3450	0.3710	0.4783	0.5382	0.3413	0.4019	0.4986	0.4839	0.4946
Adj. R^2	0.3303	0.3544	0.4677	0.5272	0.2900	0.3464	0.4882	0.4736	0.4845
RMSE	0.2560	0.2513	0.2292	0.2160	0.2779	0.2666	0.2326	0.2359	0.2335
Observations	1,050	1,050	1,160	1,160	319	319	1,479	1,479	1,479
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seniority dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm- and bond-specific variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

period than during the recession. When adding network-derived characteristics, centrality has a positive coefficient and is statistically significant. Neighbor industry distress, however, has a significant negative relationship with recovery when applying the non-recession sample, but ceases to be significant during the recession, although keeping a negative sign. Not surprisingly, due to the nature of recessions, we observe that only three defaulted bonds don't experience any distress in neighbor industries during the recession, and hence are underrepresented in the cross-variation. Thus, one cannot meaningfully interpret the insignificant coefficient of neighbor industry distress. After all, the centrality measure keeps its relationship with recovery even during the recession, and our analysis shows that centrality plays a significant role in reliably determining recovery rates as it is robust to the economic cycle.

Finally, we consider alternative networks for deriving centrality. Panel D in Table 3.6 first applies the detailed 1997 inter-industry network of trade involving 471 distinct private sector industries (Model (D1)). This model confirms the significance of centrality for increasing the granularity of the network representation compared to our main analysis. Although this specification yields a highly statistically significant positive coefficient of centrality, the magnitude of the coefficient is smaller than for the less detailed network. We then apply centrality derived from the non-bankruptcy M&A network (Model (D2)) and then from the bankruptcy M&A network (Model (D3)). By doing so, we test whether the inter-industry network of trade's capability to infer predictions on the recovery rate due to facilitating inter-industry asset transfers is robust to replacing it with the inter-industry networks of realized asset transfers. Although the M&A networks only reflect realized outcomes and hence are likely to suffer from selection bias, we do find that industry centrality in the inter-industry networks of realized M&A transactions as well as realized bankruptcy M&A transactions have a positive and highly significant effect, comparable to when we use the inter-industry trade network. Other network-derived variables that are significant in our main analysis remain stable. As the employed M&A networks cover the full period 2001–2016 and potentially contain selection bias, they are not suited for ex-ante estimation of recovery rates, but the analysis demonstrates the validity of the hypothesized inter-industry asset transfer channel for the formation of recovery rates.

Out-of-sample prediction using machine-learning methods

Most prior studies consider that corporate bond recovery rates linearly depend on the explanatory variables (see, for example, Varma and Cantor (2005), Acharya, Bharath, and Srinivasan (2007), and Jankowitsch, Nagler, and Subrahmanyam (2014)). Qi and Zhao (2011), Altman and Kalotay (2014), Yao, Crook, and Andreeva (2015), Nazemi and Fabozzi (2018), Sopitpongstorn, Silvapulle, Gao, and Fenech (2021), and Kellner, Nagl, and Rösch (2022) demonstrate that non-linear methods outperform linear methods to predict recovery rates. We apply our recovery rate estimation models in predictions via machine learning techniques in order to reflect the non-linear dependency between the recovery rate and explanatory variables, and assess the out-of-sample prediction performance in order to benchmark the real-world applicability of the different groups of variables. We estimate the recovery rate by employing least squares support vector regression, least squares support vector regression with different intercepts for seniorities, and semi-parametric least squares support vector regression, which we describe in Appendix B.4.

Table 3.7 shows the results from the out-of-sample prediction, using 10-fold cross-validation.[22] Model (1) in Table 3.7 employs only bond seniority, industry dummy variables and basic industry distress variables. To this basic specification, we alternately add the following variables: firm- and bond specific variables (Model (2)), measures derived from the inter-industry network of trade (Model (3)), and the selected macroeconomic variables (Model (4)). In the final specification (Model (5)), we combine all groups of variables. In all models, support vector regression techniques outperform the linear regressions by a large margin, a finding that is consistent with Nazemi and Fabozzi (2018) who compare the accuracy of support vector regression techniques with statistical methods. This high improvement in the determination of recovery rates by machine learning models for all individual groups of explanatory variables indicates that the relationship between recovery rates and the explanatory variables, including the network-derived variables, is not just linear, analogue to what is documented for established recovery rate drivers in earlier research. When comparing the out-of-sample fit of the different models, we observe that Model (5), that includes all variables altogether, has the best predictive power among the models. Notably, however, adding the variables derived from the inter-industry network improves model accuracy by a comparable order of magnitude as the bond- and firm-specific variables or macroeconomic variables that are established drivers of recovery rates in the literature. Our analysis demonstrates that the network-derived variables offer a robust explanation of corporate bond recovery rates even in out-of-sample settings.

In summary, we provide various tests to explore the robustness of the variables derived from the inter-industry network of trade to alternative explanations and settings. Showing that the main findings of our study remain intact under these settings makes a compelling case for the role of the inter-industry network of trade in explaining recovery rates through enabling inter-industry asset transfers.

[22] For performing a 10-fold cross-validation, we randomly split the dataset into 10 subsets while stratifying for seniority. The reported performance is obtained by averaging the 10 performances measured in the cross-validation. We run a grid search during the cross-validation process for choosing the appropriate hyper-parameters for the support vector regression specifications. The model specification for reporting is chosen based on the lowest average mean squared error (MSE).

Table 3.7: Recovery rate and centrality – Out-of-sample prediction, OLS and support vector regression results. The recovery rate is the dependent variable. All regressions include seniority, industry, and industry distress dummy variables. Model (1) represents the basic specification (seniority, industry, and industry distress dummy variables). Model (2) adds bond- and firm-specific information to the basic specification. Model (3) adds network-derived variables to the basic specification. Model (4) adds selected macroeconomic variables to the basic specification. Model (5) combines all variables employed in Models (1)–(4). The regression techniques include linear regression (Lin. Reg.), least squares support vector regression (LS–SVR), least squares support vector regression with different intercepts for seniorities (LS–SVR DB), and semi-parametric least squares support sector regression (LS–SVR SP). Out-of-sample prediction performance is based on 10-fold cross-validation. The best value for each performance measure for the respective model is highlighted in bold.

Model (1)	R^2	σ_{R^2}	$Adj.\ R^2$	$\sigma_{Adj.\ R^2}$	RMSE	σ_{RMSE}	MAE	σ_{MAE}
Lin. Reg.	0.3710	0.0561	0.3643	0.0567	0.2579	0.0194	0.1905	0.0147
LS – SVR	0.4223	0.0518	0.4162	0.0524	0.2471	0.0189	**0.1845**	0.0132
LS – SVR SP	0.3932	0.0607	0.3867	0.0613	0.2533	0.0200	0.1900	0.0135
LS – SVR DB	**0.4226**	0.0516	**0.4165**	0.0522	**0.2471**	0.0188	0.1846	0.0134
Model (2)	R^2	σ_{R^2}	$Adj.\ R^2$	$\sigma_{Adj.\ R^2}$	RMSE	σ_{RMSE}	MAE	σ_{MAE}
Lin. Reg.	0.4312	0.0536	0.4203	0.0546	0.2451	0.0182	0.1761	0.0121
LS – SVR	**0.5936**	0.0936	**0.5859**	0.0954	**0.2067**	0.0239	**0.1162**	0.0137
LS – SVR SP	0.5899	0.0902	0.5821	0.0919	0.2076	0.0224	0.1245	0.0127
LS – SVR DB	0.5725	0.0671	0.5643	0.0684	0.2122	0.0184	0.1375	0.0085
Model (3)	R^2	σ_{R^2}	$Adj.\ R^2$	$\sigma_{Adj.\ R^2}$	RMSE	σ_{RMSE}	MAE	σ_{MAE}
Lin. Reg.	0.4412	0.0497	0.4336	0.0504	0.2429	0.0173	0.1780	0.0103
LS – SVR	0.5814	0.0920	0.5757	0.0932	0.2097	0.0229	0.1354	0.0129
LS – SVR SP	**0.5898**	0.0856	**0.5842**	0.0868	**0.2076**	0.0220	**0.1308**	0.0117
LS – SVR DB	0.5768	0.0903	0.5710	0.0915	0.2108	0.0223	0.1367	0.0127
Model (4)	R^2	σ_{R^2}	$Adj.\ R^2$	$\sigma_{Adj.\ R^2}$	RMSE	σ_{RMSE}	MAE	σ_{MAE}
Lin. Reg.	0.4660	0.0562	0.4566	0.0571	0.2374	0.0175	0.1725	0.0113
LS – SVR	**0.5977**	0.1279	**0.5906**	0.1302	**0.2053**	0.0299	**0.1122**	0.0165
LS – SVR SP	0.5849	0.1276	0.5776	0.1298	0.2087	0.0298	0.1205	0.0158
LS – SVR DB	0.5878	0.1222	0.5806	0.1244	0.2080	0.0286	0.1279	0.0144
Model (5)	R^2	σ_{R^2}	$Adj.\ R^2$	$\sigma_{Adj.\ R^2}$	RMSE	σ_{RMSE}	MAE	σ_{MAE}
Lin. Reg.	0.5273	0.0733	0.5134	0.0754	0.2233	0.0204	0.1543	0.0129
LS – SVR	**0.6200**	0.1243	**0.6088**	0.1280	**0.1994**	0.0300	**0.1016**	0.0161
LS – SVR SP	0.6139	0.1249	0.6026	0.1286	0.2011	0.0300	0.1031	0.0161
LS – SVR DB	0.6000	0.1226	0.5882	0.1263	0.2049	0.0291	0.1244	0.0142

3.5 Conclusions

The large number of default events during the 2007-2008 Global Financial Crisis has spotlighted the dynamics of corporate bond recovery rates. It also became apparent that interconnections between industries and the fragile structure of the financial institutions' network were significant factors driving systemic risk and the propagation of shocks in the economy. By employing an empirical representation of the U.S. inter-industry trade network and comprehensive data on defaulted bonds in the United States over the period 2001–2016, we examine the link between the recovery rates of defaulted U.S. corporate bonds and observable inter-industry trade relationships.

This empirical study shows that corporate bond recovery rates are driven by an economic mechanism which operates beyond industry borders and which is dependent on the network structure of the economy. Although it is known that defaulted bond recovery rates vary across industries, recent studies neglect that industries do not exist in isolation from each other. Inter-industry trade relationships allow the transmission of assets, information and systematic shocks between industries. We find that industries which are more central in the network of inter-industry trade have higher recovery rates than industries that are less central. We provide evidence that the connectivity to other industries through trade relations facilitates asset disposals across industry borders in default, which is reflected in higher recoveries in central industries. By considering industry-specific asset characteristics that account for the ability to deploy assets for alternative uses, we find that the positive effect of network centrality magnifies in industries with assets that are less specialized to a specific industry's needs.

By defining a neighbor industry distress measure, we also find that bond recoveries suffer from distress in directly connected industries which propagates through trade relationships, an economic mechanism which we observe for the first time in the literature. Our robust findings are independent from the business cylce or disruptions from the financial industry during the GFC. Moreover, our findings are independent from variations in the granularity of industry definitions to create the inter-industry network for trade and further backed by actual inter-industry asset disposal data from a database of M&A transactions in bankruptcy. Our study furthermore reveals how macroeconomic drivers of recovery rate operate in central and non-central industries. Macroeconomic conditions explain a greater fraction of recovery rates' variations in central than in non-central industries. Finally, applying the network-derived characteristics in machine learning models can explain recovery rates comparably well as the established recovery rate drivers in out-of-sample prediction.

To conclude, our results provide novel evidence that an industry's position in the network of inter-industry trade is an important driver of corporate bond recovery rates, as it determines a firm's ability to dispose assets to other industries, especially when assets are not specialized to a specific industry's needs. Thus, our findings highlight a yet undiscovered channel in the formation of recovery rates.

Chapter 4

Corporate Bond Recovery Rate and Financial Markets

This chapter is joint working paper with Abdolreza Nazemi.[1][2]

4.1 Introduction

Traditional asset pricing models offer insights into the pricing of portfolios of stocks and bonds. Using factor models, researchers and practitioners identify a relationship of individual asset returns with the prevailing market-wide conditions. For stocks, such factor models have become standard tools for constructing portfolios and assessing risks. Moreover, factor models explain asset specific risk-return characteristics, allowing to identify the common return drivers and eventually infer predictions on future return performance of an asset.

Many equity factors have been identified and studied in the recent decades (see, for example, Jensen, Kelly, and Pedersen (2023)), however, the factor structure in corporate bond returns was established only lately. Bai, Bali, and Wen (2019) introduce factors for corporate bond returns, building a portfolio approach in a similar fashion as the seminal study on equity returns of Fama and French (1993). However, as Dickerson, Mueller, and Robotti (2023) show in a more recent study, these factors are erroneous and do not offer any incremental explanatory power.

[1] Abdolreza Nazemi is with the School of Economics and Management, Karlsruhe Institute of Technology, Germany.

[2] We would like to thank Prof. Marliese Uhrig-Homburg, Prof. Philipp Schuster, Dr. Marcel Müller, and Dr. Michael Reichenbacher for providing size-adapted bond liquidity data and expected liquidity measures.

Dickerson, Julliard, and Mueller (2023) revisit a multitude of potential bond return factors from the literature and find that the majority of factors are no sources of priced risk. Kelly, Palhares, and Pruitt (2023) expand the bond factor approach to including latent factors derived from an instrumented principal component approach, and show that this approach improves modeling accuracy for corporate bond returns.

As stock and bond risk factors explain the pricing of a firm's capital structure, the question of integration of bond and equity markets arises naturally. That is, do stock and bond returns depend on the same risk factors? While Choi and Kim (2018) find that risk premia in bond markets differ from those in equity markets, Kelly, Palhares, and Pruitt (2023), find evidence of a closer integration of both markets.

Once a bond defaults, its investment characteristics change significantly. Therefore, studies on recovery pricing neglect the general conditions in equity and bond markets that are captured by factor models, focusing on the unique recovery drivers, such as bond- and firm-specific variables, trading-specific information, and macroeconomic conditions.[3]

In this paper, we establish novel evidence on the relationships between corporate bond recovery rates and prevailing conditions in stock and bond markets. While bond factor models explain returns of corporate bonds in good standing, the nature of a bond changes materially in case of default from normal bond to a more equity-like asset. Whereas bonds in good standing typically offer a low-risk, low-return investment profile, defaulted bonds alter the risk-return profile more close to that of equities, with high risk, but also potentially high returns, or total loss. Examining the sensitivity of defaulted bonds to traditional asset pricing characteristics, such as equity risk factors, bond risk factors, and others, we document a close relationship between the pricing of defaulted bonds and the prevailing conditions in financial markets. The key idea of our study is to capture the dependence of defaulted bonds, which share asset characteristics of both stocks and bonds, on the pricing conditions of the markets in which stocks and bonds are traded.

Our empirical analysis uncovers a pronounced relationship of defaulted bond recovery and financial markets. As defaulted bonds trade in OTC bond markets, we investigate how bond

[3] See, for example, Acharya, Bharath, and Srinivasan (2007), Altman, Brady, Resti, and Sironi (2005), Jankowitsch, Nagler, and Subrahmanyam (2014), Mora (2015), Nazemi and Fabozzi (2018), or Baumann, Kakhbod, Livdan, Nazemi, and Schürhoff (2023)

market liquidity affects the pricing in defaulted bonds. We construct measures of market-wide liquidity relying on OTC trading data, and draw on alternative liquidity measures from the literature. While the idiosyncratic liquidity provision in defaulted bonds has been researched before (see Jankowitsch, Nagler, and Subrahmanyam (2014)), we show that when bond markets experience a liquidity crisis, the pricing of defaulted bonds will be impaired. Thus, the liquidity provision in the overall bond market affects the expected recovery in defaulted bonds. As such, the functioning of the overall bond market has a significant effect on recovery outcomes.

Comparing recovery outcomes with risk factors established in the literature for bond pricing, we find a link between defaulted bond recovery and the time-varying bond risk factors. Positive bond market returns correlate with higher bond recovery, and during times when returns of risky bonds outperform low-risk bonds, recoveries are even higher. Thus, we can draw conclusions from investors' pricing of normal bonds on the pricing of defaulted bonds. Complementary, we further augment bond market conditions with a corporate bond market distress index and a financial markets index, and find that distress in the overall bond market and tightening of financial conditions have detrimental effects on bond recoveries. These results show that even if a bond's risk-return profile changes with default, the recovery pricing will still be subject to conditions in the bond market, as well as to general financial conditions.

To assess whether the change of a bond's risk-return profile after default to a more equity-like asset is reflected in a dependence of bond recovery on prevailing pricing conditions in equity markets, we further consider a range of equity risk factors. Indeed, we find a relationship of defaulted bonds with equity risk factors. E.g., when equity markets price high-risk equities higher, the recoveries of defaulted bonds appreciate. Alternatively, when equity markets price equity portfolios with high profitability higher than those with low profitability, defaulted bonds' recoveries decline. Finally, we consider prevailing equity valuations, and further find that defaulted bonds recover more when the market values equities higher relative to industry-wide financial performance metrics such as net income, or sales. Thus, we find that defaulted bond pricing is closely related to the pricing and valuation conditions in equity markets.

With our empirical analysis utilizing a variable importance ranking methodology, we further find that many of the market-derived characteristics are more important in explaining corporate bond recovery rates than established recovery rate variables such as, for example, macroeconomic conditions, bond seniority, default type, or industry affiliation. Overall, the cumulative importance of market-derived variables is approximately 60% relative to that of the established recovery rate drivers, indicating the important role of considering market conditions for estimating defaulted bond recovery rates.

The main contribution of our paper is to expose the link between defaulted bond recovery and conditions in U.S. financial markets. Jankowitsch, Nagler, and Subrahmanyam (2014) show that the liquidity of defaulted bonds within the OTC market is related to their respective recovery. Altman, Brady, Resti, and Sironi (2005) explain recovery rate as the intersection of supply and demand in defaulted bonds. Baumann, Kakhbod, Livdan, Nazemi, and Schürhoff (2023) establish a link between the OTC market structure, defaulted bond trading, and investors' recovery rates in defaulted bonds. While these studies acknowledge the importance of defaulted bonds trading in the bond market, they neglect whether a change in the bond market's conditions may affect recovery. Controlling for a range of alternative factors, including those that might be considered related to financial markets, such as macroeconomic conditions, industry-specific distress, and time fixed effects, we consider a variety of market-derived characteristics that explain pricing in both equity and bond markets. Finding a significant effect of these variables for bond pricing, we establish a novel understanding of the market-based drivers of recovery rates. Thus, we expand the growing research on recovery drivers such as Acharya, Bharath, and Srinivasan (2007), Mora (2015), and Nazemi and Fabozzi (2018).

With our study, we also contribute to the literature of bond and equity pricing. Kelly, Palhares, and Pruitt (2023), Dickerson, Mueller, and Robotti (2023) and Dickerson, Julliard, and Mueller (2023) investigate the effects of bond factors on bond pricing. We utilize their factors in bond pricing, however, for those bonds that recently defaulted. By doing so we capture how normal bond investors' prevailing risk-return preferences are reflected in the pricing of defaulted bonds. In a similar fashion, we capture the equity pricing implications on bond recoveries from factors such as in Fama and French (2015) and Jensen, Kelly, and Pedersen (2023), showing that not only bond investors' risk-return preferences, but also those of equity investors, are reflected in the pricing of defaulted bonds.

Finally, we contribute to shaping a better understanding of the hybrid nature of defaulted bonds, which share characteristics of both bonds and equity-like investments. Choi and Kim

(2018) assess the integration of bond and equity markets, finding limited integration between the two markets. In a related study, however, Kelly, Palhares, and Pruitt (2023) find that equity risk factors explain a large fraction of bond returns when utilizing risk factors created by an instrumented principal components approach. By exposing the dependence of recovery rates on various bond and equity risk factors, bond market liquidity, bond market distress and financial conditions, as well as equity valuation levels, we add to the understanding of the integration of equity and debt markets.

The paper is organized as follows. Section 4.2 describes the data and defines the sample used in our empirical analysis, including market-derived variables. Section 4.3 contains our empirical study of recovery rates subject to conditions in financial markets. Finally, we conclude our work in Section 4.4.

4.2 Data

In the following, we present the data sources and methodologies for building the dataset of defaulted bonds and for creating the explanatory variables. We further provide definitions of the variables used in our recovery rate predictions.

4.2.1 Defaulted bond data and explanatory variables

We collect data from various sources. The defaulted bond and recovery data sample is constructed with data from Moody's Default & Recovery Database (DRD), the Mergent Fixed Income Securities Database (FISD), S&P Capital IQ, Thomson Reuters, and FINRA's Transaction Reporting and Compliance Engine (TRACE). Bond and firm characteristics are from S&P Capital IQ and FISD, and macroeconomic variables are from the Federal Reserve Bank of St. Louis (FRED). We create proxies for bond liquidity and bond market liquidity with transaction data from TRACE, and obtain additional bond market liquidity measures and both bond and equity risk factors from the literature. Equity market valuations are from the Wharton Research Data Services (WRDS). Bond market distress data is from the Federal Reserve Bank of New York, and National Financial Conditions Index data is from the Federal Reserve Bank of Chicago. In the following, we describe the dataset and data sources in more detail.

Default Events We build a large dataset of defaulted bonds over the period 2004-2016, considering a variety of default events. We retrieve default-day combinations of corporate bonds for reorganizations (Chapter 11), liquidations (Chapter 11 and Chapter 7) and distressed exchanges from Moody's DRD, Mergent FISD, S&P Capital IQ and Thomson Reuters. Consistent with Jankowitsch, Nagler, and Subrahmanyam (2014), we additionally consider bonds that were downgraded to one of the two lowest rating classes. Therefore, we utilize historical rating data of Moody's, S&P, and Fitch Ratings available via FISD. We identify bond downgrades to the second lowest rating class (unlikely-to-pay events and situations close to formal default, e.g., S&P's C rating) and downgrades to the lowest rating class (formal defaults, e.g., S&P's D rating). Due to the multitude of data sources and event definitions used for identifying default events, we yield several different default-day combinations for individual bonds that relate to the same default event. For example, rating downgrades may occur even before a bond formally defaults, and we may identify both as separate default-day combinations. Hence, in order to avoid multiple default-day representations of the same default event of a bond, we only consider the first default-day combination of a bond and thereafter only consider any consecutive default-day observation of the same bond after a time delay of at least one year.[4] For our analysis, we only consider bonds for which trading information around the default date is available in TRACE for calculating the recovery rate, and for which there is bond-specific information available in FISD. Based on this procedure, we identify 2,636 corporate bond default events. The data covers 2,425 distinct bonds of 498 distinct US-based issuer firms, and represents 679 unique default events.[5]

Recovery Rate Similar to Jankowitsch, Nagler, and Subrahmanyam (2014), we define the recovery rate of bond j as its average trading price as a percentage of par value on the default day and the 30-day period thereafter:

[4] In fact, bonds may be reinstated after default and, in some cases, default again thereafter.

[5] To the best of our knowledge, our dataset of defaulted bonds is one of the most recent and most comprehensive datasets utilized in recent recovery rate research. For comparison, Jankowitsch, Nagler, and Subrahmanyam (2014) identify 2,235 default events of 818 corporate bonds during the period 2002 to 2010, and thus consider a multitude of default-day combinations per bond that may be related to the same default event. Altman and Kalotay (2014) and Kalotay and Altman (2017) analyze 2,828 corporate bonds, among other debt instruments, over the period 1987-2011. Qi and Zhao (2011) consider 2,367 corporate bonds between 1985 and 2008, as well as other credit instruments. Mora (2015) uses data of 3,659 defaulted corporate bonds in addition to loans and preferred stock over the period 1970-2008.

$$RR_i = \frac{1}{T+1} \sum_{s=t}^{t+T} \left(\frac{1}{K_{i,s}} \sum_{k_{i,s}} \frac{price_{s,k_{i,s}}}{par_i} \right). \tag{4.1}$$

where t is the bonds' default date, K is the number of reported bond trades on day s, *price* is the transaction price, and *par* is a bond's par value. We consider TRACE data to observe prices paid on the default day t and during the 30 days after default. For bondholders that liquidate their bond holdings immediately after default, this definition of the recovery rate represents actual recovery. A further advantage of this definition is its availability close to the default event, whereas ultimate recovery may be only observed after a time period that is unknown ex ante.[6] The mean recovery rate is 38.83% with a standard deviation of 28.82%. About two thirds of the defaults occurred during the global financial crisis. In 2008, the recovery rates are the lowest, with a mean recovery rate of only 22.61%, and further show also the lowest standard deviation in the low 20s as percentage of par value. In contrast, the two years before exhibit the highest average recovery rates (66.96% in 2006 and 63.21% in 2007). The majority of default events are Chapter 11 default events (reorganizations and liquidations), which recover about 37–38% of par value on average. Distressed exchanges recover about 60% of par value and are the highest recoveries in our sample, which is consistent with the literature (see, for example, Mora (2015)), followed by default risk rating events which recover about 59%. Default ratings and Chapter 7 liquidations are the most severe default events, with each recovering about 26% of par value. Figure 4.1 shows the histogram of the recovery rate distribution over the whole sample and the yearly time-variation, and the variation by default event type of the recovery rate can be found in Table 4.1.

Bond Characteristics In order to account for cross-sectional characteristics of the defaulted bonds, we include a range of bond-specific variables. We collect each bond's industry affiliation and seniority ranking from FISD. Consistent with the literature, the most senior bonds show the highest recovery, at about 65% of par value. The recovery rate drops for more junior bonds, with subordinate and junior bonds recovering the least (about 27.8%). The recovery rates vary

[6] Metz and Sorensen (2012) show that 30-day post-default bond trading prices are powerful predictors of ultimate recovery.

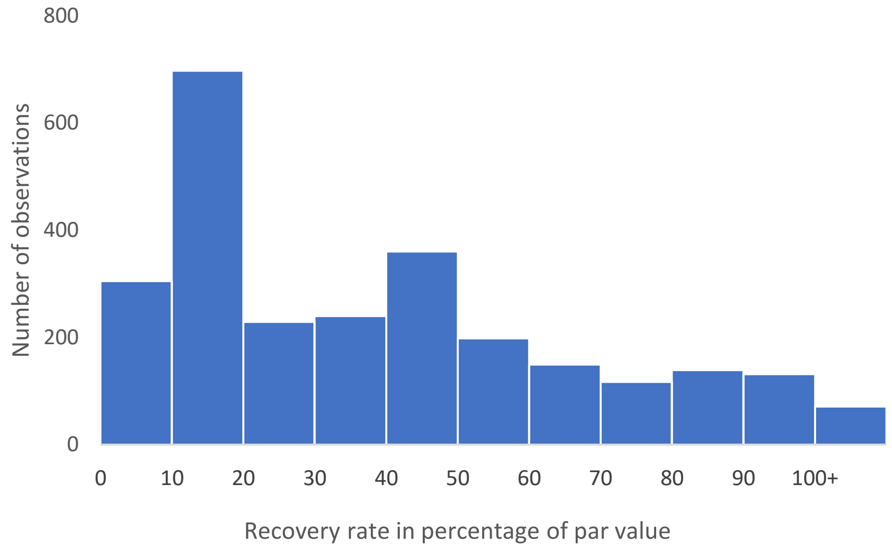


Figure 4.1: Distribution of the recovery rates of the defaulted U.S. corporate bonds from 2004 to 2016.

by industry. Bonds from the utilities sector recover the most, a finding that is consistent with early studies such as Altman and Kishore (1996). The lowest recovery rates are observed in the financials sector. Table 4.2 shows recovery rate summary statistics by bond seniority and industry.

As the TRACE data contains comprehensive bond trading information for all defaulted bonds, we are able to observe bond price changes over time. In Figure 4.2, we show the time variation in trading prices around the default date for all bonds, and by industry affiliation during the 180 days before to 180 days after default. Whereas the average trading prices drop around the default event and follow a V-shaped curve, the price trajectories differ substantially for different industries. The communication services and utilities industries show the least severe price drops of less than 25% of par value, whereas bonds from financials, information technology, and energy industries show a steep price decline, each dropping by more than 50% in par value.

Following Acharya, Bharath, and Srinivasan (2007), we consider two industry distress dummy variables which indicate whether an industry has suffered declines in sales, and whether an index of stocks of firms within the same industry has dropped recently before default. We create these

Table 4.1: Summary statistics of recovery rate distribution by time and default event type 2004–2016. The recovery rate is calculated as the average daily trading price of transactions on the default day and during the 30-day period thereafter as a percentage of par value.

	Bond defaults	**Firm defaults**	**Median**	**Mean**	**SD**
Panel A: Overall recovery rate					
All bonds	2,636	679	34.22	38.83	28.82
Panel B: Recovery rate by year					
2004	124	71	59.80	58.60	29.56
2005	125	51	60.95	58.10	29.15
2006	49	36	74.07	66.96	26.46
2007	41	27	66.83	63.21	33.98
2008	1,010	96	14.60	22.61	20.21
2009	657	151	43.57	43.58	23.20
2010	67	43	65.51	59.92	33.75
2011	93	45	41.49	46.52	30.10
2012	75	48	47.50	51.03	30.71
2013	48	31	43.13	49.21	36.12
2014	86	45	59.82	62.21	33.13
2015	109	54	30.10	36.69	24.32
2016	152	100	39.50	43.52	30.88
Panel C: Recovery rate by default event type					
Distressed exchange	197	95	59.60	59.23	29.75
Default risk rating	306	129	58.74	57.51	27.34
Chapter 11 reorganization	1,428	453	37.59	37.44	27.91
Chapter 11 liquidation	92	61	29.93	38.03	28.66
Default rating	542	35	15.86	26.32	20.64
Chapter 7 liquidation	71	37	11.61	26.31	33.73

measures utilizing industry-specific sales and stock indices growth information from S&P Capital IQ. In addition, we obtain the following information from FISD: a bond's offering amount, the days until maturity, the coupon rate, and a dummy variable that indicates whether the bond contains covenants. Information about the availability of CDS contracts of a given bond is added as a dummy variable, and is retrieved from S&P Capital IQ. Finally, we consider the rating one year prior to default with detailed ratings information that we collect from FISD. Therefore, we encode each rating class as an integer, where the best rating has the lowest ranking value (i.e., AAA = 1, AA+ = 2, etc.).

Firm Fundamentals We consider a range of firm fundamentals as explanatory variables which we retrieve from S&P Capital IQ. We use data from bond issuers' annual reports that were most recently available prior to default. Therefore, we consider profitability (net income as percentage of revenue), total assets and the number of employees. Following Jankowitsch, Nagler, and Subrahmanyam (2014), we furthermore add proxies of structural credit risk, such

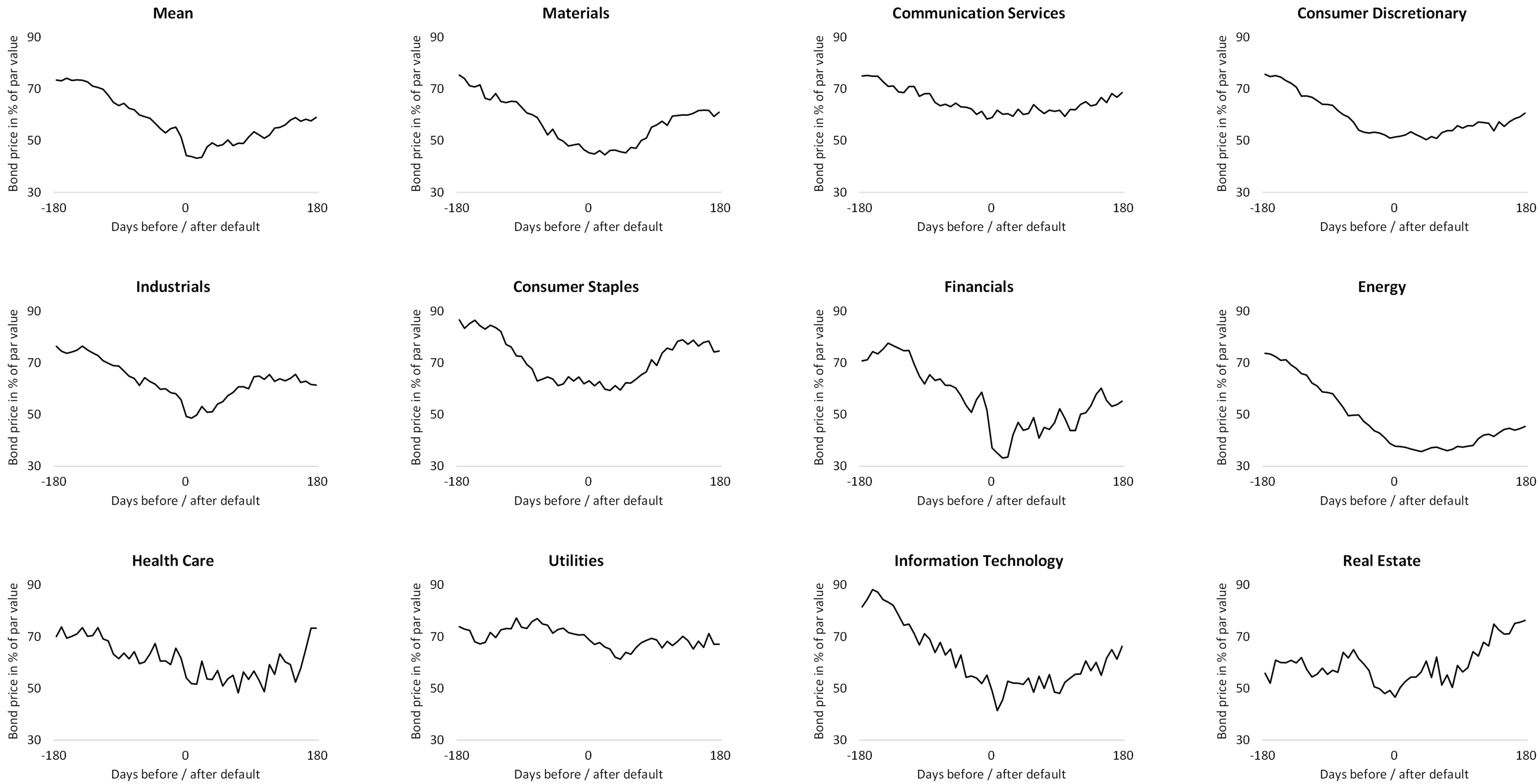


Figure 4.2: Defaulted bond trading prices during 180 days before and 180 days after default by industry based on data reported to TRACE. The figure shows trading prices in fractions of par value.

Table 4.2: Summary statistics of recovery rates 2004–2016 by seniority and by industry. The recovery rate is calculated as the average trading price of transactions on the default day and during the 30-day period thereafter. Panel A shows recovery rate summary statistics by bond seniority. Panel B shows recovery rates by industry classification.

	SD	*Min*	*q25*	*Median*	*Mean*	*q75*	*Max*	*N*
Panel A: Recovery rate by seniority								
Senior Secured	31.5	0.0	40.2	66.5	64.9	92.6	119.6	276
Senior Subordinate	31.3	0.0	13.9	38.6	42.5	67.2	105.4	200
Senior Unsecured	26.2	0.0	13.4	29.2	35.3	49.5	119.8	2,113
Subordinate & Junior	31.5	0.0	1.8	15.9	27.8	54.8	91.7	47
Panel B: Recovery rate by industry								
Materials	29.53	0.05	17.07	40.13	42.97	63.53	109.97	149
Communication Services	34.63	0.00	14.27	49.93	49.39	83.25	108.13	233
Consumer Discretionary	30.73	0.00	19.69	39.91	45.55	71.92	119.80	332
Industrials	28.76	0.01	26.77	44.82	49.62	72.30	106.59	162
Consumer Staples	33.96	3.66	23.81	58.55	57.37	89.77	105.41	44
Financials	22.47	0.01	12.81	18.08	29.79	43.67	101.12	1,265
Energy	25.35	0.00	17.03	30.55	37.61	55.11	103.44	232
Health Care	30.05	2.75	12.01	36.22	42.74	74.65	99.57	31
Utilities	31.64	8.03	45.22	78.72	71.51	100.63	119.62	102
Information Technology	31.20	0.45	10.93	38.98	41.07	65.93	101.68	58
Real Estate	22.83	15.37	35.52	41.78	48.54	57.56	99.85	28

as equity value and default barrier, that measures a firm's debt relative to its assets.

Macroeconomic Variables The macroeconomic conditions at default have been proven to be a key determinant of corporate bond recovery (see, for example, Mora (2015) or Nazemi and Fabozzi (2018)). We follow Jankowitsch, Nagler, and Subrahmanyam (2014) and add the federal funds rate, the slope of the interest rate curve, the default rate, and the industry default rate as macroeconomic predictors of recovery rates. From FRED, we retrieve the federal funds rate and the 10-year minus the 3-month U.S treasury notes yield as the slope. We calculate the default rate as the percentage of default events that we observe in a 90-day rolling window relative to the total number of bonds for which we observe trades in TRACE during that time, and the industry default rate is collected from Moody's.

Liquidity of Defaulted Bonds Jankowitsch, Nagler, and Subrahmanyam (2014) show that a defaulted corporate bond's trading liquidity is correlated with its recovery rates. They find that the higher the liquidity of a defaulted bond is during the 30 days after its default day, the higher is its recovery rate. We replicate their measures by relying on trading data from

TRACE. The average daily trading volume ($Volume$) $v_{i,t}$ of bond i is defined as

$$v_{i,t} = \frac{1}{T+1} \sum_{s=t}^{t+T} \left(\frac{1}{K_{i,s}} \sum_{k_{i,s}} v_{s,k_{i,s}} \right), \tag{4.2}$$

and is measured as the average over the period starting with the default day t to $T = 30$ days thereafter.

The number of trades ($Trades$) $n_{i,t}$ measures the total number of trades executed for bond i during the period starting with the default day t to $T = 30$ days thereafter, with $K_{i,s}$ as the number of reported bond trades on day s:

$$n_{i,t} = \frac{1}{T+1} \sum_{s=t}^{t+T} K_{i,s}. \tag{4.3}$$

The average Amihud (2002) measure $Amihud$ of bond i over the period starting with the default day t to $T = 30$ days thereafter is defined as

$$Amihud_{i,t} = \frac{1}{T+1} \sum_{s=t}^{t+T} \left(\frac{1}{N_{i,s}} \sum_{k_{i,s}} \frac{|r_{k_{i,s}}|}{v_{k_{i,s}}} \right), \tag{4.4}$$

with $N_{i,s}$ as the number of observed returns r on a given day s.

Finally, we implement the price dispersion measure $d_{i,s}$ on day s defined as

$$d_{i,s} = \sqrt{\frac{1}{\sum_{k_{i,s}} v_{k_{i,s}}} \cdot \sum_{k_{i,s}} \left(\frac{p_{k_{i,s}}}{m_{i,s}} - 1 \right)^2 \cdot v_{k_{i,s}}}, \tag{4.5}$$

with $m_{i,s}$ as the mean of the transaction prices reported to TRACE as an estimate of bond i's fair value and $p_{i,s}$ as the bond's individual trading prices. We use the average price dispersion measure $PriceDispersion$ over the period starting with the default day t to $T = 30$ days thereafter.

4.2.2 Financial markets data

In the following, we describe the data and the construction of the key variables that represent conditions in financial markets, which we employ in our empirical analysis for the first time in recovery rate literature. We not only consider market data related to debt instruments, but also conditions in equity markets.

Bond Market Liquidity In addition to a defaulted bond's liquidity implications on its recovery rate, we consider the liquidity conditions in the overall secondary bond market. We utilize a variety of measures that capture different dimensions of bond market liquidity. To do so, we calculate bond market liquidity ourselves by utilizing the trading information in TRACE, but also complement our data by collecting bond market liquidity information available through the literature.

In order to observe the bond market's liquidity at the time of a bond's default event, we consider the most recent available market liquidity observation on the default date. For liquidity that we calculate using TRACE data, this represents the average daily one-month trailing market liquidity, that is, the average of the daily liquidity of all individual bonds traded during the month before a default event. In order to be included for calculating market-wide liquidity using TRACE data, a bond has to be actively traded (see, for example, Reichenbacher and Schuster (2022) or Müller, Reichenbacher, Schuster, and Uhrig-Homburg (2023)), hence, we only include bonds that are traded at least on ten days in a given month. Because we are only interested in the liquidity available to investors, rather than the liquidity of the inter-dealer market, we also remove all inter-dealer trades before we derive daily bond liquidity. In addition, we filter and adjust the data in order to yield a coherent dataset that allows the construction of meaningful measures of market-wide liquidity.[7] By doing so, our final TRACE sample contains 45,777,649 bond transactions of 28,255 corporate bonds over the period 2004-2016.

We then calculate the bond market liquidity measures for all bonds separately, and consider

[7] We remove all bonds that are not available in FISD, and we match data from TRACE with bond characteristics in FISD based on unique CUSIP identifiers. We include only USD denominated bonds, and remove all bonds with a variable coupon rate, and with an issue size below USD 100m. We furthermore remove transactions when a bond's time to maturity is less than one year, and transactions that are related to the initial bond issuance, i.e., the primary market. Finally, we remove all bonds in default. Because TRACE data contains realized prices that include accrued interest, we finally adjust prices based on information of the coupon frequency, coupon rate, and coupon cash and payment in kind components from FISD. This allows us to correctly identify price returns within the data.

their average liquidity as the market liquidity. We consider the $Volume$, $Trades$, $Amihud$, and $PriceDispersion$ defined in Section 4.2.1, using transaction data of the bond market over the month before the default observation. We furthermore replicate the liquidity measures of Corwin and Schultz (2012) ($CorwinSchultz$) and Roll (1984) (*Roll*) that were shown by Schestag, Schuster, and Uhrig-Homburg (2016) to be robust estimators of transaction costs in the corporate bond market. We define $CorwinSchultz_{i,s}$ of bond i on day s as:

$$CorwinSchultz_{i,s} = 2 \cdot \frac{e^{\alpha_{i,s}} - 1}{1 + e^{\alpha_{i,s}}}, \tag{4.6}$$

where

$$\alpha_{i,s} = \frac{\sqrt{2 \cdot \beta_{i,s}} - \sqrt{\beta_{i,s}}}{3 - 2 \cdot \sqrt{2}} - \sqrt{\frac{\gamma_{i,s}}{3 - 2 \cdot \sqrt{2}}}, \tag{4.7}$$

$$\beta_{i,s} = \sum_{j=0}^{1} \left(\log \left(\frac{H_{i,s+j}}{L_{i,s+j}} \right) \right)^2, \tag{4.8}$$

and

$$\gamma_{i,s} = \left(\log \left(\frac{H_{i,s,s+1}}{L_{i,s,s+1}} \right) \right)^2. \tag{4.9}$$

Here, $H_{i,s}$ and $L_{i,s}$ are the highest and the lowest trade prices of bond i on day s, respectively, and $H_{i,s,s+1}$ and $L_{i,s,s+1}$ are the highest and the lowest trade prices of bond i on two consecutive days.

We furthermore define $Roll_{i,s}$ of bond i on day s as:

$$Roll_{i,s} = \begin{cases} 2 \cdot \sqrt{-Cov\left(r_{i,s}, r_{i,s-1}\right)}, & \text{if } Cov\left(r_{i,s}, r_{i,s-1}\right) < 0 \\ 0, & \text{otherwise,} \end{cases} \tag{4.10}$$

with return r_s on day s. Figure 4.3 illustrates the time variation of bond market liquidity measures derived from transaction data in TRACE.

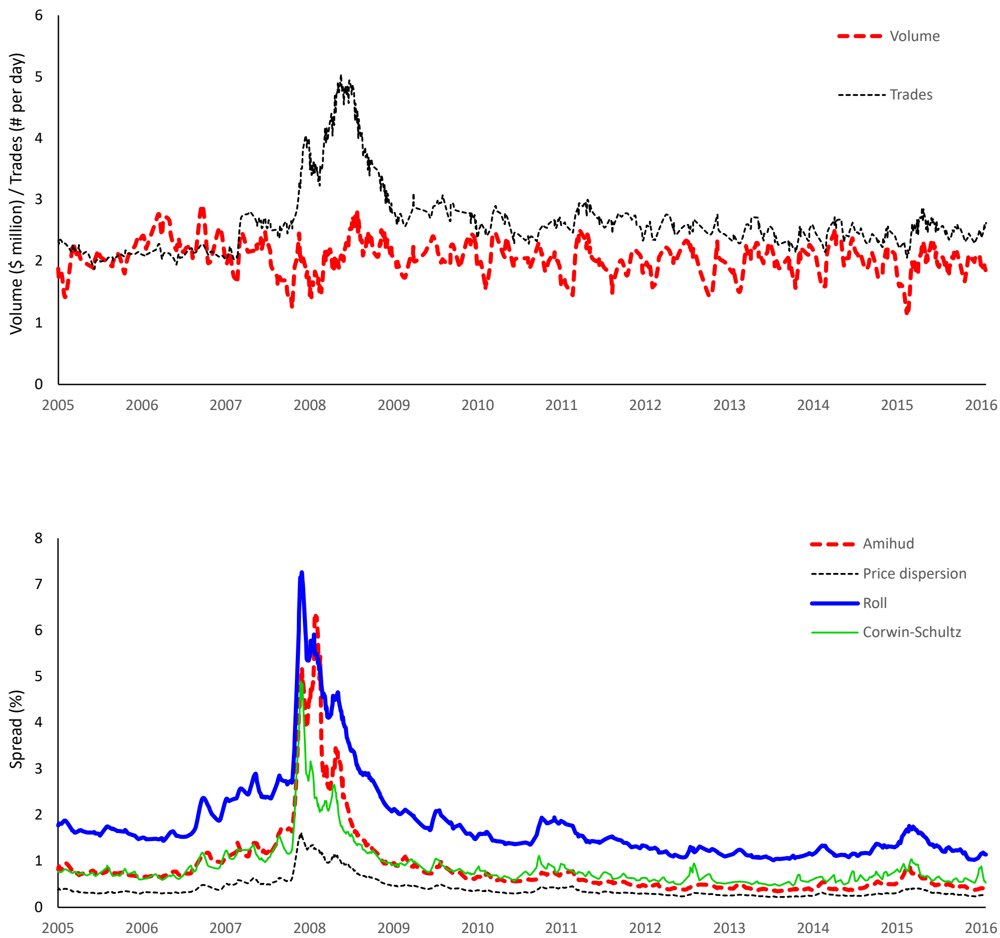


Figure 4.3: Time series of bond market liquidity measures derived from transaction data in TRACE.

In addition, we utilize liquidity data provided in the literature. We obtain the liquidity price and liquidity quantity measures of Goldberg and Nozawa (2021). Liquidity price *LiquidityPrice* represents the "noise" in corporate bonds, that is, deviations from fitted issuer-level yield curves, whereas liquidity quantity *LiquidityQuantity* proxies the liquidity provision by dealers through use of their balance sheets, measuring their gross inventory positions. Reichenbacher and Schuster (2022) examine bond transaction costs subject to trade size. Because time-varying changes in trade size may influence the time variation in liquidity proxies of traditional approaches, they provide size-adapted estimates of the transaction cost measure described by Schultz (2001) (*SizeAdaptedCost*), along with size-adapted estimates of the average bid-ask spread *SizeAdaptedSpread* (see, for example, Hong and Warga (2000)). We obtain these measures from them and employ them in our recovery rate analysis.

Müller, Reichenbacher, Schuster, and Uhrig-Homburg (2023) predict liquidity distributions one month ahead of the majority of individual bonds available in TRACE via machine learning. We obtain the predicted liquidity measures from them, which are the predicted bid-ask spread *PredictedSpread* (see, for example, Hong and Warga (2000)), and the predicted size adapted bid-ask spread of Reichenbacher and Schuster (2022) (*PredictedSizeAdaptedSpread*). Here, we consider the average of the predicted liquidities of individual bonds as the market-wide aggregate liquidity. In total, we utilize twelve corporate bond market liquidity measures, six of which we create using transaction data from TRACE, and six of which we obtain from the existing literature.

Table 4.3: Summary statistics of bond market liquidity measures. Volume is in $ million, Trades is in # trades per day, and others are % spread

	Mean	Standard deviation	10th percentile	Lower quartile	Median	Upper quartile	90th percentile
Volume	1.94	0.36	1.27	1.83	1.90	2.24	2.34
Trades	3.30	0.79	2.38	2.55	3.40	3.96	4.02
Amihud	2.10	1.52	0.48	0.73	1.64	3.93	3.93
PriceDispersion	0.75	0.38	0.29	0.39	0.68	1.25	1.25
Roll	3.24	1.59	1.28	1.69	2.91	5.35	5.35
CorwinSultz	1.62	0.88	0.62	0.85	1.34	2.70	2.87
LiquidityPrice	25.36	14.23	12.65	14.72	24.85	35.02	44.77
LiquidityQuantity	16.69	0.20	16.46	16.55	16.63	16.89	16.99
SizeAdaptedSpread	0.49	0.24	0.19	0.27	0.48	0.69	0.83
SizeAdaptedCost	0.35	0.17	0.13	0.18	0.33	0.58	0.58
PredictedSpread	1.77	0.66	0.85	1.10	1.83	2.56	2.61
PredictedSizeAdaptedSpread	0.50	0.27	0.17	0.25	0.46	0.72	0.92

Corporate Bond Market Distress Index Boyarchenko, Crump, Kovner, and Shachar (2022) create an index that quantifies corporate bond market distress. Their Corporate Bond

Market Distress Index (CMDI) estimates the conditions in the primary and secondary U.S. corporate bond markets, that is, not only capturing bond trading conditions. They aggregate measures of primary bond market issuances, measures of primary market spreads, and measures of volume, liquidity, duration-matched spreads, default-adjusted spreads, and conditions of non-traded bonds for the secondary market. Their index identifies deteriorating market conditions when several of the individual measures indicate distress. We obtain the weekly market-wide CMDI, as well as its subindices for investment-grade and high-yield corporate bond markets, from the Federal Reserve Bank of New York.[8]

National Financial Conditions Index Brave and Butters (2011) use a dynamic factor model with dimension reduction applied on a large number of financial indicators to gauge conditions in U.S. financial markets. Their measure, the National Financial Conditions Index (NFCI), accounts for conditions in money markets, debt and equity markets, and the banking system, and is classified into risk, credit and leverage subcategories. In addition, Brave and Kelley (2017) create an adjusted National Financial Conditions Index (ANFCI) that captures only the financial conditions that are uncorrelated with the prevailing economic conditions. We obtain weekly NFCI, ANFCI and data for the subcategories from the Federal Reserve Bank of Chicago.[9]

Bond Market Risk Factors We consider bond factor models that explain the variation in the cross-section of corporate bond returns:

1. *Traded bond factors*: We collect time series data of U.S. corporate bond risk factors from Dickerson, Julliard, and Mueller (2023) who provide 14 tradable factors from the corporate bond literature.[10] These factors include:

    i. *CRF*, a credit risk factor that measures the return difference between portfolios of low- and high-rated bonds

[8] The CMDI data is available via https://www.newyorkfed.org/research/policy/cmdi (Retrieved on October 25th, 2023).

[9] The NFCI data is available via https://www.chicagofed.org/research/data/nfci/current-data (Retrieved on October 25th, 2023).

[10] The traded bond factors are available via https://openbondassetpricing.com (Retrieved on October 25th, 2023).

ii. *CRY*, a bond carry factor that measures the average return difference of bond portfolios that are double sorted by ratings and credit spreads

iii. *DEF*, a bond default risk factor as the return difference of a long-term corporate bond portfolio and long-term government bonds

iv. *DRF*, a downside risk factor that is defined as the return difference between a high value-at-risk (VaR) and a low VaR portfolio across ratings

v. *DUR*, a duration factor that captures the return difference between high and low duration portfolios across ratings

vi. *HMLB*, a book-to-market factor that captures return differences of bond portfolios double sorted by bond size and book-to-market ratio

vii. *LTREVB* and *STREVB* which are factors that capture long- and short-term reversals that are dependent sorted by return reversals, ratings, and maturities (*LTREVB*) and double sorted by return reversals and ratings (*STREVB*)

viii. *MKTB* (and *MKTBD*) which captures the (duration adjusted) corporate bond market excess return

ix. *MOMB*, a bond momentum factor that captures the return difference across cumulative historic bond returns and ratings

x. *PEADB*, a bond earnings announcement drift factor that captures return differences across high and low earnings surprises, as well as ratings

xi. *TERM*, a bond term structure factor that captures the difference between long-term government bond returns and the one-month treasury bill returns

xii. *VAL*, a bond value factor that captures return differences of 'fair' and actual credit spreads, double sorted by bond value and bond size

2. *IPCA factors*: Kelly, Palhares, and Pruitt (2023) create five bond risk factors using a latent factor approach with time-varying factor loadings to model corporate bond returns. Their approach employs instrumented principal components analysis (*IPCA*) following Kelly, Pruitt, and Su (2021) that aggregates 29 established corporate bond risk factors in

five instrumented principal components. We obtain these factors for estimating recovery rates.[11]

For applying the bond risk factors in recovery rate estimation, we transform monthly observations to trailing quarterly observations in order to not only capturing short term-fluctuations.

Equity Market Risk Factors We also consider equity risk factors that explain stock returns for estimating corporate bond recovery rates. We employ the following equity market risk factor models from the literature:

1. *Fama French*: Fama and French (2015) and consecutive studies such as Feng, Giglio, and Xiu (2020) provide five risk factors in equity markets. Their factors capture the return difference between small and big firms (SMB), firms with high and low book-to-market ratios (HML), firms with robust and weak profitability (RMW), and firms with conservative and aggressive investment practice (CMA).[12]

2. *Factor themes*: Jensen, Kelly, and Pedersen (2023) examine a large set of 153 global risk factors that previous research has shown to explain returns in equities. By clustering and weighting individual factors across themes and countries, they provide 13 unique equity risk factors for the U.S. stock market. These factors represent returns of the top minus the bottom terciles in each of the themes *low risk*, *momentum*, *profit growth*, *profitability*, *quality*, *seasonality*, and *value*, and returns of the bottom minus the top terciles in each of the themes *accruals*, *debt issuance*, *investment*, *leverage*, *size*, *skewness*.[13]

As for bond risk factors, we transform monthly observations of equity risk factors to trailing quarterly observations.

Equity Market Valuations Furthermore, we consider industry-specific equity valuation ratios, which are intended to capture the effect of prevailing valuation levels of comparable companies (i.e., companies that operate within the same industry) at the time of a bond's default

[11] The IPCA data is available via https://sethpruitt.net/2022/03/29/reconciling-trace-bond-returns (Retrieved on October 25th, 2023).

[12] The Fama French factors are available via https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (Retrieved on October 25th, 2023).

[13] The factor theme data is available via https://jkpfactors.com (Retrieved on October 25th, 2023).

on its recovery. We collect time series data of the stock valuations in the U.S. for 48 different industries from WRDS. Specifically, we consider *price / cash flow*, *price / earnings*, *price / operating earnings*, *price / sales*, and *price / book ratios*.

4.3 Empirical Analysis of Recovery Rates

In this section, we first examine how financial market conditions shape corporate bond recovery rates. We start by analyzing the influence of conditions in the corporate bond market, followed by conditions in equity markets. Finally, we examine how different market conditions jointly affect corporate bond recovery rates, and we determine which of the new variables that capture market conditions contribute the most to reliably determining the market conditions' effect on recovery rates.

4.3.1 Does the bond market affect bond recovery rates?

The conditions in the corporate bond market reflect bond pricing and market functioning. For example, the bid-ask spread as a bond market liquidity measure represents the portion of a transaction that is consumed by the execution of the trade (see, for example, Schestag, Schuster, and Uhrig-Homburg (2016)). Time-varying bond risk factors indicate the current pricing conditions in bond market portfolios. We aim to expose how these bond market-wide conditions furthermore drive the pricing of defaulted corporate bonds, that is, the recovery rate.

Bond market liquidity and bond recovery rates

Jankowitsch, Nagler, and Subrahmanyam (2014) show that the liquidity in a defaulted bond influences its recovery. In contrast, we consider the market-wide liquidity to capture not only the idiosyncratic liquidity effects, but also liquidity in the overall bond market, and thus, the prevailing conditions of market efficiency or transaction costs. To begin with, we examine the recovery rate implications of the idiosyncratic defaulted bond liquidity together with bond market liquidity, relying on the four liquidity measures employed by Jankowitsch, Nagler, and Subrahmanyam (2014): trading volume (*Volume*), number of trades (*Trades*), the liquidity defined in Amihud (2002) to which we refer as *Amihud*, and the price dispersion (*PriceDispersion*). Therefore, we employ the following OLS regression:

$$
\begin{aligned}
RecoveryRate = \alpha &+ \beta(\text{Defaulted bond liquidity}) \\
&+\gamma(\text{Bond market liquidity}) \\
&+\delta(\text{Seniority}) \\
&+\zeta(\text{Default type}) \\
&+\eta(\text{Bond characteristics}) \\
&+\kappa(\text{Firm characteristics}) \\
&+\lambda(\text{Macroeconomic conditions}) \\
&+\mu(\text{Industry distress dummies}) \\
&+\nu(\text{Year dummies}) + \epsilon,
\end{aligned}
\tag{4.11}
$$

and we cluster standard errors at the firm-default event level. Table 4.4 presents the regression results. In specification (1), we employ a standard recovery rate estimation model based upon widely established recovery rate predictors, such as characteristics of the bond and the issuer firm, and macroeconomic conditions. These variables include seniority, default type, bond characteristics, firm characteristics, macroeconomic conditions, industry distress variables, and year dummy variables. We also add the liquidity measures of defaulted bonds, to replicate the model of Jankowitsch, Nagler, and Subrahmanyam (2014). Similar to their study, we find that one liquidity measure is significant. The trading volume *Volume* in a defaulted bond is significantly positively related to its recovery rate, and it is in line with the findings of Jankowitsch, Nagler, and Subrahmanyam (2014) that a defaulted bond recovers more when its bond issue experiences more liquid trading immediately after default.

We now benchmark each of the defaulted bonds' liquidity measures with the corresponding liquidity in the corporate bond market. Therefore, we alternate *Volume*, *Trades*, *Amihud*, and *PriceDispersion* in specifications (2)–(5) in Table 4.4. In each of the specifications, we compare the effect of liquidity in the bond market with the effect of the defaulted bond's liquidity. In specifications (2) and (3), we find a positive significant effect of *Volume* and *Trades* in a defaulted bond on its recovery rate. These two measures proxy for the trading intensity. A one standard deviation change in the trading volume or number of trades of the defaulted bonds implies a six percentage point higher, or a one percentage point higher recovery rate of the corresponding defaulted bonds, respectively. The bond market's trading volume or number of trades, however, is insignificant.

In specifications (4) and (5), we consider *Amihud* and *PriceDispersion*, which harness differences in the observed pricing of a bond to gauge its liquidity. While we do not find any significant effect of *Amihud* and *PriceDispersion* of defaulted bonds on recovery rates, we find that the bond market's liquidity proxied by either of these two measures has indeed an economically and statistically significant effect on recovery. A one standard deviation increase in *Amihud* and *PriceDispersion* corresponds to an 11 and 12 percentage point lower recovery rate, respectively. As both measures capture the illiquidity dimension within the bond market, our results show that when the overall bond market illiquidity increases, defaulted bonds recover less. Interestingly, the variation in market liquidity has a more pronounced effect on recovery than the defaulted bond's liquidity, given the larger coefficients that capture the effect of a one standard deviation change of market liquidity on recovery rates.

Finally, in specification (6), we employ all liquidity measures that represent liquidity of the defaulted bonds and of the overall bond market altogether. While the defaulted bond *Volume* remains a significant positive driver of bond recovery, we now find only *PriceDispersion* in the bond market to be significantly correlated with recovery. The signs of the significant measures are as expected, as higher liquidity in the defaulted bond or within the bond market is correlated with higher recovery. Our analysis confirms the relevance of defaulted bond-level liquidity for recovery rates, described by Jankowitsch, Nagler, and Subrahmanyam (2014). Still, it shows that bond-market liquidity also has a determining effect on corporate bond recovery rates.

Having found above that bond market liquidity is a significant determinant of recovery rates, we now examine a wide range of different bond market liquidity measures. The liquidity in bond markets is difficult to quantify (see, for example, Schestag, Schuster, and Uhrig-Homburg (2016) and Hendershott, Li, Livdan, and Schürhoff (2022)). Thus, numerous studies attempt to accurately capture the liquidity from different perspectives. Thus, these different liquidity measures may capture different liquidity dimensions and henceforth provide alternative explanations of the relationship between bond market liquidity and recovery rates. To test this, we utilize a specification that is similar to that defined in Equation 4.11. In particular, we consider specification (1) in Table 4.4 as the basic specification, adding all four bond-level liquidity measures,

Table 4.4: Results of OLS regression that estimates the recovery of defaulted corporate bonds. The table benchmarks liquidity of the defaulted bonds with liquidity of the bond market, utilizing four liquidity measures employed by Jankowitsch, Nagler, and Subrahmanyam (2014). Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by firm-default event. Significance is denoted *** (1%), ** (5%), and * (10%).

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intercept	0.42***	0.41***	0.41***	0.36**	0.36**	0.35**
Defaulted bonds liquidity						
Volume	0.06***	0.06***				0.07***
Trades	−0.01		0.01*			−0.01
Amihud	0.00			−0.01		0.00
PriceDispersion	−0.01				−0.01	−0.01
Bond market liquidity						
Volume		0.00				−0.02
Trades			−0.04			0.05
Amihud				−0.11***		0.04
PriceDispersion					−0.12***	−0.21**
# observations	2.211	2.211	2.211	2.211	2.211	2.211
R-squared	0.563	0.562	0.535	0.539	0.541	0.573
Seniority	Yes	Yes	Yes	Yes	Yes	Yes
Default type	Yes	Yes	Yes	Yes	Yes	Yes
Bond characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic conditions	Yes	Yes	Yes	Yes	Yes	Yes
Industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

as in Jankowitsch, Nagler, and Subrahmanyam (2014), as controls and then examine the effects of a variety of bond market-based liquidity measures. As many of the liquidity measures are highly correlated with each other, and to avoid multicollinearity, we add all bond market liquidity measures recursively in separate specifications. By doing so, we check for the robustness of our findings to the choice of the liquidity measure.

In Table 4.5, the specifications (1)–(4) iterate the bond market liquidity measures introduced in Table 4.4. Here, we do not benchmark the bond market liquidity measures with their respective defaulted bond liquidity measures. We yield consistent results for all measures, and find a negative significant influence of *Amihud* and *PriceDispersion* derived from the bond market on recovery rates, but we find no significant influence of *Volume* or *Trades*. We further add bond market liquidity measures *Roll* following Roll (1984) and *CorwinSchultz* as in Corwin and Schultz (2012), *LiquidityPrice* and *LiquidityQuantity* as provided by Goldberg and Nozawa (2021), and two normalized size-adapted liquidity proxies *SizeAdaptedSpread* and *SizeAdaptedCost* provided by Reichenbacher and Schuster (2022) in specifications (5)–(10). We find similar significant directional effects for all of these alternative liquidity measures, except for *LiquidityQuantity* for which we find a positive significant effect. The findings are consistent, as all measures except for *LiquidityQuantity* reflect the illiquidity in the market. *Roll* and *CorwinSchultz* correspond with lower recoveries, similar as a higher *LiquidityPrice* measure does. However, when *LiquidityQuantity* rises, i.e., dealer liquidity provision through their inventories rises in the bond market, recovery rates benefit. Relatedly, if the size-adapted liquidity in the bond market declines (specifications (9) and (10)), so do the recovery rates of defaulted corporate bonds.

We then consider expected bond market liquidity *PredictedSpread* and *PredictedSizeAdaptedSpread* based upon one-month ahead liquidity predictions performed by Müller, Reichenbacher, Schuster, and Uhrig-Homburg (2023) via machine learning in specifications (11) and (12). Interestingly, we find that even the predicted bond market liquidities have a consistent and significant effect on corporate bond recovery. Finally, in specification (13) we consider the first principal component of all liquidity measures employed in specifications (1)–(12). A one standard deviation increase in the first principal component corresponds to a 14 percentage point lower recovery of defaulted bonds, a relationship that is consistent with the previous findings. Overall, we find that an increase in bond market liquidity is correlated with higher recovery of defaulted bonds, and we find that the variation in bond market liquidity has a more pronounced effect on recovery than the liquidity of the defaulted bonds themselves. All tested bond market liquidity measures, except for the trading volume and the number of trades, are highly significant for explaining recovery rates in our OLS regression specifications.

Table 4.5: Results of OLS regression that estimates the recovery of defaulted corporate bonds. The table benchmarks bond market liquidity proxied by various liquidity measures. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by firm-default event. Significance is denoted *** (1%), ** (5%), and * (10%).

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**
Intercept	0.42***	0.40***	0.34**	0.33**	0.33**	0.33**	0.27*	0.29*	0.33**	0.28*	0.26*	0.26*	0.26*
Volume	0.00												
Trades		−0.03											
Amihud			−0.11***										
PriceDispersion				−0.13***									
Roll					−0.14***								
CorwinSchultz						−0.12***							
LiquidityPrice							−0.06**						
LiquidityQuantity								0.08**					
SizeAdaptedSpread									−0.12***				
SizeAdaptedCost										−0.12***			
PredictedSpread											−0.14***		
PredictedSizeAdaptedSpread												−0.12***	
First principal component													−0.14***
# observations	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211
R-squared	0.563	0.564	0.569	0.572	0.572	0.573	0.566	0.567	0.570	0.571	0.569	0.571	0.572
Seniority	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Default type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bond characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bond liquidity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Do bond market risk factors explain bond recovery rates?

Having shown that recovery rates are dependent on bond market liquidity, we now consider bond risk factors from corporate bond pricing models in the literature. Although these risk factors are commonly used in estimating portfolio returns in the corporate bond market, we aim to derive information about the prevailing conditions in the corporate bond market to explain recovery rates. In the following, we consider bond risk factors employed by Dickerson, Julliard, and Mueller (2023), and by Kelly, Palhares, and Pruitt (2023). We employ the following OLS regression:

$$\begin{aligned} RecoveryRate = &\alpha + \beta(\text{Bond risk factor}) \\ &+\gamma(\text{Defaulted bond liquidity}) \\ &+\delta(\text{Seniority}) \\ &+\zeta(\text{Default type}) \\ &+\eta(\text{Bond characteristics}) \\ &+\kappa(\text{Firm characteristics}) \\ &+\lambda(\text{Macroeconomic conditions}) \\ &+\mu(\text{Industry distress dummies}) \\ &+\nu(\text{Year dummies}) + \epsilon, \end{aligned} \tag{4.12}$$

in which we cluster standard errors at the firm-default event level. The results are shown in Table 4.6. We iterate 14 bond risk factors of Dickerson, Julliard, and Mueller (2023) in specifications (1)–(14), and consider their first principal component in specification (15). In the final specification (16), we further add the bond risk factors created via IPCA by Kelly, Palhares, and Pruitt (2023). We find that the credit risk factor *CRF*, the bond carry factor *CRY*, the bond default risk factor *DEF*, and the bond downside risk factor *DRF* in specifications (1)–(4) are significant and positively related to the recovery rate. These factors measure the excess returns of more risky bonds over less risky bonds during the last quarter from different perspectives. Their positive relation with recovery rates is reasonable. When market conditions were beneficial for risky bonds in the preceding quarter, these beneficial conditions appear to be also reflected in higher pricing of defaulted bonds, which are themselves risky investments.

As such, these bond risk factors seem to capture the market's increased risk tolerance.

The duration factor *DUR* in the specification (5) is also significant and negatively related to recovery. That is, during times when bonds with high sensitivity to interest rate changes outperform bonds with low sensitivity, the recovery rate will be impaired. Although the relationship seems questionable, as bonds with high sensitivity to interest rate changes typically perform well during times of decreasing interest rates, the negative coefficient may capture coinciding deteriorating economic conditions. The *HMLB* factor is found significant and positive in specification (6), and it bears the largest coefficient of 0.12 among the bond risk factors. Because *HMLB* captures the excess returns of bonds of small firms, which are typically more risky, over those of large, less risky firms, it captures the market's increased risk appetite, which is likely to also benefit defaulted bonds. While the long-term reversal factor *LTREVB* (specification (7)) is positive and significant, the short-term reversal factor *STREVB* (specification (8)) is rendered insignificant. The positive relationship between the recovery rate and *LTREVB* suggests that when the return difference between the worst and the best performing bond portfolio has widened, recovery rates benefit. Thus, bonds which have performed poorly over a prolonged historic time period increasingly outperform bonds that have already performed well historically. Relatedly, if the prospects of poorly performing assets improve, this apparently also support the pricing of bonds that recently defaulted.

Next, in specifications (9) and (10), we consider the bond market excess returns over the risk-free rate *MKTB* and a variant *MKTBD* that is adjusted for duration. While both bear a positive sign, only the duration adjusted *MKTBD* is significant. Plausibly, defaulted bonds achieve higher prices during times when the overall bond market experienced positive returns in the preceding quarter. Among the remaining bond risk factors, the bond momentum factor *MOMB* (specification (11)), the bond earnings announcement drift factor *PEADB* (specification (12)) and the bond term structure factor *TERM* (specification (13)) are significant and negatively related to the recovery rate, and the bond value factor *VAL* (specification (14)) is significant and positively related with recovery. The negative effect of *MOMB* is comparable to the long-term reversal factor *LTREVB*, as *MOMB* captures the excess returns of recent outperformers over underperformers, whereas *LTREVB* captures the excess returns of historic underperformers over historic outperformers. Thus, the negative sign of *MOMB* is expected. While for *PEADB*

Table 4.6: Results of OLS regression that estimates the recovery of defaulted corporate bonds. The table benchmarks bond risk factors. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by firm-default event. Significance is denoted *** (1%), ** (5%), and * (10%).

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**
Intercept	0.32**	0.40***	0.41***	0.43***	0.40***	0.37***	0.41***	0.43***	0.43***	0.41***	0.47***	0.34**	0.35**	0.40***	0.39***	0.34**
CRF	0.08***															
CRY		0.11***														
DEF			0.05**													
DRF				0.08***												
DUR					−0.04**											
HMLB						0.12***										
LTREVB							0.08***									
STREVB								0.00								
MKTB									0.03							
MKTBD										0.06**						
MOMB											−0.06***					
PEADB												−0.06***				
TERM													−0.03**			
VAL														0.09***		
First principal component															−0.11***	−0.17***
IPCA1																−0.01
IPCA2																−0.01
IPCA3																−0.01
IPCA4																−0.09***
IPCA5																−0.06**
# observations	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211
R-squared	0.570	0.573	0.566	0.572	0.566	0.575	0.575	0.563	0.564	0.565	0.570	0.574	0.566	0.573	0.573	0.584
Seniority	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Default type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bond characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bond liquidity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

and *TERM* the observed negative relationships with recovery are not immediately conceivable, the positive sign of *VAL* seems reasonable. It captures the excess returns of bonds that have a credit spread that is above a model-implied 'fair' credit spread, thus, it captures that recovery rates rise during times when more risky bonds (i.e., bonds with a higher than 'fair' credit spread) achieve higher returns than less risky bonds. In specification (15), the economically and statistically significant coefficient of the first principal component that combines the risk factors examined in specifications (1)–(14) confirms the relevance of the bond risk factors for defaulted bond recovery.

In the final specification (16), we further add instrumented principal components of Kelly, Palhares, and Pruitt (2023). These factors estimate latent factors and bond's exposure to these factors, potentially better capturing bonds' risk-return characteristics. We find that two of these IPCA factors are significant, in addition to the significant principal component of the previously analyzed bond risk factors. However, the sign and composition of the IPCA factors cannot be reliably interpreted due to their latent characteristics and varying weights of the underlying factors that are transformed when aggregated via instrumented principal component analysis. Nevertheless, our results show that bond market conditions are still captured with these factors and affect corporate bond recovery rates.

Corporate bond market distress and bond recovery rates

In the next step, we investigate the influence of corporate bond market distress, captured by three factors developed by Boyarchenko, Crump, Kovner, and Shachar (2022), which measure overall bond market distress (*CMDI*), and distress in the investment grade (*IG CMDI*), and non-investment grade bond markets (*NIG CMDI*). We employ the following OLS regression:

$$
\begin{aligned}
RecoveryRate = &\alpha + \beta(\text{Corporate bond market distress}) \\
&+\gamma(\text{Defaulted bond liquidity}) \\
&+\delta(\text{Seniority}) \\
&+\zeta(\text{Default type}) \\
&+\eta(\text{Bond characteristics}) \\
&+\kappa(\text{Firm characteristics}) \\
&+\lambda(\text{Macroeconomic conditions}) \\
&+\mu(\text{Industry distress dummies}) \\
&+\nu(\text{Year dummies}) + \epsilon,
\end{aligned} \tag{4.13}
$$

and again we cluster standard errors at the firm-default event level. Plausibly, the results in specifications (1)–(3) in Table 4.7 demonstrate lower recovery rates when any of the corporate bond distress indices rises, and consistent results when employing their principal component in specification (4). Moreover, the coefficients are statistically significant at the 1%-level and magnitudes are comparable, i.e., we do not observe any considerably higher exposure of recovery rates to one of these measures. The results show that bonds recover less when the corporate bond market experiences distress. Importantly, because we consider year fixed effects, industry distress dummies, and macroeconomic factors that capture the economic cycle, the coefficients of the bond market distress measures do not simply capture the time varying economic conditions or the impact of the financial crisis.

Financial conditions in US financial markets and bond recovery rates

The National Financial Conditions Index (*NFCI*) captures conditions in debt markets such as the corporate bond market, but also considers money markets, equity markets, and the banking system through combining various financial market factors (Brave and Butters (2011)). For recovery rate estimation, we consider the *NFCI*, as well as the subcategories related to risk, credit, and leverage conditions, and the *adjusted NFCI* (*ANFCI*) that adjusts the *NFCI* for potential interference from economic conditions in financial conditions. Finally, we also consider the first principal component of these measures as an aggregate measure. We employ the

Table 4.7: Results of OLS regression that estimates the recovery of defaulted corporate bonds. The table benchmarks bond market distress estimated by Boyarchenko, Crump, Kovner, and Shachar (2022). Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by firm-default event. Significance is denoted *** (1%), ** (5%), and * (10%).

Model	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	0.34**	0.33**	0.38***	0.33**
CMDI	−0.11***			
IG CMDI		−0.09***		
NIG CMDI			−0.09***	
First principal component				−0.12***
# observations	2.211	2.211	2.211	2.211
R-squared	0.571	0.570	0.570	0.572
Seniority	Yes	Yes	Yes	Yes
Default type	Yes	Yes	Yes	Yes
Bond characteristics	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes
Bond liquidity	Yes	Yes	Yes	Yes
Macroeconomic conditions	Yes	Yes	Yes	Yes
Industry distress dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

following OLS regression in which we cluster standard errors at the firm-default event level:

$$\begin{aligned} RecoveryRate = {} & \alpha + \beta(\text{Financial conditions}) \\ & +\gamma(\text{Defaulted bond liquidity}) \\ & +\delta(\text{Seniority}) \\ & +\zeta(\text{Default type}) \\ & +\eta(\text{Bond characteristics}) \\ & +\kappa(\text{Firm characteristics}) \\ & +\lambda(\text{Macroeconomic conditions}) \\ & +\mu(\text{Industry distress dummies}) \\ & +\nu(\text{Year dummies}) + \epsilon. \end{aligned} \tag{4.14}$$

Table 4.8 presents the results. As a rise in each of these measures represents a tightening in the financial conditions in U.S. financial markets as measured by the respective underlying factors, we find that in each specification a tightening in the financial conditions is both statistically and economically significant, and correlated with lower recovery rates of defaulted corporate bonds. This is reasonable and comparable with previous findings. When financial markets are under pressure, investors recover less on their defaulted bonds, which is in line with the previous finding on the negative impact of bond market distress on recovery rates. Again, considering year fixed-effects, industry distress dummies and macroeconomic conditions allow us to disentangle the impact of financial market conditions from the impact of the general business cycle. Thus, our results demonstrate that financial conditions in bond and other financial markets matter for recovery rates.

Table 4.8: Results of OLS regression that estimates the recovery of defaulted corporate bonds. The table benchmarks financial conditions captured by indices developed by Brave and Butters (2011) and Brave and Kelley (2017). Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by firm-default event. Significance is denoted *** (1%), ** (5%), and * (10%).

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intercept	0.22	0.26*	0.23	0.22	0.07	0.17
NFCI	−0.16***					
ANFCI		−0.14***				
Risk			−0.15***			
Credit				−0.15***		
Leverage					−0.17***	
First principal component						−0.17***
# observations	2.211	2.211	2.211	2.211	2.211	2.211
R-squared	0.572	0.571	0.572	0.572	0.576	0.574
Seniority	Yes	Yes	Yes	Yes	Yes	Yes
Default type	Yes	Yes	Yes	Yes	Yes	Yes
Bond characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Bond liquidity	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic conditions	Yes	Yes	Yes	Yes	Yes	Yes
Industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Overall, our analysis shows that various characteristics which represent bond market conditions are significant factors for explaining recovery rates. Bond market liquidity, bond risk factors, but also bond market distress and a financial conditions index demonstrate the impact of time-varying bond market conditions on corporate bond recovery rates. In the next step, we investigate whether equity market conditions are also relevant for the determination of recovery rates.

4.3.2 Does the stock market affect bond recovery rates?

When a bond defaults, the bond's risk-return profile changes materially. As long as not in default, corporate bonds typically have a predictable cash flow in the form of stable recurring coupon payments and a fixed principal payment at maturity. Moreover, bond prices are less volatile than equity prices, and difficult to observe due to the decentralized OTC nature of bond markets. Despite a price sensitivity to interest rate changes, the downside risk is, if not due to default, limited, given the higher claim ranking of debt within a firm's capital structure. Nevertheless, bonds are limited on the upside as no payments beyond interest and the principal are expected. Stocks, however, offer both significant up- and downside risk, and while high cash generating businesses or mature companies may offer steady cash flows to investors via dividends, stock valuations are much more volatile and often observable real-time on centralized exchanges.

A default event, however, alters a bond's expected cash flow and risk profile. While bonds do often cease paying coupons when entering default, the default event exposes additional downside risk up to total loss. In addition, defaulted bonds offer a valuation upside in case that the firm can successfully reorganize or if residual claims are higher than initially expected and priced in the market. After all, defaulted debt is often exchanged with equity during the resolution process in many default events. Thus, the risk-return profile of defaulted bonds diverges from that of non-defaulted bonds and at least converges towards the profile of stocks. This motivates us to scrutinize whether prevailing conditions in the stock market are correlated with the recovery of defaulted bonds. More precisely, we are interested in understanding whether drivers of stock pricing are also relevant for defaulted bond pricing.

Do stock market risk factors explain bond recovery rates?

Fama and French (2015) build a five-factor pricing model to estimate stock returns. Jensen, Kelly, and Pedersen (2023) investigate a multitude of equity risk factors that were deployed in academic research on the drivers of equity returns. Through clustering, they extract 13 themes of factors, condensing 153 individual factors from the literature. We employ these factors for our recovery rate analysis. Thus, we utilize the following OLS regression and we cluster standard errors at the firm-default event level:

$$
\begin{aligned}
RecoveryRate = \alpha &+ \beta(\text{Equity risk factor}) \\
&+\gamma(\text{Defaulted bond liquidity}) \\
&+\delta(\text{Seniority}) \\
&+\zeta(\text{Default type}) \\
&+\eta(\text{Bond characteristics}) \\
&+\kappa(\text{Firm characteristics}) \\
&+\lambda(\text{Macroeconomic conditions}) \\
&+\mu(\text{Industry distress dummies}) \\
&+\nu(\text{Year dummies}) + \epsilon.
\end{aligned} \tag{4.15}
$$

We provide regression results in Table 4.9. We find that eight out of the thirteen themes of Jensen, Kelly, and Pedersen (2023) in specifications (1)–(13) are significantly related to corporate bond recovery. The *Accruals* factor in specification (1), which measures the excess equity returns of companies with small accruals over those with large accruals, captures whether the stock market environment in the preceding quarter offered excess returns for companies that bear lower accruals. That is, companies which collect cash from earnings earlier achieved higher excess returns. While the positive significant relationship with corporate bond recovery is not immediately intuitive, defaulted firms must enforce more strict cash management and timely cash collection, and thus may benefit from the prevailing elevated equity market pricing of firms that do not defer the collection of cash.

While the factors *Debt issuance*, *Investment*, and *Low leverage* in specifications (2)–(4) are not significant, although they are related to investors' views on the debt capital structure of firms, we find that the *Low risk* factor in specification (5) is economically and statistically significant and negatively related to recovery rates. When excess returns of low risk equities increased in the preceding quarter, the recovery rates of defaulted bonds decrease. The relation is plausible, as it implies that when we observe high excess returns of more risky equities, we would also observe higher recovery rates of defaulted bonds. Thus, the equity market's compensation for owning risky assets translates into similar directional pricing effects of defaulted bonds. In fact, the *Low risk* factor bears the largest coefficient among the equity risk factors, and thus a variation in the equity market's excess returns has the largest effect on defaulted bond pricing among the examined equity risk factors.

In specifications (6)–(9), we find that the *Momentum*, *Profit growth*, *Profitability*, and *Quality* factors are all significant and negatively related with recovery. Whenever the equity market's excess returns increased over the last quarter for portfolios of firms that showed valuation increases, superior profit growth, relatively high profitability, or high quality, the recoveries of defaulted bonds are lower. This pattern appears plausible, as distressed companies typically perform poorly in these metrics, thus, when the market places more value on them, distressed firms' defaulted bond prices decline. Vice versa, if the excess returns of such portfolios decrease, indicating a growing indifference of the market with regards to these factors, recoveries are higher.

Among the remaining factors, the *Skewness* (specification (11)) and *Size* (specification (12)) factors are significant, but *Seasonality* (specification (10)) and *Value* factors (specification (13)) are not significant. When equities with low *Skewness* experienced excess returns during the last quarter, bonds recover less. Under these circumstances, assets with asymmetric returns could be perceived as more risky, similar to defaulted bonds. Thus, the negative coefficient appears plausible. The *Size* factor is positively related to recovery rate, and it represents that if returns of smaller firms outperform those of large firms, defaulted bonds will recover more. As smaller firms are often attributed with higher risk, it may represent that more risky assets experience more demand, which would also benefit defaulted bond pricing.

We now investigate the five factors of Fama and French (2015). In specifications (14), we find that recovery rates are higher when the stock market's excess returns grew over the last quarter, given the significant and positive coefficient of *MktEquity*. The *HML* factor in specifications (15) rises when value stocks outperform growth stocks. We find that this factor is significant and positively related with recovery rates. Contradicting with the positive coefficient, a rise in the *HML* factor could be related to the increasing preference of investors for lower risk investments. Alternatively, and in line with the positive coefficient, prices of bonds of firms in distress are often pushed downwards below fair value due to temporary selling pressure. Thus, when a distressed firm is able to perform a successful turn-around, this situation offers opportunities for investors that prefer value over growth, and demand for these defaulted bonds is plausibly higher under these market conditions. The significant *SMB* factor of Fama and French (2015) in specification (16) represents a similar market characteristic as the size factor of Jensen, Kelly, and Pedersen (2023) discussed above, and the sign of both factors is consistent.

The *RMW* factor tracks excess returns of firms with robust profitability over firms with weak profitability. Similar to the other factors related to firm profitability, we find that the *RMW* factor in specification (17) is significant and negatively related to recovery rates. That is, defaulted bonds recover less when equity markets placed more value on firms with stable profitability within the most recent quarter, and thus, we see lower bond recoveries when markets are more risk averse with regards to firms' profitability profile. In specification (18), we find the *CMA* factor tracking return differences in firms with high and low investment activity insignificant, similar to the closely related investment factor of Jensen, Kelly, and Pedersen (2023) in specification (3).

Finally, the principal component of all factors capturing prevailing pricing conditions in equity markets is significant in specification (19). Overall, our results show that the recovery rates of defaulted bonds are in fact related to equity risk factors, and that the observed relationships are reasonable to the degree that equity market conditions are also reflected in the pricing of defaulted bonds, which are themselves assets with equity-like risk-return profiles.

Equity industry valuations and bond recovery rates

Finding equity risk factors significant for estimating defaulted bond recovery in the previous section, it comes natural to also examine equity market valuations in estimating recovery rates. While the previous section considers stock returns, we now investigate whether prevailing market valuations of stocks are significant in our recovery rate regression models. We consider *Price / cash flow*, *Price / earnings*, *Price / operating earnings*, *Price / sales*, and *Price / book* ratios in equity markets. More precisely, we obtain these valuation metrics for stocks of firms that operate within the same industry as a defaulted firm. By doing so, we capture the average equity valuations of companies that perform comparable business activities to the respective issuer firm of a defaulted bond. The following OLS regression is utilized:

Table 4.9: Results of OLS regression that estimates the recovery of defaulted corporate bonds. The table benchmarks equity risk factors. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by firm-default event. Significance is denoted *** (1%), ** (5%), and * (10%).

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Intercept	0.39***	0.42***	0.42***	0.42***	0.32**	0.40***	0.42***	0.39***	0.33**	0.41***	0.41***	0.41***	0.42***	0.28*	0.37***	0.46***	0.36**	0.40***	0.35**
Accruals	0.04***																		
Debt issuance		0.00																	
Investment			−0.01																
Low leverage				0.00															
Low risk					−0.11***														
Momentum						−0.04**													
Profit growth							−0.02**												
Profitability								−0.06***											
Quality									−0.05***										
Seasonality										−0.01									
Skewness											−0.02**								
Size												0.06***							
Value													0.01						
MktEquity														0.10***					
HML															0.03**				
SMB																0.04***			
RMW																	−0.05***		
CMA																		−0.02	
First principal component																			−0.10***
# observations	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211	2.211
R-squared	0.569	0.563	0.563	0.563	0.576	0.565	0.565	0.573	0.573	0.564	0.566	0.572	0.564	0.575	0.566	0.567	0.568	0.564	0.574
Seniority	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Default type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bond characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bond liquidity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic conditions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

$$
\begin{aligned}
RecoveryRate = \alpha &+ \beta(\text{Equity industry valuation}) \\
&+\gamma(\text{Defaulted bond liquidity}) \\
&+\delta(\text{Seniority}) \\
&+\zeta(\text{Default type}) \\
&+\eta(\text{Bond characteristics}) \qquad (4.16) \\
&+\kappa(\text{Firm characteristics}) \\
&+\lambda(\text{Macroeconomic conditions}) \\
&+\mu(\text{Industry distress dummies}) \\
&+\nu(\text{Year dummies}) + \epsilon,
\end{aligned}
$$

and we cluster standard errors at the firm-default event level. As presented in Table 4.10, the defaulted bonds recover more when comparable companies listed on U.S. stock exchanges are priced higher with respect to the valuation metrics employed in specifications (1)–(5), or their principal component in specification (6). While each measure relates the prevailing industry-wide stock valuation to different financial performance characteristics, such as cash flow or earnings, they all exhibit a positive and significant relationship with defaulted bond recovery. Thus, we find that the recovery rates of defaulted bonds are indeed depending on the prevailing stock valuation levels of companies that operate within the same industry.

Our analysis across different dimensions of the conditions of U.S. bond and equity markets shows that the recovery of defaulted bonds is indeed related to the prevailing market conditions. Not only the bond market, but also the stock market has an influence on investors' recoveries. We show that bond market liquidity, bond risk factors, bond market distress indices, financial market conditions indices, equity risk factors, and equity valuations play an important role in explaining corporate bond recovery rates.

4.3.3 Which variables matter?

Having established the relevance of bond- and equity-market variables, we now examine the importance ranking of these variables for explaining recovery rates. We apply permutation importance ranking described by Altmann, Toloşi, O.Sander, and Lengauer (2010) in order to

Table 4.10: Results of OLS regression that estimates the recovery of defaulted corporate bonds. The table benchmarks industry-wide equity valuation metrics. Non-binary explanatory variables are normalized with center 0 and standard deviation 1. Standard errors are adjusted for heteroskedasticity and clustered by firm-default event. Significance is denoted *** (1%), ** (5%), and * (10%).

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intercept	0.44***	0.46***	0.44***	0.45***	0.34**	0.45***
Price / cash flow	0.03*					
Price / earnings ratio		0.03**				
Price / operating earnings ratio			0.03**			
Price / sales				0.04***		
Price / book					0.06***	
First principal component						0.05***
# observations	2.211	2.211	2.211	2.211	2.211	2.211
R-squared	0.565	0.566	0.566	0.568	0.573	0.570
Seniority	Yes	Yes	Yes	Yes	Yes	Yes
Default type	Yes	Yes	Yes	Yes	Yes	Yes
Bond characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Bond liquidity	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic conditions	Yes	Yes	Yes	Yes	Yes	Yes
Industry distress dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

understand the importance of each group of the newly introduced variables. By permuting the response vector, this algorithm allows to obtain robust importance estimates for different variables or subgroups of variables.

We evaluate groups of variables for the following bond and equity market characteristics: corporate bond market liquidity, corporate bond market distress, national financial conditions, bond market risk factors, equity market risk factors, and equity market valuations. In addition, we consider the commonly used recovery rate explanatory variables in groups for bond seniority, industry and industry distress, default event type, bond characteristics, firm fundamentals, macroeconomic variables, defaulted bond liquidity, and year.

For examining permutation importance, we first randomly split the data into a training set that contains 70% of the observations, and a test set that contains the remaining 30%. In the training set, we train a random forest via grid search across 5-folds in cross validation for hyperparameter tuning. The trained random forest is then applied for permutation importance estimation on the test set. We scale the importance results so that the importance of the most important variable group equals to one. In Table 4.11, we present the results.

We find that bond characteristics and bond liquidity are the two most influential variables

Table 4.11: Permutation importance ranking for groups of recovery rate explanatory variables. Variables derived from financial markets are highlighted in bold.

Rank	Variable group	Permutation importance
1	Bond characteristics	1.000
2	Bond liquidity	0.919
3	**National financial conditions**	**0.511**
4	Firm fundamentals	0.483
5	**Equity market risk factors**	**0.390**
6	**Bond market risk factors**	**0.273**
7	**Equity market valuations**	**0.174**
8	**Bond market liquidity**	**0.123**
9	Macroeconomic conditions	0.051
10	Default type	0.044
11	Seniority	0.043
12	**Bond market distress**	**0.015**
13	Industry affiliation and industry distress	0.012
14	Year	0.003

for explaining recovery rates. The national financial conditions in debt markets, money markets, equity markets, and the banking system rank third, higher than the defaulted firms' fundamentals, showing that in our dataset of defaulted bonds, market conditions matter even more than firm fundamentals for determining recovery outcomes. Interestingly, equity market risk factors are ranked fifth, and their importance score indicates they are even more important than bond market risk factors that follow ranked sixth. Equity market valuations are ranked seventh, and are more important than bond market liquidity, that are ranked eighth. Since the equity market risk factors, bond market risk factors, equity market valuations, and bond market liquidity measures are ranked considerably higher than most of the remaining commonly used recovery rate drivers, we find that financial market conditions, captured from different dimensions in bond and equity markets, are indeed important determinants for explaining recovery rates.

The other used variables, such as macroeconomic conditions, default type, bond seniority, industry affiliation and industry distress, and default year, bear much lower importance rankings. Bond market distress measures are ranked as the third lowest group of variables. Comparing the cumulative importance of variables derived from financial markets (1.49) and the established recovery rate explanatory variables (2.55), we find that the cumulative importance ranking of market-derived variables is approximately 60% relative to that of the established recovery rate drivers. Our results show that corporate bond recovery rates depend to a large degree on prevailing conditions in stock and bond markets rather than just on firm and bond characteristics, and macroeconomic conditions. Thus, deteriorating financial markets impair investors' recovery on investment, highlighting the importance of considering market conditions for understanding the formation of corporate bond recovery rates.

4.4 Conclusions

To improve the understanding of recovery rates as one of the key parameters in credit risk, we investigate the hybrid nature of defaulted bonds which share characteristics of both stocks and bonds. Therefore, we consider both conditions in bond and equity markets for estimating defaulted bond recovery rates, which is a novelty in the literature. Capturing the relationships between bond and equity investors' risk-return preferences and corporate bond recovery rates gives interesting new insights for both regulators and practitioners.

While controlling for a variety of alternative explanations, we show that recovery rates of defaulted bonds are driven by prevailing conditions in both stock and bond markets. Equity and bond risk factors, which are established in the pricing of stocks and non-defaulted bonds, are significant drivers of recoveries in defaulted bonds. Our analysis shows that the traditional asset pricing factors capture pricing implications in defaulted bonds, given that defaulted bonds are assets that share characteristics of both stocks and bonds.

The importance of defaulted bond liquidity on recovery outcomes has been demonstrated previously (see, for example, Jankowitsch, Nagler, and Subrahmanyam (2014)). In contrast, our study shows that liquidity in the overall bond market affects recovery rates of defaulted bonds. Our finding demonstrates that not only the liquidity provision in the defaulted bond itself is a relevant recovery determinant, but that the overall functioning of the bond market, captured by bond market liquidity proxies, affects recovery outcomes significantly. In addition, bond market distress, general conditions in U.S. financial markets, and prevailing valuation levels of stocks affect recovery rates.

We further draw conclusions from applying an importance ranking methodology. As we find that market-derived variables represent about 60% of the importance of traditional recovery rate drivers, our analysis highlights the necessity to consider the prevailing conditions in equity and bond markets to obtain adequate recovery estimates.

Overall, we renovate the approach to recovery rate estimation in credit risk by including determinants derived from financial markets, capturing the market conditions' effects on recovery rates. Our study sheds new light on the integration of debt and equity markets by exposing the pricing implications of equity-like and bond-like characteristics of the defaulted bond securities. As such, our study provides interesting insights into distressed debt pricing subject to overall financial market conditions.

Chapter 5

Intertemporal Defaulted Bond Recoveries Prediction via Machine Learning

This chapter is joint work with Abdolreza Nazemi and Frank J. Fabozzi.[1] It was published in 2022 as: *Intertemporal defaulted bond recoveries prediction via machine learning* in European Journal of Operational Research (EJOR), Volume 297, Issue 3, Pages 1,162-1,177.[2] Preliminary explorations of some sections of this paper were published in Heidenreich (2019).

5.1 Introduction

The determinants of recovery rates play an important role in the valuation of default risk insurance. Moreover, the advanced internal ratings-based approach under the Basel Accord II and III allows financial institutions to use their own estimates for credit risk parameters. Consequently, accurate and reliable estimates for recovery rates are needed. Although studies by practitioners and academics have investigated recovery rate determinants of defaulted corporate bonds and loans, as well as alternative prediction methods for estimating recovery rates, these

[1] Abdolreza Nazemi is with the School of Economics and Management, Karlsruhe Institute of Technology, Germany. Frank J. Fabozzi is with the Carey Business School, Johns Hopkins University, Baltimore, USA. We gratefully thank Dr. Konstantin Heidenreich whose support and inputs were very valuable for completing this study.

[2] Nazemi, Baumann, and Fabozzi (2022), available via https://doi.org/10.1016/j.ejor.2021.06.047

studies seldomly fully account for the time variation of recovery rates.

Recent studies have examined out-of-sample or in-sample settings to analyze the determinants of recovery rates.[3] According to Kalotay and Altman (2017), the applicability of conventional out-of-sample estimation to the field of recovery rate prediction is questionable. In particular, the k-fold cross-validation method, the commonly used performance measure for evaluating the predictive accuracy for the recovery rate, has shortcomings. For the k-fold cross validation method the dataset is randomly divided into k subsamples. Each subsample is used for out-of-sample prediction once, while the remaining k-1 subsamples are used for training. The performance measurement is the average of the predictions for the k-th subsample.

Even though conventional out-of-sample estimation has been established as the standard procedure in academia, only out-of-time prediction is feasible in real-world applications of corporate debt recovery rate prediction. While it is acknowledged that out-of-sample estimation makes a distinction between training and testing data for recovery rate prediction, it suffers from two main shortcomings. First, as the dataset is randomly partitioned for out-of-sample estimation, it is virtually inevitable that defaults of bonds used for training the model have occurred after defaults of bonds used for testing the model. Consequently, out-of-sample prediction assumes the data-generating process to be time-invariant, leading to a look-ahead bias in the predictions. The second shortcoming is that out-of-sample prediction implicitly makes the questionable assumption that recovery rates of two defaulted bonds issued by the same company are independent of each other. For instance, when two bonds from the same issuer have defaulted at the same time, only for out-of-time estimation it is ensured that these two bonds are either both in the training set or both in the test set, such that the recoveries in the test set are independent from the training set.[4]

In this paper, we address these shortcomings by comparing a wide range of statistics and machine learning methods – inverse Gaussian regression, random forest, sparse power expectation propagation, and support vector regression – not only for out-of-sample but also for out-of-time

[3] See, for example, Qi and Zhao (2011), Yao, Crook, and Andreeva (2015) and Nazemi, Heidenreich, and Fabozzi (2018).

[4] In this respect, recovery rates of corporate bonds and loans are different to recovery rates of consumer credit as these have a less time-varying distribution and the interdependence of multiple defaults is a minor aspect. Betz, Kellner, and Rösch (2021) investigate how default resolution times impact final loss rates of loans.

prediction of recovery rates on defaulted corporate bonds. We predict out-of-time by ensuring that only sample points observed before the default event were used during the training process (see Section 5.5.3 for more details).

The literature on recovery rates has identified both cross-sectional factors that are time-invariant and systematic factors with a time-varying dependency as determinants of recovery rates. In order to create better models of recovery rates, new insights into the drivers of the time variation of recovery rates are needed.[5] We extend the sparse existing research devoted to modeling recovery rates' time variation. Therefore, we include text-based measures extracted from front-page articles published in *The Wall Street Journal* as independent variables. By including these text-based measures, we consider aggregate uncertainty about future economic conditions prevalent at the time of default. As bond prices shortly after default represent expected recovery, these are driven by investors' expectations about future cash flows. These, in turn, are subject to uncertainty, hence the time-variant distribution of recovery rates is ultimately connected to economic uncertainty. Furthermore, we compare selection techniques such as stability selection, MC+ algorithm, and SparseStep algorithm for selecting a subset of macroeconomic variables from a large set of macroeconomic measures in order to identify those which are most closely related to the recovery rate. This study is the first to compare these econometric and machine learning methods in empirical finance.

Our primary contribution to the recovery rate literature is threefold. First, in addition to presenting a machine learning framework for out-of-sample recovery rate prediction, we evaluate the intertemporal prediction performance of a wide range of parametric and non-parametric techniques across various out-of-time prediction setups. Surprisingly, these have attracted less attention in the literature than out-of-sample prediction techniques. Our findings demonstrate that machine learning techniques deliver superior predictive performance compared to traditional techniques not only out-of-sample but also out-of-time. Second, this study is the first to apply sparse power expectation propagation for modeling the recovery rate. The best out-of-time prediction accuracy is achieved using a sparse power expectation propagation approach, outperforming support vector regression-based approaches by Yao, Crook, and Andreeva (2015) and Nazemi, Heidenreich, and Fabozzi (2018) published in this journal. Third, our study in-

[5] See, for example, Doshi, Elkamhi, and Ornthanalai (2018) and Kalotay and Altman (2017).

cludes news-based measures that have been extracted and categorized with machine learning techniques as an alternative group of independent variables to account for the time variation in recovery rate estimation. By incorporating news-based variables, we show that these variables are significant drivers of recovery rates.

Besides these main contributions, this study is the first to benchmark several selection techniques from the machine learning and econometrics literature for out-of-sample identification and selection of the most informative macroeconomic variables for recovery rate prediction from high-dimensional data. To the best of our knowledge, this study is the first to investigate a ranking of the groups of all independent variables including bond characteristics, seniority, news-based, industry, and seven groups of macroeconomic variables. This ranking, from most informative to least informative, is based on the groups' permutation importance for predicting recovery rates of U.S. corporate bonds with the random forest method. It provides the interesting insight that some of the most informative variables for recovery rate prediction have attracted less attention from previous research than their importance suggests.

We organize the remainder of the paper as follows. A review of the literature is presented in Section 5.2. In Section 5.3 we describe the modeling techniques and selection algorithms we applied. We describe the data we used in Section 5.4 and present our empirical results in Section 5.5. Our conclusions are provided in Section 5.6.

5.2 Literature Review

Altman and Kishore (1996) show that the defaulted debt from public utilities (70%) and chemical, petroleum, and related products (63%) exhibits the highest average recovery rates. Moreover, they find that after controlling for seniority, the original rating of a defaulted bond has no impact on the recovery rate. Altman, Brady, Resti, and Sironi (2005) find that default rates, seniority, and collateral levels are important determinants of recovery rates of corporate bonds. Focusing on the macroeconomic determinants of recovery rates, they find that while there is a significant negative relationship between realized default rates and recovery rates, other macroeconomic variables such as the growth rate of the gross domestic product and the return of the stock market have only a weak correlation with the average recovery rate. Acharya, Bharath, and Srinivasan (2007) document that creditors recover less if the industry of the defaulted firm is in distress. In particular, they show using a dataset for the years 1982 to 1999 that defaulted corporate bonds in distressed industries exhibit 10% to 15% lower average recovery rates.

Altman and Kalotay (2014) introduce a modeling approach based on mixtures of Gaussian distributions conditioned on borrower characteristics, instrument characteristics, and credit market conditions. They show that the forecasts generated by this method are more accurate than parametric regression-based forecasts during out-of-time estimation. In an in-sample study, Jankowitsch, Nagler, and Subrahmanyam (2014) examine the recovery rates of defaulted bonds while paying special attention to the trading microstructure around various types of default events. They find in an in-sample analysis that (1) trading volume in the 30 days after the default is high while trading activity decreases after this time period and (2) bond characteristics (e.g., coupon and covenants) and CDS availability have a significant impact on market-based recovery rates.

Jansen, Das, and Fabozzi (2018) and Schläfer and Uhrig-Homburg (2014) use the term structure model for the recovery rate of credit default swaps. Calabrese and Zenga (2010) suggest a beta regression model for the estimation of bank loan recovery rates. Hartmann-Wendels, Miller, and Töws (2014) forecast recovery rates on a dataset of defaulted leasing contracts provided by three German leasing companies. In their study, model trees outperform regression-based approaches out-of-sample. They emphasize the importance of out-of-sample estimation for appropriate risk management. Yao, Crook, and Andreeva (2017) suggest incorporating a two-stage modeling framework to predict recovery rates of credit cards. Krüger and Rösch (2017) study the downturn loss-given-default employing the quantile regression technique for both in-sample and out-of-sample estimation. Hurlin, Leymarie, and Patin (2018) apply six models for modeling LGD of almost 10,000 defaulted Brazilian credit and leasing contracts. Cheng and Cirillo (2018) investigate a nonparametric survival approach to estimate the recovery rate and recovery time of private loans.

Mora (2015) argues that macroeconomic conditions do matter for recovery rate prediction. She shows how recovery rates in different industries are impacted by macroeconomic conditions in different ways. Studies such as Varma and Cantor (2005), Acharya, Bharath, and Srinivasan (2007), Qi and Zhao (2011), Jankowitsch, Nagler, and Subrahmanyam (2014), and Yao, Crook, and Andreeva (2015) use only a few macroeconomic variables. Nazemi, Heidenreich, and Fabozzi (2018) report that models for estimating recovery rates significantly outperform by adding principal components derived from 104 macroeconomic variables. Nazemi and Fabozzi (2018) investigate the relationship between recovery rates of corporate bonds and macroeconomic variables out-of-sample. They implemented the least absolute shrinkage and selection operator (LASSO) for determining the most relevant macroeconomic variables from a comprehensive macroeconomic dataset applied to recovery rates. The models including the macroeconomic variables selected by LASSO outperform the models including a few macroeconomic variables which are typically used in the literature on recovery rates. In our study, we expand their work by comparing the empirical performance of several selection techniques.

The literature documents time variation of corporate bond recovery rates and demonstrates that the factors which explain recovery rates depend on both cross-sectional features, e.g. bond-specific characteristics, as well as on systematic features, e.g. macroeconomic time series data.[6] Doshi, Elkamhi, and Ornthanalai (2018) note that there is a general agreement among academics that corporate debt recovery rates are time-varying, but the empirical work on this characteristic is limited. Chen (2010), highlighting the dependence of recovery rates' time variation on the effects of systematic variables such as the economic cycle, states that the average values of recovery rates during recessions (1982, 1990, 2001, and 2008) are smaller than during economic upswings. Bruche and Gonzalez-Aguado (2010) argue that in recessions more firms default while the average recovery rate decreases. They propose an econometric model incorporating the credit cycle as an unobserved Markov chain to account for time variation in the probability of default and the recovery rate. Baumann and Nazemi (2023) demonstrate the relation of corporate bond recovery rates to time-varying stock- and bond risk factors. Gambetti, Gauthier, and Vrins (2019) employ the concept of economic uncertainty in recovery rate prediction and find it to explain a larger fraction of the systematic variation of recovery rates than

[6] Several studies emphasize the time-varying characteristic of corporate debt recovery rates. Varma and Cantor (2005) mention considerable time-variation for recovery rates of defaulted bonds and loans. Acharya, Bharath, and Srinivasan (2007) describe the time-series behavior of recovery rates of defaulted loans and bonds over the period 1982–1999. Jankowitsch, Nagler, and Subrahmanyam (2014) mention that recovery rates exhibit substantial variation over time. Mora (2015) evaluates recovery rates of bonds, loans and preferred stock within the time-dependent cyclicality of economic conditions.

other time-variant proxies for the business cycle. The authors explain the relationship between uncertainty and recovery by arguing that recovery, measured by the bond price shortly after default, represents investors' expectations about future cash flow, which in turn depends on economic conditions and is therefore subject to uncertainty. In their study, they employ different uncertainty measures and find that increasing economic uncertainty is significantly connected to decreasing recovery rates even when controlling for the business cycle. One of their explanatory variables, an economic policy uncertainty measure (EPU) constructed by Baker et al. (2016), considers news-based uncertainty that captures the monthly number of newspaper articles containing specific expressions related to economic policy uncertainty. Baker, Bloom, and Davis (2016) create this measure through human audit, analyzing over 12,000 newspaper articles by hand and manually selecting policy-related expressions.

We incorporate five text-based news measures created by Manela and Moreira (2017), who emphasize the possibility of explaining asset price fluctuations with time-varying levels of uncertainty that are contained within news articles. They create their measures by extracting information from front-page articles published in *The Wall Street Journal* using machine learning techniques. Their approach is based on the co-movement of word frequencies with option-implied volatility. Using *WordNet* and *WordNet::Similarity* for classification, they construct five interpretable word categories. Whereas Baker, Bloom, and Davis (2016) exclusively consider news which specifically contain the words "uncertainty" or "uncertain" and therefore create a measure that, by intuition, has a negative connotation, the methodology of Manela and Moreira (2017) doesn't capture any predefined sentiment. Because the news categories which they identify are not unique to uncertainty from economic policy concerns, their measures account for more diverse sources of time-varying uncertainty. They further argue that their approach does not depend on human interaction or judgment in the expression selection process, and is therefore more objective than the EPU measure. While the text-based news measures are able to capture uncertainty that is priced by the stock market, EPU does not.[7]

Compelled by the sparse literature on out-of-time estimation of recovery rates, Kalotay and Altman (2017) investigate the time variation of recovery rates.[8] They report that paramet-

[7] Cortes and Weidenmier (2019) also find no significant association of EPU with stock volatility.

[8] Besides this paper, Doshi, Elkamhi, and Ornthanalai (2018) estimate a time-varying recovery rate of credit default swaps.

ric methods outperform the non-parametric methods for intertemporal prediction of recovery rates. Comparing cross-sectional and intertemporal predictive performance they conclude that machine learning techniques such as the regression tree fail to outperform traditional techniques such as inverse Gaussian regression in an intertemporal setting. Further, applying conditional mixture models, they improve estimates of expected credit losses by taking the time variation of the recovery rate distribution into account. A fast maximum-likelihood approach for the estimation of conditional mixtures of distributions is employed in their analysis.

Bastos (2014) illustrates how ensembles of models derived from the same regression method yield more accurate forecasts of recovery rates than a single model. In particular, using bootstrap aggregation (bagging) to build an ensemble of regression trees, he shows that his results are valid for both corporate bonds and loans both during out-of-sample estimation and cross-validation. Qi and Zhao (2011), Yao, Crook, and Andreeva (2015), Nazemi, Fatemi Pour, Heidenreich, and Fabozzi (2017) and Nazemi, Heidenreich, and Fabozzi (2018) report that non-parametric techniques such as regression trees and support vector regressions outperform parametric methods for predicting recovery rates of corporate bonds in an out-of-sample study. Table 5.1 provides an overview of the prevalence of recovery rate model validation techniques in recent studies for U.S. corporate bonds. The k-fold cross-validation is the most frequently used method for out-of-sample evaluation and has been established as standard procedure in recovery rate estimation. The k-fold cross-validation has two mains drawbacks described in Section 1. None of these studies make comparisons between machine learning techniques and statistical models with attention to intertemporal forecasting performance, and with the exception of Kalotay and Altman (2017) who compared statistical models with just regression trees. Our principal contribution relative to the literature applies yet unrecognized machine learning techniques for intertemporal analysis of U.S. corporate bonds' recovery rates.

Our study is closest to the study by Kalotay and Altman (2017) and Nazemi and Fabozzi (2018). We provide five main contributions as compared to previous studies. First, our paper investigates the ability of parametric and non-parametric models to predict recovery rates for corporate bonds in several out-of-time prediction setups, as well as out-of-sample. In contrast to Kalotay and Altman (2017), we find that machine learning techniques also outperform in in-

Table 5.1: Overview of recovery rate models for U.S. corporate bonds in the literature

Author(s)	Data	Model(s)	In-sample	Out-of-sample	Out-of-time
Frye (2000)	Corporate bonds (1983-1997)	Conditional model	yes	no	no
Altman, Brady, Resti, and Sironi (2005)	Corporate bonds (1982-2002)	Univariate and multivariate, logistic, logarithmic and linear regression	yes	no	no
Varma and Cantor (2005)	Bonds and loans from c. 1,100 corporate issuers (1983-2003)	Univariate and multivariate regression	yes	no	no
Acharya, Bharath, and Srinivasan (2007)	1,511 loans and bonds of over 300 firms (1982-1999)	Multivariate regression	yes	no	no
Bruche and Gonzalez-Aguado (2010)	2,000 bonds (1974-2005)	Markov chain	yes	yes	no
Chava, Stefanescu, and Turnbull (2011)	Corporate bonds (1980-2008)	Linear, logit and probit	yes	yes	no
Jacobs and Karagozoglu (2011)	Corporate loans and bonds (1985-2008)	Beta-link generalized linear model	yes	yes	no
Qi and Zhao (2011)	3,751 defaulted bonds and loans (1985-2008)	Regression, fractional response regression, transformation regressions, tree, neural network	yes	yes	no
Altman and Kalotay (2014)	4720 debt instruments, of which 60% are bonds (1987-2011)	Parametric regressions, regression trees Bayesian conditional mixture	yes	yes	yes
Jankowitsch, Nagler, and Subrahmanyam (2014)	Corporate bonds, 1,270 default events of 534 firms (2002-2010)	Multivariate regression	yes	no	no
Donovan, Frankel, and Martin (2015)	Several instruments of 347 firms (1994-2011)	Univariate and multivariate regression	yes	no	no
Mora (2015)	4,422 bonds, loans and preferred stock (1970-2008)	Univariate and multivariate regression	yes	no	no
Yao, Crook, and Andreeva (2015)	1,413 corporate bonds (1985-2012)	Linear regression, fractional response regression, SVRs, two-stage model	yes	yes	no
Kim and Kung (2016)	Secured corporate loans and bonds (1989-2009)	Multivariate linear regressions	yes	no	no
Kalotay and Altman (2017)	2,828 non-financial corporate bonds (1987-2011)	Inverse Gaussian regressions, mixture models, regression trees	yes	yes	yes
Nazemi and Fabozzi (2018)	794 corporate bonds (2002-2012)	Linear regression, SVRs, bagging, boosting, LASSO, ridge regression	yes	yes	no
Gambetti, Gauthier, and Vrins (2019)	1,831 corporate bonds (1990-2013)	Beta regression, mixture models, regression trees	yes	no	no

tertemporal prediction. Second, we introduce the sparse power expectation propagation method to the credit risk literature and find that it yields the most compelling results for out-of-time recovery rate prediction, thereby outperforming support vector regression methods which are the most accurate methods applied in Yao, Crook, and Andreeva (2015), Nazemi, Heidenreich, and Fabozzi (2018) and Nazemi and Fabozzi (2018). Third, our paper includes several text-based variables of time-varying uncertainty constructed via machine learning techniques for explaining the recovery rates of U.S. corporate bonds. We rely on five different news categories and extend the uncertainty concept in recovery rate prediction compared to previous work. Fourth, whereas Nazemi and Fabozzi (2018) extensively investigate macroeconomic variables in recovery rate prediction and apply the LASSO for identifying and selecting the most informative macroeconomic predictors, we benchmark the performances of several advanced selection techniques for selecting macroeconomic variables from a large set of macroeconomic variables. Lastly, we investigate the permutation importance of groups of explanatory variables in order to rank the major determinants of recovery rates.

5.3 Corporate Bond Recovery Rate Modeling

For recovery rate modeling, we apply linear regression, inverse Gaussian regression, regression tree, random forest, semiparametric least-squares support vector regression, and power expectation propagation techniques. Our benchmark model is linear regression. In the following, we describe the power expectation propagation technique which we apply in recovery rate modeling for first time in the literature. We provide a description of the other modeling techniques, the macroeconomic variables selection techniques utilized (LASSO, SparseStep, and MC+ algorithm), as well as a variable ranking method in Appendix C. In our recovery rate models, we control for recovery rates determinants with dummy variables for industry, seniority, coupon type, and instrument type. We further use industry distress dummy variables and the news-based measures that can capture uncertainty about the future.

5.3.1 Power expectation propagation

According to Bui, Yan, and Turner (2017), Gaussian processes are flexible distributions over functions that are used for a wide range of applications such as regression, representation learning and state space modeling. They introduce a unifying framework for sparse Gaussian process pseudo-point approximation using power expectation propagation.[9] Their novel approach to sparse Gaussian process regression, a power expectation propagation framework, subsumes expectation propagation and the sparse variational free energy method into a unified framework for pseudo-point approximation.

In particular, if power expectation converges, its updates are equivalent to the original expectation propagation procedure while substituting the Kullback-Leibler divergence minimization with an alpha-divergence minimization. As alpha approaches zero, the power expectation propagation solution becomes the minimum of a variational free energy approach. In contrast, when alpha is equal to one, the solution from the original expectation propagation approach is recovered. Bui, Yan, and Turner (2017) show that their innovative algorithm for Gaussian process regression outperforms both expectation propagation and variational free energy approaches. To the best of our knowledge, our paper is the first to apply sparse power expectation propagation in credit risk.

[9] We use the MATLAB implementation from Bui, Yan, and Turner (2017) of the algorithm for power expectation propagation.

5.4 Data

We merge several data sources such as S&P Capital IQ, Bloomberg, Federal Reserve Bank of St. Louis, and news from front-page articles of *The Wall Street Journal* for analyzing the recovery rate of U.S. corporate bonds in this study. Our initial dataset which consists of 2,080 bonds that defaulted between 2001 and 2016 is retrieved from the S&P Capital IQ database (Capital IQ). Bond data are retrieved from S&P Capital IQ. We consider market-based recovery rates based on bond pricing data 30 days after default which we retrieve from Capital IQ. The bond prices available through Capital IQ are obtained from the Intercontinental Exchange (ICE) and are, depending on availability of respective data sources, based upon evaluations of dealer quotes, live trading levels and trade execution data from the Trade Reporting and Compliance Engine (TRACE).[10] In our sample, there is only one default event observation for each bond. All bonds are denominated in US dollar. Industry variables are retrieved from Bloomberg (BBG). A default event occurs when a company files for a Chapter 11 bankruptcy petition or is assigned a rating of 'D' (meaning that the debtor is in default) or 'SD' (selective default) by Standard & Poor's. The issuers of the bonds in our study were assigned to the following industries: industrial, consumer discretionary, consumer staples, telecommunications, raw materials, utilities, energy, financial services, and information technology.

We exclude one bond from our analysis because the data was corrupt. The remaining 2,079 bonds exhibit an average recovery rate of 45.57% and a sample standard deviation of 35.04% as reported in Table 5.2. We combine the seniority classes junior subordinate and subordinate to one class because these two classes contain the fewest bonds. In general, the expectation (senior creditors have the highest recovery rate) regarding the average recovery rates within the seniority classes is met. Subordinated bonds exhibit the lowest average recovery rate of 8.15%, while senior subordinated bonds have the second lowest average recovery rate. Accordingly, senior secured bonds have the highest average recovery rate of 61.91%. Defaulted bonds from the utility sector have the highest average recovery rate whereas defaulted bonds from the

[10] While Khieu, Mullineaux, and Yi (2012) argue that the recovery rate based on 30-day-prices is biased, Metz and Sorensen (2012) find that the bond price 30 days after default serves as a powerful predictor of the mean and variability of ultimate recovery. Moreover, the 30-day market convention represents the actual recovery for investors who sell timely after default, and it provides the advantage to be observable early after default while ultimate recovery can only be observed after the issuer's emergence from default (see, for example, Mora (2015)).

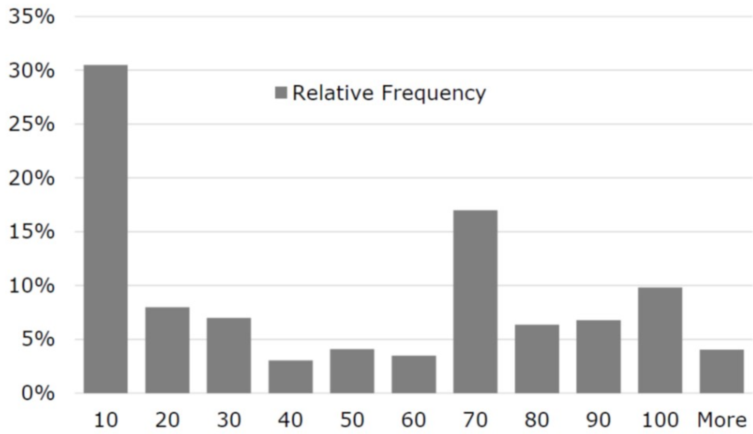


Figure 5.1: Relative frequency of the recovery rates for the defaulted U.S. corporate bonds from 2001 to 2016.

telecommunications sector have the lowest average recovery rate (71.61% vs. 18.54%).

Table 5.2: Descriptive statistics of the recovery rates for all bonds (Panel A) and across seniority classes (Panel B). We report the mean, standard deviation (Std), 10th percentile (p_{10}), first quartile (p_{25}), median, third quartile (p_{75}), 90th percentile (p_{90}), and number of bonds (#).

	Mean	Std	p_{10}	p_{25}	Median	p_{75}	p_{90}	#
Panel A								
All bonds	45.57	35.04	5.00	10.00	43.50	71.96	95.57	2079
Panel B: Recovery rates across seniority								
Senior Unsecured	46.25	34.51	7.50	10.00	48.00	71.00	95.41	1715
Senior Subordinated	24.10	28.34	0.50	2.25	15.50	36.00	72.33	158
Subordinated	8.15	11.98	0.13	0.13	3.00	12.50	18.00	21
Senior Secured	61.91	35.28	5.00	30.00	70.25	94.75	101.15	185

The histogram of the relative frequency of the observed recovery rates in our sample exhibits two peaks and is presented in Figure 5.1. The class between 0% and 10% contains around 640 defaulted bonds. There is another peak of the distribution at the class of values between 60% and 70%. However, the observed distribution does not look like a bimodal distribution.

Evidence for the importance of macroeconomic variables in credit risk management can be found in Bruche and Gonzalez-Aguado (2010), Varma and Cantor (2005), Chava, Stefanescu, and Turnbull (2011), Jankowitsch, Nagler, and Subrahmanyam (2014), Mora (2015), Nazemi, Fatemi Pour, Heidenreich, and Fabozzi (2017), and Nazemi and Fabozzi (2018). We use the database from the Federal Reserve Bank of St. Louis (FRED, Federal Reserve Economic Data) complemented by aggregate default data from Fitch Ratings to retrieve 182 macroeconomic variables used in the credit risk literature such as Acharya, Bharath, and Srinivasan (2007), Varma and Cantor (2005), Jankowitsch, Nagler, and Subrahmanyam (2014), Mora (2015), and Nazemi and Fabozzi (2018). The macroeconomic data covers the time period from 2000 (one year before the start of the recovery rate observation period) until 2016. We list the macroeconomic variables in Appendix C.

A novel data source for recovery rate estimation is news from front-page news articles of *The Wall Street Journal*. To the best of our knowledge, our study is the first study to use any kind of text-based variable constructed via machine learning techniques in recovery rate estimation. News ideally fit intertemporal prediction setups since the data needed can be easily collected immediately by keeping track of the media, whereas the most recent macroeconomic data may not be available at the desired prediction date as it is often published only with a time lag. We choose the text-based news measures that were shown to reflect investors' concerns about the future in a study by Manela and Moreira (2017). They incorporate a measure of the investors' mood that goes beyond commonly used hard data. The relationship between investors' uncertainty and implied volatility is also robust after controlling for realized stock market volatility. Their work is based on the premise that news reflect the interest of readers and that words used by the business press express the concerns of the average investor.

Manela and Moreira (2017) create their text-based measures by first collecting headlines and abstracts of *The Wall Street Journal* articles, then decomposing them into one- and two-word n-grams, and eventually using out-of-sample support vector regression to regress the commonly used implied volatility indices, the CBOE Volatility Index (VIX) and the CBOE OEX Implied Volatility Index (VXO), on monthly normalized n-grams. By doing so, they yield a forward-looking news-implied volatility measure that is capable of serving as a proxy for investor uncertainty. The authors demonstrate this capability by tracing back news headlines and abstracts starting in 1889, then creating a long history of news-implied volatility and showing that their measure peaks in a number of market turmoils and economic crises. In the next step, Manela and Moreira (2017) classify words in the front-page articles to determine the different sources of uncertainty inherent in the news articles. Therefore, they apply commonly used text analytics methods *WordNet* and *WordNet::Similarity*, by which the words are classified according to semantic similarity and relatedness. This step yields five distinct measures from different news categories: government, intermediation, securities markets, war, and unclassified news. The news categories can represent various sources of uncertainty. Manela and Moreira (2017) provide possible explanations, e.g. they associate news from the war category with uncertainty about the destruction of human and physical capital, whereas they relate government news to uncertainty about policy changes. Intermediation news are connected to banking crises and bank failure, and securities markets news to stock price movements.

We further obtain the EPU measure which captures economic policy uncertainty derived from manually screened news articles provided by Baker, Bloom, and Davis (2016) in order to control for the effects from EPU on recovery rates in our analysis. Baker, Bloom, and Davis (2016) rely on ten of the largest U.S. newspapers and therein count the number of occurrences of three policy-related term categories: "uncertainty" or "uncertain", "economic" or "economy", and policy-related terms "Congress", "deficit", "Federal Reserve", "legislation", "regulation", or "White House". The selection of these policy terms is the result of an extensive supervised human audit of over 12,000 news articles, in which the auditors manually code EPU = 1 or EPU = 0, depending on the perceived economic uncertainty. The authors then identify policy terms that appear frequently in the articles coded as contributing to economic policy uncertainty and select the above conceptual term sets.

We merge our dataset with news-based measures that are reported by Baker, Bloom, and Davis (2016) as well as by Manela and Moreira (2017). We use the monthly time series data as a gauge for the investors' uncertainty. In Figure 5.2, we plot the EPU as well as the text-based news measures. The correlation matrix in Table 5.3 demonstrates that there is only limited correlation between EPU and the text-based news variables in our data, suggesting that the latter contain information that is distinct from the information conveyed via the EPU measure.

Table 5.3: Correlation matrix of text-based news variables and the economic policy uncertainty (EPU) measure from 2001 to 2012.

	Government	Intermediation	Securities Markets	War	Unclassified	EPU
Government	1.00					
Intermediation	0.44	1.00				
Securities Markets	0.58	0.43	1.00			
War	-0.11	-0.17	-0.07	1.00		
Unclassified	0.41	0.61	0.57	0.17	1.00	
EPU	0.38	0.19	0.40	0.27	0.68	1.00

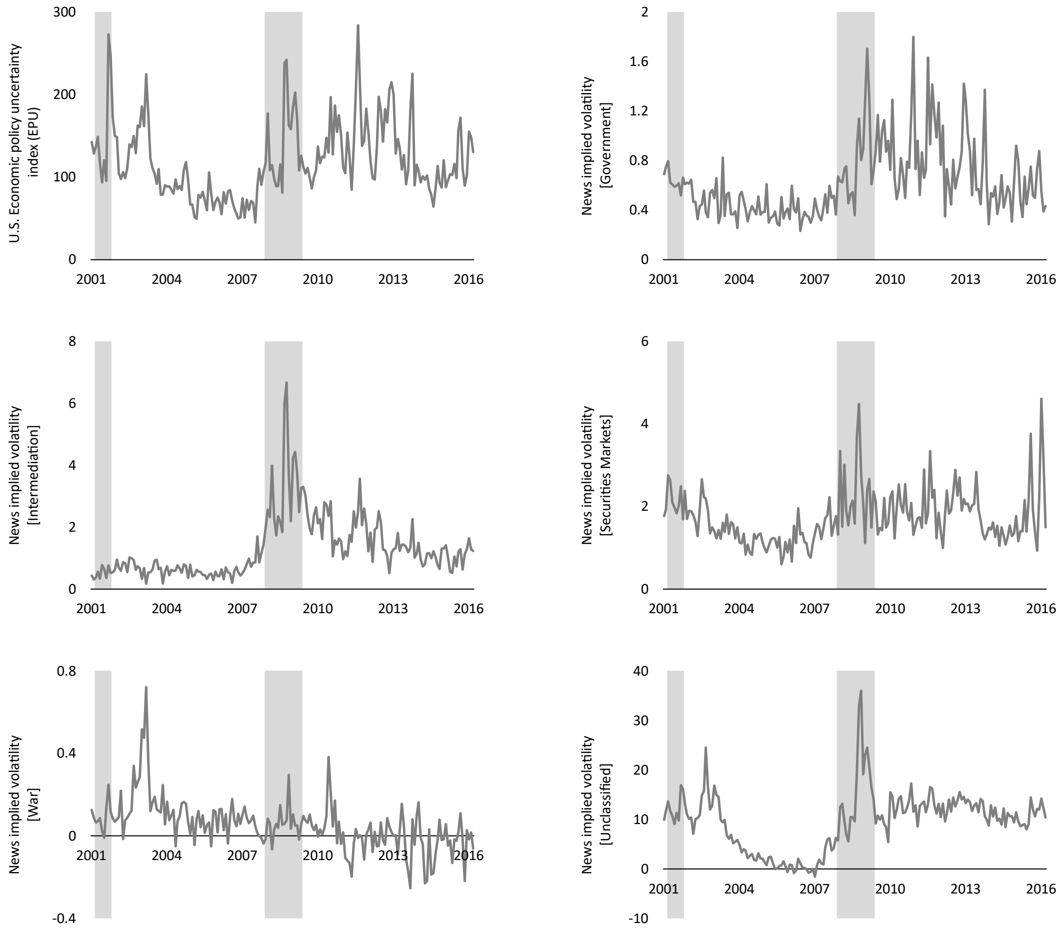


Figure 5.2: Economic policy uncertainty (EPU) and categorized news-based implied volatility during the period 2001–2016. Shaded areas represent U.S. recession periods as defined by the National Bureau of Economic Research (NBER)

5.5 Empirical Analysis of Recovery Rates' Prediction

We first examine the relation between news and the recovery rate. Second, we select macroeconomic variables for recovery rate prediction and benchmark advanced selection techniques. Third, we investigate the recovery rate estimation in out-of-sample and out-of-time (intertemporal) settings. Finally, we rank the groups of explanatory variables by their permutation importance for recovery rate prediction.

5.5.1 Analyzing the news' impact on recovery rates

In Table 5.4 we present an overview of the linear regression specifications based on the entire dataset of 2,079 corporate bonds. The recovery rate of the defaulted U.S. corporate bond is the dependent variable. In model (1), we consider a basic specification including seniority dummies, industry variables, and bond characteristics as independent variables. In model (2), we add the EPU measure from Baker et al. (2016) to the basic specification. In contrast, in model (3) we consider the five news-related measures from Manela and Moreira (2017). In model (4), we use the seven macroeconomic variables selected by stability selection in addition to seniority, industry, and bond variables. Finally, we combine the independent variables from models (3) and (4) in model (5).

Table 5.4: This table presents the results of the linear regression specifications. The recovery rate of the respective bond is the independent variable. In (1) we use seniority dummies, industry variables, and bond characteristics as independent variables. In (2) we add add the economic policy uncertainty (EPU) measure from Baker, Bloom, and Davis (2016). In (3), we replace the EPU with news-based measures. In contrast, in (4) we consider the macroeconomic variables selected by stability selection in addition to the seniority dummies, industry variables, and bond characteristics. In (5) we add the combination of news-based measures and the selection of macroeconomic variables to the base model. The respective t-statistics for each variable are presented in parentheses. Statistical significance at the 99% level is indicated with ***, significance on the 95% level is indicated with ** and significance on the 90% level is marked with *.

Variable	(1)	(2)	(3)	(4)	(5)
Intercept	44.5654***	47.1636***	38.1582***	46.0907***	33.1257***
	(22.4612)	(18.5472)	(11.8707)	(18.5324)	(8.0126)
EPU		-0.0301*			
		(-1.6564)			
Government			31.7482***		39.6178***
			(10.9206)		(11.7446)
Intermediation			-3.579***		-6.1713***
			(-3.2054)		(-3.9478)
Securities Markets			-0.646		-0.822
			(-0.3849)		(-0.4804)
War			4.4547		-11.011
			(0.5826)		(-1.2071)
Unclassified			-1.392***		-0.547**
			(-6.3147)		(-2.2647)
Manufacturers: Inventories to Sales Ratio				61.4796***	50.183**
				(2.9021)	(2.3043)
Number of Civilians Unemployed for Less Than 5 Weeks				-0.0167***	-0.0148***
				(-3.6623)	(-3.1979)
30-Year Conventional Mortgage Rate				8.2292***	10.0959***
				(5.8475)	(7.0106)
3-Month Commercial Paper Minus Federal Funds Rate				-5.7185**	4.2306
				(-1.9842)	(1.3994)
Light Weight Vehicle Sales: Autos & Light Trucks				-0.1929	1.294*
				(-0.3101)	(1.9188)
Nonfarm Business Sector: Unit Labor Cost				-1.4842***	-1.3715***
				(-3.8899)	(-3.5514)
Trade Weighted U.S. Dollar Index: Major Currencies				-1.1194***	-1.2015***
				(-6.6632)	(-7.1092)
Adj. R^2	0.4179	0.4184	0.4536	0.4462	0.4826
RMSE	26.6211	26.6034	25.7603	25.9202	25.0247
MAE	20.7074	20.6564	19.9213	20.1421	19.2381
AIC	1.96E+04	1.96E+04	1.95E+04	1.95E+04	1.93E+04
BIC	1.97E+04	1.97E+04	1.96E+04	1.96E+04	1.95E+04
Number of bonds	2079	2079	2079	2079	2079
Seniority	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes
Bond Characteristics	Yes	Yes	Yes	Yes	Yes

Combining news-related and macroeconomic variables in model (5) generates a further improvement of in-sample fit to an adjusted R-squared of 48.26%. We show the significance of three out of five text-based measures of news even when controlling for the effects of macroeconomic variables in model (5). Similar to Acharya, Bharath, and Srinivasan (2007), industry distress variables have a significant negative impact on recovery rates. In line with the results from Varma and Cantor (2005), bonds with a higher payment rank in the seniority structure recover on average significantly more than bonds with a lower rank. Finally, we confirm the significance of bond characteristics reported by Jankowitsch, Nagler, and Subrahmanyam (2014).

The intuition behind applying news-based variables to recovery rate prediction is that news-based variables, which were shown to reflect investors' time-varying concerns in previous asset pricing research, are likely to also drive the prices of defaulted debt securities. Given that the 30-day bond prices as the recovery rate represent expected future cash flows, we anticipate it to depend on the news that proxy investors' expectations about the future.

Following Gambetti, Gauthier, and Vrins (2019), we add the EPU measure to the basic specification in model (2). Gambetti, Gauthier, and Vrins (2019) argue that the bond price after default as a representation of recovery reflects expectations about future cash flows, and as such is influenced by uncertainty. Furthermore, they emphasize a cause-effect between uncertainty and economic downturns, an economically coherent notion that surfaces in their observations of a significant negative impact of uncertainty on recovery. Similar to their study, we observe a significant and negative impact of the EPU on recovery rates. However, we find only a very marginal improvement of recovery estimation accuracy.

We then replace the EPU with text-based variables as measures for investors' expectations about the future in model (3). By doing so, we include a more diverse universe of uncertainty for recovery rate estimation that relies not only on uncertainty related to economic policy. The estimation accuracy improves by about 3.5 percentage points, showing that news-based variables outperform the EPU in recovery rate prediction.[11] Moreover, we also yield a better prediction power than when using selected macroeconomic variables instead in model (4). A substantial advantage of text-based news variables over macroeconomic variables is not only due to their superior predictive power, but also that news are available immediately whereas macroeconomic

[11] In a linear regression analysis not reported here, but whose results are available from the authors, we combine models (2) and (3) of Table 5.4, controlling for the EPU in the presence of the five news-related measures. In this setting, the EPU becomes insignificant while the variables from government, intermediation and unclassified categories keep their significant relationships with recovery.

data may be published with a time lag, making news-based variables particularly useful in out-of-time predictions. This reveals the substantial importance of news-based variables for predicting U.S. corporate bond recovery rates.

Five categories of text can be identified as distinct origins of uncertainty which we include in our analysis: government, intermediation, stock markets, war, and unclassified. We observe uncertainty related intermediation having a strongly significant negative impact in models (3) and (5). The most frequent word counts in the intermediation category are the following: "financial", "business", "bank", "credit", and "loan". Intermediation-related news spikes mostly during financial crises and periods of bank failures. Thus, the significantly negative impact on recovery rates observed is in accordance with the intuitive expectation of lower recovery rates during times of financial distress.

The news-related variable of the unclassified category has a significantly negative coefficient in the model specifications (3) and (5). The most frequently occurring words of the unclassified category are "U.S.", "Washington", "gold", "special", and "treasury". The occurrence of the terms "gold" and "treasury" points to macroeconomic uncertainty as these assets are often regarded as safe havens. Assuming that recovery rates are lower in an environment with increased macroeconomic uncertainty, this interpretation of the unclassified category would explain the significantly negative coefficient of this source of uncertainty.

News-related to the government category is the only news source that has a significantly positive coefficient in our analysis. The most frequently occurring words of this category are "tax", "money", "rates", "government", and "plan". These terms do not necessarily bear a negative connotation. For instance, the prospects of tax cuts or a more expansive fiscal policy might be reflected in news from the government category. So, the expectation of government policies which are perceived as positive is one possible explanation for the significantly positive impact on recovery rates in our analysis.

Stock market related uncertainty is represented most frequently through the following words: "stock", "market", "stocks", "industry", and "markets". With uncertainty about financial crises already reflected through the highly significant intermediation-related uncertainty, we observe the stock market-related news with a negative but insignificant coefficient in models (3) and

(5). Further, war-related news has very little variance over our observation period and is not a significant determinant in the linear regression analysis.

Similar to Nazemi and Fabozzi (2018), we select a small number of macroeconomic variables from a large collection of macroeconomic variables, but further compare several different selection techniques for this step as described in Section 5.5.2. By selecting a small subset of the macroeconomic variables and eliminating the rest of the variables, the outcome model becomes more interpretable compared to using principal component analysis of the macroeconomic variables in Nazemi, Heidenreich, and Fabozzi (2018). Adding seven selected macroeconomic variables within model (4) presented in Table 5.4, we achieve a prediction improvement over the basic model specification. Nevertheless, the improvement in prediction accuracy is smaller than achieved by Nazemi and Fabozzi (2018). With regards to macroeconomic variables' significance, we observe the following drivers of recovery rates in our analysis. An increase in the inventories-to-sales ratio in the manufacturing industry by 10% coincides with an increase in the recovery rate of 5.0% in model (4). This is contrary to the notion that a lower sales turnover indicates macroeconomic weakness and might lead to a lower recovery rate. Further, in accordance with macroeconomic intuition, the average recovery rate decreases by 1.48% at a time when the number of unemployed civilians rises by 100,000.

Additionally, when inflation of the unit labor cost in the business sector increases by 1%, the average recovery rate decreases by 1.37%. An increase in the 30-year mortgage rate by 1% coincides with a 10%-increase of the recovery rate in model (4). Moreover, when the U.S. dollar strengthens by 1% against a trade-weighted basket of foreign currencies, we observe a decrease of 1.2% in the recovery rate.

We provide more evidence on the effects of text-based news variables by examining how they interact with a recession indicator. In Table 5.5 we present additional regression results including terms for the interaction effects between news-based variables and a recession indicator. Simultaneously, we allow for the respective direct effects of both the recession indicator and the news-based variables. In model (1), we add the recession indicator to a basic specification including seniority dummies, industry variables, and bond characteristics. The recession indicator is a dummy variable indicating recession in the U.S. economy for the periods that are defined as recessions by the National Bureau of Economic Research (NBER). Our dataset covers two recessions according to the NBER definition: March 2001 – November 2001 and December 2007 – June 2009. During these recessions, about 900 of the defaults in our dataset occurred (i.e. 44% of the defaulted corporate bonds). It is worthwhile to mention that while our analysis involving the recession indicator allows us to better understand the economic effects of text-based news variables on recovery rates of defaulted U.S. corporate bonds, it cannot be included into out-of-time predictions because it is only fully determined in retrospect to a recession. We observe a significant negative relationship between the recession indicator and recovery. This is in line with the economic intuition that recovery rates are lower during times of distress. We point out that the industry distress dummy variables are still significant and maintain negative coefficients in this specification, hence we conclude that an economy-wide recession operates as a superordinate factor in addition to industry-specific distress factors.

As we have demonstrated the significance of news-based variables in the first part of this section, we now expand our analysis on the effects from text-based news variables on recovery rates of defaulted bonds by iterating interaction terms between text-based news variables and the recession indicator in models (2)–(6) in Table 5.5. In general, we find that the interaction effects are all positive. Yet, consistent with our previous analysis, news variables and interaction terms related to securities markets and war are insignificant, with the small exception that the interaction between news related to securities markets and recession is significant at the 10% confidence level in model (4). More importantly, we find that the interaction terms in models (2), (3) and (6) are significant with 1% confidence level, i.e. the interactions between recession and those text-based news variables that we found to be significantly related to recovery rate earlier in this section (news from the government, intermediation and unclassified categories). At the same time, the direct effects of news related to government, intermediation and industry distress remain intact in all models. Interestingly, the magnitude of the positive effect from news related to the government in model (2) almost doubles during the recession period as compared to times of economic growth. Apparently, as recovery is expressed by bond prices in our analysis, news related to the government are perceived even more positive by investors during a recession. This is plausible, as substantial actions have been taken by U.S. authorities during the recent global financial crisis in order to mitigate distress, support the economy and provide stability to the financial system. News about these actions will likely lead to increasing investors' confidence.

We also perceive that the positive coefficient of the interaction term of intermediation news and recession (models (3)) has a similar magnitude as the negative coefficient of the direct effect from intermediation news, but is slightly smaller. This shows that during the crisis, intermediation news still negatively affect recovery rates of defaulted U.S. corporate bonds, but to a much smaller extent than during non-recession periods. While the text-based news variable of the unclassified category is significant when including its own interaction term in model (6), it becomes insignificant in most of the other models. Moreover, when comparing the coefficient of the interaction effect of unclassified news with the unclassified news' direct effect, we find that the directional effect of unclassified news on recovery rates turns from negative in times of economic growth to positive during a recession. Although the reasoning behind this is somewhat opaque, it is not implausible given that unclassified news cover the bulk of all news collected, and therefore should include both news that increase concerns but also news that increase the confidence of investors. In summary, the analysis of interaction between news and the recession indicator is an interesting piece of empirical evidence that shows how news can reflect investors' concerns, but are also positively perceived by investors if the news are related to government actions during economic recessions.

We further elaborate whether our observations are rooted in the large fraction of defaulted bonds issued by firms from the financial industry. About 51% of defaulted bonds in our data sample are attributable to the financial industry, hence we remove these bonds and conduct linear regression analysis on the remaining sample, involving 1,020 defaulted bonds that were issued by non-financial firms. In model (1) of Table 5.6, we consider the basic specification that includes seniority dummies, industry variables, and bond characteristics as independent variables. Model (2) adds the EPU measure and confirms its significant negative relationship to recovery rates that has been previously identified for the whole sample. In model (3), we replace the EPU with text-based news variables and observe significant positive effects of news from the government category, and significant negative effects of news from the intermediation and unclassified categories, which is similar to our findings when involving the whole sample. We find, however, that the coefficients of these variables are substantially smaller as compared to

Table 5.5: This table presents the results of the linear regression involving interaction terms between a recession indicator and news-based variables. The recovery rate of the respective bond is the independent variable. In (1) we use seniority dummies, industry variables, and bond characteristics as independent variables, and further add a dummy variable indicating recession in the U.S. economy for periods as defined by the National Bureau of Economic Research (NBER). In (2) we add add the news-based variables and account for the interaction between the recession indicator and the news-based variable from the government category. In models (3), (4), (5) and (6) we iterate the interaction terms of the recession indicator and news from intermediation category, securities market category, war category and the unclassfied category, respectively. The interaction terms and their respective singular compontents are highlighted in bold. The respective t-statistics for each variable are presented in parentheses. Statistical significance at the 99% level is indicated with ***, significance on the 95% level is indicated with ** and significance on the 90% level is marked with *.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	50.8709***	42.3065***	40.0089***	40.9739***	35.7024***	40.1866***
	(19.7543)	(9.5466)	(9.0664)	(8.0498)	(8.5479)	(9.1934)
Government		**26.7863*****	34.9647***	35.8834***	37.5727***	34.8412***
		(6.3399)	(10.0266)	(10.2786)	(11.0248)	(10.0359)
Intermediation		-8.6486***	**-9.6021*****	-7.3071***	-6.2321***	-7.4705***
		(-5.2494)	(-4.9892)	(-4.4309)	(-3.9725)	(-4.6907)
Securities Market		1.3763	0.6433	**-2.0550**	-0.7461	0.5046
		(0.7789)	(0.3651)	(-1.1095)	(-0.4372)	(0.2902)
War		-11.6967	-12.2428	-13.7096	**-12.9651**	-14.9453
		(-1.2937)	(-1.3512)	(-1.5046)	(-1.3721)	(-1.6430)
Unclassified		-0.2945	-0.4609*	-0.3090	-0.2576	**-0.7210****
		(-1.1815)	(-1.7580)	(-1.2176)	(-0.9500)	(-2.4718)
Recession	-17.0835***	**-26.8014*****	**-19.8319*****	**-25.9045*****	**-13.4593*****	**-23.7072*****
	(-6.3700)	(-6.2730)	(-5.5059)	(-3.4451)	(-4.4292)	(-5.5678)
Government*Recession		**22.2482*****				
		(4.1927)				
Intermediation*Recession			**5.2463*****			
			(2.9068)			
Securities Market*Recession				**6.5848***		
				(1.8361)		
War*Recession					**7.4322**	
					(0.3138)	
Unclassified*Recession						**0.9869*****
						(3.2719)
Adj. R^2	0.4567	0.4921	0.4898	0.4885	0.4877	0.4904
RMSE	25.6678	24.7820	24.8370	24.8677	24.8876	24.8234
MAE	19.7555	18.8180	18.8880	18.9178	18.9596	18.8674
AIC	1.94E+04	1.93E+04	1.93E+04	1.93E+04	1.93E+04	1.93E+04
BIC	1.96E+04	1.95E+04	1.95E+04	1.95E+04	1.95E+04	1.95E+04
Number of bonds	2079	2079	2079	2079	2079	2079
Seniority	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Bond Characteristics	Yes	Yes	Yes	Yes	Yes	Yes

using the whole dataset, indicating that the news variables have a smaller effect on non-financial bonds. As *The Wall Street Journal* is predominantly focused on business and financial news, we are not surprised of this finding. Opposite to involving the whole sample, however, securities markets news become significant, having also a negative effect on recovery rates. In summary, the analysis highlights that government related news keep their unique characteristics, having a significant positive effect on recovery rates of defaulted bonds for both samples including or excluding defaulted bonds issued by financial firms. Although the other text-based news variables consistently have negative coefficients, they appear to be more interchangeable with each other, depending on alternations in the data sample.

Table 5.6: This table presents the results of the linear regression considering only defaulted bonds issued by non-financial firms. The recovery rate of the respective bond is the independent variable. In (1) we use seniority dummies, industry variables, and bond characteristics as independent variables. In (2) we add the economic policy uncertainty (EPU) measure from Baker, Bloom, and Davis (2016). In (3), we replace the EPU with news-based measures. The respective t-statistics for each variable are presented in parentheses. Statistical significance at the 99% level is indicated with ***, significance on the 95% level is indicated with ** and significance on the 90% level is marked with *.

Variable	(1)	(2)	(3)
Intercept	37.3171***	45.3711***	45.9041***
	(17.8758)	(14.9568)	(13.5024)
EPU		-0.0693***	
		(-3.6397)	
Government			11.9157***
			(3.2147)
Intermediation			-2.4522**
			(-2.1189)
Securities Markets			-3.6010**
			(2.0810)
War			-2.3412
			(-0.3011)
Unclassified			-0.8354***
			(–3.7856)
Adj. R^2	0.3571	0.3649	0.3885
RMSE	25.3207	25.1551	24.6324
MAE	19.3735	19.1727	18.9089
AIC	0.95E+04	0.95E+04	0.95E+04
BIC	0.96E+04	0.96E+04	0.96E+04
Number of bonds	1020	1020	1020
Seniority	Yes	Yes	Yes
Industry	Yes	Yes	Yes
Bond Characteristics	Yes	Yes	Yes

Overall, taking into account the significance of three out of the five text-based measures even when controlling for macroeconomic effects points to a time-varying influence of investors' mood on recovery rates. The further increase of prediction performance when combining news-related measures with macroeconomic variables demonstrates that the news-based variables contain additional predictive information compared to macroeconomic variables. This is in line with the finding of Gambetti, Gauthier, and Vrins (2019) when controlling for the business cycle in their analysis of uncertainty measures. Hence, we can conclude that the effect measured by the text-based measures has additional predictive power for recovery rates and is not simply mirroring the already known significance of macroeconomic variables for recovery rate prediction. We further find that news related to the government generally has a unique positive effect on recovery rates of defaulted corporate bonds. The effect magnifies during recession periods, a characteristic that possibly accounts for an increasing confidence among investors, conveyed through the news and reflected in 30-day bond prices. While effects from other news categories tend to be negative, their magnitude decreases during recessions. For unclassified news, the effects even become positive, allowing us to conclude that news does not only have a unidirectional effect on recovery rates of defaulted corporate bonds. We also find that financial news has a smaller effect on recovery rates of defaulted bonds from non-financial issuers.

5.5.2 Benchmark of variable selection techniques

In this study, we employ machine learning techniques using two different prediction settings. First, we predict out-of-sample. We sort the dataset randomly stratified for the seniority classes. After using 10 folds cross-validation to select the hyperparameters based on the root mean-squared errors (RMSEs) on the training set (70% of the data), we predict out-of-sample on the test set (30% of the data). By following this procedure, we determine the optimal number of trees and minimum leaf size for the random forest as well as the cost C and the kernel width γ for the SP LS-SVR. We follow the recommendation of Bui et al. (2017) by using α=0.5 for a MSE loss when applying their power expectation propagation approach.

In Table 5.7, we demonstrate that the machine learning techniques outperform traditional statistical techniques during out-of-sample prediction. Using a random partition of 70% of the dataset as the training set, we mitigate the risk of overfitting. In addition to evaluating a wide range of prediction methods, we compare the performance using stability selection, the SparseStep algorithm, and the MC+ algorithm for selecting the most important macroeconomic variables. Without regard to the selection technique used for selecting the macroeconomic variables, all four machine learning techniques (i.e., regression tree, a power expectation propagation approach, SP LS-SVR, and random forest) outperform the two traditional techniques in both performance evaluation metrics, RMSE, and mean absolute error (MAE). Independent of which selection technique is applied, random forest shows the best predictive out-of-sample performance.

Nazemi and Fabozzi (2018) demonstrate that recovery rate models incorporating LASSO-selected macroeconomic variables outperform those from previous research which include only few macroeconomic variables or principal components. We apply and benchmark three different selection techniques, SparseStep, MC+ and stability selection, for identifying the most relevant macroeconomic variables. For all six prediction techniques, macroeconomic variables selected by SparseStep appear to provide the best predictive accuracy. Thus, finding that selecting the macroeconomic variables with SparseStep instead of LASSO increases predictive accuracy, we yield an improvement on the study from Nazemi and Fabozzi (2018). However, the difference between SparseStep and the two remaining selection techniques, MC+ and stability selection is modest.

Table 5.7: This table shows the performance measures from out-of-sample prediction on the testing set which is a random partition of the dataset (30%) while the remaining 70% of the dataset were used for training and determining the hyperparameters during cross-validation. (SP LS-SVR: Semi-Parametric Least-Squares Support Vector Regression; Lin. Reg.: Linear Regression; Reg. Tree: Regression Tree; PEP: Sparse Gaussian Process Approximation with Power Expectation Propagation; RF: Random Forest; IG Reg.: Inverse Gaussian Regression)

	SparseStep		MC+		Stability Selection	
Model	RMSE	MAE	RMSE	MAE	RMSE	MAE
SP LS-SVR	20.9890	13.2027	20.9971	13.5843	21.4105	13.5146
Lin. Reg.	24.8969	18.9199	25.1544	19.2876	25.2331	19.3116
Reg. Tree	22.4956	14.0037	22.5373	14.4637	23.3830	14.8230
PEP	21.2664	14.0712	21.3650	13.8618	21.2667	13.8177
RF	20.6838	13.2145	20.7231	13.2625	21.0394	13.5151
IG Reg.	24.0352	17.9865	24.2890	18.1841	24.4879	18.2376

The lowest RMSE (20.6838) is observed when selecting the macroeconomic variables with SparseStep and using random forest for prediction. Using SP LS-SVR (20.9890) and the power expectation propagation approach (21.2664), the predictive accuracy decreases slightly. The regression tree (best RMSE of 22.4956) has the lowest predictive power of the machine learning techniques. Among the traditional approaches, the inverse Gaussian regression has a minor advantage in predictive capacity compared with linear regression for all three selection techniques. Applying SparseStep for the macroeconomic variables' selection yields the lowest RMSE for the linear regression and inverse Gaussian regression techniques. For this reason, we will use SparseStep during out-of-time prediction.

Even though comparability of performance measures across datasets is limited, our results for out-of-sample estimation are on par with the best results in the literature. The lowest RMSE reported by Yao, Crook, and Andreeva (2015) is 0.2136 for SP LS-SVR during an out-of-sample prediction study. Nazemi, Heidenreich, and Fabozzi (2018) report the lowest RMSE of 0.1750 for LS-SVR with different intercepts for each seniority class during 10 folds cross-validation. The lowest RMSE during 10 folds cross-validation in the study by Kalotay and Altman (2017) is 0.27 for the regression tree.

In summary, during out-of-sample estimation all four machine learning techniques outperform the two traditional approaches (linear regression and inverse Gaussian regression), irrespective of which selection technique is utilized. While this relationship has been documented by Qi and Zhao (2011), Yao, Crook, and Andreeva (2015), Kalotay and Altman (2017), and Nazemi and Fabozzi (2018), the literature on corporate bonds' recovery rate prediction out-of-time is sparse. In the following, we address this gap in the literature.

5.5.3 Intertemporal prediction of the recovery rate

Having predicted out-of-sample in the previous section, we now predict out-of-time. We first evaluate the machine learning methods in out-of-time prediction of recovery of portfolios of defaulted bonds in accordance with Kalotay and Altman (2017), and thereafter predict recovery rates on the instrument-level by applying five out-of-time prediction setups inspired by approaches suggested in asset pricing literature.

Intertemporal prediction of defaulted bond portfolio recovery As cogently outlined by Kalotay and Altman (2017), prediction out-of-time instead of out-of-sample addresses several issues. Taking into consideration the likelihood of time variation in recovery rates, they point out that reporting forecast performance on a random partition of the dataset is less appropriate. Kalotay and Altman (2017) emphasize the importance of accounting for time variation in recovery rates. In particular, testing performance out-of-time ensures that only sample points observed before the default event were used for training. Further, only investigating performance out-of-time prevents data points from the same issuer and the same exposure to be part of both the training and test set, therefore satisfying the condition that observations in the test set are independent from observations in the training set. We train our models by including data from 2001 until 2011 and use data from the remainder of the sample period (2012 to 2016) as the test set. Following Kalotay and Altman (2017) for ease of comparison, we predict mean recovery rates of portfolios of defaulted bonds. Therefore, we draw a sample of 100 bonds from the test set and calculate the average recovery rate on this sample, weighting the bonds equally. This procedure is repeated 10,000 times. We also repeat this analysis moving through time. Starting with the training set from 2002 to 2011, we add another year of data to the training set until we have reached the end of the dataset using training data up to 2014. The bonds from the two years following the training period are used as the test set whereby we sample nine bonds from the respective two-year period and repeat this step 2,000 times.

The out-of-time performance of our models is presented in Table 5.8.[12] The bonds from 2001 to 2011 are used as the training set while the bonds from 2012 to 2016 are used as the test set for sampling. Again, machine learning techniques outperform the traditional approaches for all prediction techniques. In particular, the predictive accuracy of inverse Gaussian regression and linear regression decreases significantly. In contrast to out-of-sample prediction, random forest is the worst performing machine learning technique for out-of-time prediction with an RMSE of 13.5294. Interestingly, in the out-of-time prediction setup, the power expectation propagation approach yields the lowest RMSE of 2.6887 while SP LS-SVR (4.2736) and the regression tree (5.1717) exhibit a slightly lower predictive capacity.

Table 5.8: This table shows the performance measures from out-of-time prediction sampling from the testing set (from 2012 to 2016) while the data from 2001 to 2011 are used for training and determining the hyperparameters during cross-validation. The SparseStep algorithm is used to select the most informative macroeconomic variables. The best performance measures are highlighted in bold. (IG Reg.: Inverse Gaussian Regression; Lin. Reg.: Linear Regression; Reg. Tree: Regression Tree; SP LS-SVR: Semi-Parametric Least-Squares Support Vector Regression; PEP: Sparse Gaussian Process Approximation with Power Expectation Propagation; RF: Random Forest)

		SparseStep					
	Actual	IG Reg.	Lin. Reg.	Reg. Tree	SP LS-SVR	PEP	RF
Mean	32.4095	76.7641	78.2838	37.1176	36.2062	34.1537	45.7951
Std	2.1677	1.3181	1.5030	1.6219	0.7990	0.9622	0.8272
1%	27.4195	73.7258	74.8435	33.3696	34.3637	31.9405	43.9059
		168.88%	172.96%	21.70%	25.33%	16.49%	60.13%
5%	28.8675	74.6184	75.8140	34.4713	34.9157	32.5985	44.4500
		158.49%	162.63%	19.41%	20.95%	12.92%	53.98%
10%	29.6229	75.0592	76.3542	35.0428	35.1853	32.9160	44.7320
		153.38%	157.75%	18.30%	18.78%	11.12%	51.01%
25%	30.9674	75.8629	77.2716	36.0258	35.6595	33.4930	45.2336
		144.98%	149.53%	16.33%	15.15%	8.16%	46.07%
50%	32.4085	76.7598	78.2634	37.1102	36.2066	34.1568	45.7914
		136.85%	141.49%	14.51%	11.72%	5.39%	41.29%
75%	33.8399	77.6559	79.2965	38.2048	36.7370	34.8010	46.3507
		129.48%	134.33%	12.90%	8.56%	2.84%	36.97%
90%	35.2057	78.4484	80.2251	39.1859	37.2386	35.3864	46.8660
		122.83%	127.88%	11.31%	5.77%	0.51%	33.12%
RMSE		44.4119	45.9333	5.1717	4.2736	**2.6887**	13.5294
MAE		44.3545	45.8743	4.7300	3.8397	**2.1923**	13.3856

[12] As we yield the most accurate predictions with SparseStep during out-of-sample prediction, we report only the results applying SparseStep for macroeconomic variable selection during out-of-time prediction. The results using MC+ and stability selection are consistent with the results reported for SparseStep. These results are not reported here but are available from the authors.

In Table 5.9 we show the out-of-time performance of our models when retraining the models each year. Starting with a training set including bonds until 2011, we extend the training set with new bonds each year and use the bonds from the following two years as test set for sampling. For instance, in the first step we use the bonds from 2001 to 2011 as the training set and sample from the bonds from 2012 and 2013 for prediction. In the next iteration, we extend our training set to include the bonds from 2012 and use the bonds from 2013 and 2014 for sampling.

Table 5.9: This table shows the performance measures from out-of-time prediction when all models are retrained every year. Starting with a training set including bonds until 2011 we extend the training set with new bonds each year and use the bonds from the following two years as the test set. So, in the first step we use the bonds from 2001 to 2011 as the training set and sample from the bonds from 2012 and 2013. In the next iteration, we extend our training set to include the bonds from 2012 and use the bonds from 2013 and 2014 as the test set. The SparseStep algorithm is used to select the most informative macroeconomic variables. The best performance measures are highlighted in bold. (IG Reg.: Inverse Gaussian Regression; Lin. Reg.: Linear Regression; Reg. Tree: Regression Tree; SP LS-SVR: Semi-Parametric Least-Squares Support Vector Regression; PEP: Sparse Gaussian Process Approximation with Power Expectation Propagation; RF: Random Forest)

		SparseStep					
	Actual	IG Reg.	Lin. Reg.	Reg. Tree	SP LS-SVR	PEP	RF
Mean	35.4642	65.1487	64.5576	42.6204	37.5660	40.493	46.4706
Std	13.2443	6.5806	6.8764	11.2331	4.1321	8.16991	4.1480
1%	8.7095	49.2976	48.9102	19.8052	28.2487	17.6113	37.4106
		466.02%	461.57%	127.40%	224.34%	102.21%	329.54%
5%	13.8126	53.8581	53.2014	25.3224	30.9192	24.9364	40.0503
		289.92%	285.17%	83.33%	123.85%	80.53%	189.96%
10%	17.3750	56.3814	55.5540	28.7439	32.4446	29.5408	41.3762
		224.50%	219.74%	65.43%	86.73%	70.02%	138.14%
25%	25.7049	60.8349	59.8974	34.5368	34.9132	36.1703	43.6340
		136.67%	133.02%	34.36%	35.82%	40.71%	69.75%
50%	35.8299	65.3895	64.6008	41.6541	37.6385	41.4314	46.3077
		82.50%	80.30%	16.26%	5.05%	15.63%	29.24%
75%	44.5556	69.8571	69.2250	50.2680	40.3493	45.8537	49.1103
		56.79%	55.37%	12.82%	-9.44%	2.91%	10.22%
90%	52.6980	73.5200	73.5126	57.9349	43.0418	49.7292	51.8733
		39.51%	39.50%	9.94%	-18.32%	-5.63%	-1.56%
RMSE		33.0883	32.6534	13.7023	13.1569	**11.7634**	17.2256
MAE		29.7743	29.2172	11.0205	10.6562	**9.41088**	13.9478

Based on the root mean square error (RMSE) and mean absolute error (MAE), the best performing method in Table 5.9 is again the power expectation propagation approach with an RMSE of 11.7634, followed by SP LS-SVR (13.1569) and regression tree (13.7023). However, the prediction performance on the quantiles of the recovery rate distribution offers further insight. While the power expectation propagation approach is the best performing model in terms of RMSE and MAE, it has the lowest percentage deviation among all techniques only for the 1st-percentile, 5th-percentile, and 75th-percentile. In contrast, the regression tree has the lowest percentage deviation for the 10th- (deviating 127.4%) and 25th- (deviating 34.36%) percentiles, while SP LS-SVR has the lowest percentage deviation for the median (5.05%) and the random forest has the lowest percentage deviation for the 90th-percentile (deviating -1.56%).

While Kalotay and Altman (2017) report their lowest RMSE of 6.8 for out-of-time estimation without retraining for a mixture model with bagging, we report lower RMSEs of 2.7 for the power expectation propagation approach and 4.3 for the SP LS-SVR. The result is similar for out-of-time prediction when yearly retraining the models. Each of our three best-performing machine learning techniques – the power expectation propagation approach (RMSE = 11.8), the SP LS-SVR (RMSE = 13.2), and the regression tree (RMSE = 13.7) – outperforms their best-performing technique, a mixture model with bagging (RMSE = 15.5). Performing the comparison on a percentile-level, our best techniques outperform the best techniques reported by Kalotay and Altman (2017) for the median and the higher percentiles (75% and 90%) while for the lower percentiles (1%, 5%, 10%, and 25%) the techniques from Kalotay and Altman (2017) are more accurate.

More traditional approaches such as linear regression and inverse Gaussian regression experience significant deterioration during the out-of-time prediction compared with their out-of-sample performance. In contrast, the predictive accuracy of the machine learning techniques such as the newly proposed power expectation propagation approach and SP LS-SVR does not decline when switching from out-of-sample estimation to out-of-time estimation. From these results we conclude that the non-linear relationships between the recovery rate and the explanatory variables are more stable during our observation period than the linear dependencies between the recovery rate and the explanatory variables. In general, we find that the power expectation propagation approach provides the most compelling out-of-time prediction results.

Although we report the average performance measures across all time steps in Table 5.10, we show the predictive performance for each of the four two-year ahead sub-periods following the respective period used for training each model. Hence, we are able to demonstrate the consistency of our modeling approaches.

Table 5.10: This table shows the performance measures for each two-year ahead subperiod from out-of-time prediction when all models are retrained every year. The first column shows the last year that is included in the training set. # of bonds denotes the number of bonds in each two-year ahead period which is used as test set for sampling. Starting with a training set including bonds until 2011 we extend the training set with new bonds each year and use the bonds from the following two years as the test set. So, in the first step we use the bonds from 2001 to 2011 as training set and the bonds from 2012 and 2013 as the test set. In the next iteration, we extend our training set to include the bonds from 2012 and use the bonds from 2013 and 2014 as test set. The SparseStep algorithm is used to select the most informative macroeconomic variables. The best performance measures are highlighted in bold. (IG Reg.: Inverse Gaussian Regression; Lin. Reg.: Linear Regression; Reg. Tree: Regression Tree; SP LS-SVR: Semi-Parametric Least-Squares Support Vector Regression; PEP: Sparse Gaussian Process Approximation with Power Expectation Propagation; RF: Random Forest)

			SparseStep					
	# of bonds		IG Reg.	Lin. Reg.	Reg. Tree	SP LS-SVR	PEP	RF
2011	75	RMSE	28.4971	29.8182	9.2928	7.8582	8.4563	9.3318
		MAE	27.6984	28.9866	7.3983	6.3000	6.8337	7.6770
2012	62	RMSE	17.8793	16.8617	12.3584	12.1457	9.5560	9.6305
		MAE	15.6871	14.6412	9.9761	9.7989	7.6453	7.7579
2013	91	RMSE	30.3518	28.3460	16.6306	9.5900	11.4688	15.9382
		MAE	28.1792	26.1179	13.5029	7.6879	9.2671	13.6061
2014	104	RMSE	48.2324	47.8334	15.3412	19.7783	16.0980	27.4414
		MAE	47.5327	47.1232	13.2046	18.8379	13.8974	26.7500
Mean		RMSE	31.2402	30.7148	13.4058	12.3431	**11.3948**	15.5855
		MAE	29.7744	29.2172	11.0205	10.6562	**9.4109**	13.9478

Intertemporal prediction of instrument-level recovery rate In the following, we focus on intertemporal recovery rate prediction of individual bonds rather than defaulted bond portfolios. We compare the performance of the machine learning methods for individual bonds' recovery rate prediction across five intertemporal prediction setups. We employ intertemporal prediction setups that have been previously used in the asset pricing literature. The standard approach is to divide the data into a subsample for model training, a validation subsample for hyperparameter selection, and a test subsample for prediction performance evaluation. Bali, Goyal, Huang, Jiang, and Wen (2022) use fixed consecutive time windows for training, validation and prediction testing. Bianchi, Büchner, and Tamoni (2020) use an annual rolling prediction window of fixed size that is preceded by a training sample which increases as they move in time, performing cross-validation on the training set for hyperparameter selection. Gu, Kelly, and Xiu (2020) combine both approaches by using a rolling prediction window of fixed size with an increasing training set, however selecting hyperparameters on a fixed-sized validation set which is located between the training and test sets and which they roll forward as they move in time.

First, we train the models on the data 2002 to 2011 while relying on 10-fold cross-validation for tuning hyperparameters, and then predict on the test set 2012 to 2016. This setup is comparable to that used in Table 5.8 for the portfolio approach. For the subsequent setups, we replace the cross-validation for tuning the hyperparameters with a validation set that is located between the training and the test set. Hence, for the second setup, we split the data into the training set ranging from 2002 to 2010, the year 2011 as the validation set, and the years 2012 to 2016 remain as the test set. Third, we suggest a setup in which we move through time. We start with the years 2002 to 2010 as the training set, 2011 as the validation set and the two subsequent years 2012 and 2013 as the test set. We then move one year in time, increasing the training data by one year but keeping the lengths of the validation and test sets as one year and two years, respectively. The fourth setup is similar to the previous setup except that we also keep the length of the training data fixed and drop older training data , instead of increasing it as we move through time. In the final setup, we apply a daily rolling prediction window that consecutively predicts recovery rates of default-day combinations, i.e. we predict all recovery rates of bonds that defaulted on a given day, and thereafter move on to the next default date in our data on which one or more bonds defaulted. We then retrain the model and predict recovery rates of all bonds which defaulted on that day. For this setup, the validation set considers the 120 most recently defaulted bonds, of which the oldest defaults are added successively to the training set as we move through time, while the last predicted defaulted bonds feed into the validation set.

We denote the training set τ_1, the validation subsample for hyperparameter selection τ_2, and the test set for prediction performance evaluation τ_3. We measure performance on the test sets with RMSE and MAE:

$$\text{RMSE} = \sqrt{\sum\nolimits_{i \in \tau_3} \frac{\left(RR_i - \hat{RR_i}\right)^2}{n}} \tag{5.1}$$

$$\text{MAE} = \sum\nolimits_{i \in \tau_3} \frac{\left|RR_i - \hat{RR_i}\right|}{n} \tag{5.2}$$

where $\hat{RR_i}$ is the out-of-time predicted recovery rate and RR_i the actual recovery rate of bond i, and n is the total number of bonds in the test set τ_3.

The results are shown in Table 5.11. Again, the machine learning methods outperform linear regression and inverse Gaussian regression techniques. Furthermore, consistent with the previous analysis, the power expectation propagation approach performs best and delivers the lowest forecast errors in four of five setups. It yields the best result (RMSE = 23.8) when applied in the setup with a daily rolling prediction window and increasing test set length (Setup (5)). Only in Setup (2), where we apply a fixed prediction window and select hyperparameters during a fixed validation year, SP LS-SVR is performing better than the power expectation propagation approach.

Moreover, we find that for fixed prediction windows, parameter tuning via 10-fold cross-validation across the full historic data (Setup (1)) yields better recovery rate forecasts than using only the last year before the test set as a validation set (Setup (2)). Likewise, the rolling window approach performs better when increasing the training size, considering all historic recovery rate observations (Setup (3)), instead of keeping the length of the training set fixed by adding new training data and dropping old training data while moving through time (Setup (4)). Both of these observations indicate that incorporating the full historic information in calibrating the models is more valuable than calibrating with more recent data. Overall, our analysis demonstrates the benefits of applying the power expectation propagation approach for out-of-time recovery rate prediction.

5.5.4 Permutation importance of groups of explanatory variables

Here we rank groups of variables to elaborate on the degree of feature importance for recovery rate prediction. We investigate the permutation importance according to Altmann, Toloşi, O.Sander, and Lengauer (2010) of each group of variables for the performance of the random forest technique in recovery rate prediction. Therefore, we build 11 groups of independent variables as detailed in Appendix C: industry, bond characteristics, seniority, news, and the macroeconomic variables which are separated into groups (financial conditions, micro-level factors, business cycle, monetary measures, corporate profitability (on a macro level), international competitiveness, and stock market). We scale the permutation importance of each group such that the importance of the most important group of variables equals 100. We examine the im-

Table 5.11: This table shows the performance measures of machine learning methods for five different out-of-time prediction setups. In setup (1), we train the models on the data 2002 to 2011 while relying on 10-fold cross-validation for tuning hyperparameters, and then predict on the test set 2012 to 2016. In (2) we use the year 2011 as the validation set and the years. In (3), we use a rolling prediction window, with 2011 as the validation set and the two subsequent years test set, consecutively moving one year in time, increasing the training data by one year in each step. Setup (4) is similar to setup (3) with the exception of fixed training set length. In setup (5), we apply a daily rolling prediction window that consecutively predicts recovery rates of default-day combinations. For this setup, the validation set considers the 120 most recently defaulted bonds, of which the oldest defaults are added successively to the training set as we move through time. The best performance measures per prediction setup are highlighted in bold. The SparseStep algorithm is used to select the most informative macroeconomic variables. (IG Reg.: Inverse Gaussian Regression; Lin. Reg.: Linear Regression; Reg. Tree: Regression Tree; SP LS-SVR: Semi-Parametric Least-Squares Support Vector Regression; PEP: Sparse Gaussian Process Approximation with Power Expectation Propagation; RF: Random Forest)

Out-of-time prediction setup			IG Reg.	Lin. Reg.	Reg. Tree	SP LS-SVR	PEP	RF
(1)	Fixed window;	RMSE	54.3040	56.1158	29.9666	27.6113	**27.2993**	30.2622
	fixed training length;	MAE	47.0216	48.2788	23.3972	23.4763	**22.0009**	26.2950
	cross-validation							
(2)	Fixed window;	RMSE	55.4472	59.2014	34.3494	**27.9730**	29.0718	31.2301
	fixed training length;	MAE	48.7604	52.2000	26.1498	24.4059	**20.7947**	26.8246
	one year validation							
(3)	Annual rolling window;	RMSE	47.5039	47.4405	29.8983	30.7911	**26.2828**	30.7124
	increasing training length;	MAE	41.0943	41.2887	22.5964	26.0794	**19.9878**	26.1667
	one year validation							
(4)	Annual rolling window;	RMSE	48.2685	49.6850	32.8556	32.9354	**27.6137**	32.8795
	fixed training length;	MAE	41.6908	43.0925	25.7331	28.0902	**22.4910**	28.2036
	one year validation							
(5)	Daily rolling window;	RMSE	51.1180	52.5641	28.0918	26.7455	**23.7689**	31.0743
	increasing training length;	MAE	44.8800	46.6117	21.6381	21.9732	**18.1355**	26.6197
	120 defaults for validation							

portance ranking of groups of variables for the U.S. corporate bonds that defaulted from 2001 to 2016.

As illustrated in Table 5.12, bond characteristics are the most important group of variables for recovery rate prediction in our analysis. So, the significance of bond characteristics reported by Jankowitsch, Nagler, and Subrahmanyam (2014) is confirmed by our study. The importance of the seniority of the defaulted bond (ranked second, 30.7945) is in accordance with the significance of the seniority reported in, for example, Varma and Cantor (2005) and Jankowitsch, Nagler, and Subrahmanyam (2014). The importance of stock market indicators (ranked third, 14.3448) confirms the significance of the return on the market index reported by Varma and Cantor (2005).

Interestingly, the group of text-based news variables is ranked higher than the widely used industry variables (9.5908 compared with 6.1447), which confirms our findings in Section 5.5.1

Table 5.12: Ranking groups of variables by permutation importance for all defaulted bonds from 2001 to 2016

Rank	Entire dataset	Importance
1	Bond Characteristics	100.0000
2	Seniorities	30.7945
3	Stock Market Indicators	14.3448
4	International Competitiveness	13.2206
5	News	9.5908
6	Industry	6.1447
7	Micro-Level Factors	4.5070
8	Corporate Profitability (Macro)	4.3065
9	Financial Conditions	3.4836
10	Business Cycle	2.9321
11	Monetary Measures	2.3154

that this group of variables is an important driver of recovery rates. Similarly, the literature has paid little attention to variables indicating international competitiveness which ranked fourth in our analysis with an importance of 13.2206. Having an importance of 2.9321, business cycle variables which include commonly used variables such as GDP growth and the unemployment rate (see, for example, Altman, Brady, Resti, and Sironi (2005) and Yao, Crook, and Andreeva (2015)) ranked only second to last in our analysis.

Micro-level factors such as the federal funds rate and the term structure reported to be significant by Jankowitsch, Nagler, and Subrahmanyam (2014), Nazemi and Fabozzi (2018), and considered by Qi and Zhao (2011) are ranked seventh with an importance of 4.5070 in our analysis. However, among the macroeconomic variables, micro-level factors constitute the group with the third-highest rank. Financial conditions and monetary measures have not been investigated in the literature but are also not important in our ranking for the entire dataset (3.4836 respectively 2.2154).

The industry of the defaulted bond is reported to be an important determinant of recovery rates by Altman and Kishore (1996). Further, Acharya, Bharath, and Srinivasan (2007) introduce two industry distress dummy variables indicating a negative sales growth of the respective industry and a performance of the industry index worse than -30% in the preceding year. These industry distress dummy variables are part of the industry group in our analysis. In our analysis however, industry variables have an importance of 6.1447 and rank only sixth.

The ranking groups of variables provides insights into which groups of covariates have more information for predicting recovery rates for corporate bonds. Interestingly, groups of variables involving text-based news or international competitiveness, which have been neglected by previous research, have higher importance ranks than industry-factors or macroeconomic variables that have previously been extensively studied by researchers. The finding suggests that these unexplored higher ranking groups of variables provide potentially promising fields of research on the economic mechanisms in recovery rate determination.

5.6 Conclusions

The recovery rate is a key risk parameter in credit risk. Though there is substantial literature on out-of-sample recovery rate estimation for corporate bonds, most approaches employed suffer from two main shortcomings. The assumption of a time-invariant recovery rate distribution is unrealistic. Moreover, assuming the independence of samples when multiple defaulted bonds from the same issuer are part of both training and test set results in unrealistically accurate predictions. Therefore, it is essential to examine the estimation of this risk factor for defaulted U.S. corporate bonds in an intertemporal setting.

In this study, we investigate the prediction of recovery rates for defaulted U.S. corporate bonds over the period 2001-2016 in several intertemporal setups to address these issues. We find that machine learning techniques outperform traditional approaches such as inverse Gaussian regression and linear regression during out-of-time prediction. Employing semiparametric least-squares support vector regression, a power expectation propagation approach, regression tree, and random forest yields significantly higher predictive out-of-time accuracy than the traditional statistical techniques. In particular, the newly proposed power expectation propagation approach achieves the most compelling prediction results under several different out-of-time prediction setups. Interestingly, we also find that out-of-time prediction accuracy benefits from considering a longer history of data for model generation, rather than merely using more recent data and dropping older data points.

We test whether news-implied measures and its five components can predict the recovery rates of corporate bonds. These measures relied on machine learning techniques to uncover information from the front-page coverage of *The Wall Street Journal*. Interestingly, we find that investors' uncertainty about the government, intermediation, and the economy are significant drivers of recovery rates. Government-related news are associated with higher recovery rates, especially during recessions. News that are generally negatively associated with recovery rates tend to be less harmful or even turn out to be supportive for recovery rates in times of economic downturns. We further discover that recoveries of bonds issued by non-financial firms are less impacted by financial news.

We benchmark three techniques for selecting the most informative macroeconomic factors from a broad range of macroeconomic variables. Among the selection techniques examined, the SparseStep algorithm selects those macroeconomic variables which contribute the most to recovery rate prediction accuracy. Lastly, studying the permutation importance of the groups of variables, we find that bond characteristics, seniority dummy variables, and stock market indicators are the most important groups of variables for corporate bonds' recovery rate prediction. However, groups of variables involving text-based news or international competitiveness, which have drawn little or no attention in previous research, appear to be more important in explaining recovery rates than previous studies would suggest.

Chapter 6

Conclusions and Research Outlook

Since the introduction of the Basel II and III accords, Banks in the G20 countries have been encouraged to employ their own IRB approaches for determining regulatory capital and for stress testing. Moreover, buy-and-hold investors in corporate bonds need adequate tools to assess and understand the credit risk associated with their investments. For these reasons, reliable estimates of recovery rates are needed. In this dissertation, the main focus is to improve our understanding and assessment of credit risk in corporate bonds.

While the probability of default (PD) has been researched extensively in the past, and exposure at default (EAD) is investor-specific, the third parameter to estimate credit risk, the recovery rate (RR), is the subject of this research. As the 2007-2008 Global Financial Crisis and the recent COVID-19 pandemic have shown, contagion and cascade effects can induce stress within financial markets and the real economy due to interconnectedness among financial institutions or within globalized value chains. Thus, for adequate risk assessment, practitioners, regulators, and academics need to consider the connectivity within financial markets and economic activities. Moreover, defaulted bonds trade in OTC markets, however, traditional recovery rate models do not account for implications from the bond trading microstructure. In addition, recovery rate estimation models suggested in the literature do not reflect that defaulted bonds share characteristics of both stocks and bonds, and prediction models insufficiently account for the time-varying structure of recovery rates. Using data-science based approaches, Chapters 2–5 address these shortcomings and investigate the formation of recovery rates of defaulted corporate bonds from various perspectives.

Chapter 2 analyzes the intermediation of recently defaulted bonds in the opaque OTC corporate bond market. Because bond default events have a surprise character and initiate the need for an ownership change from buy-and-hold investors to specialized vulture investors, bond dealers play an important role in timely matching supply and demand in the defaulted securities. Chapter 2 utilizes detailed bond transaction data from TRACE to identify dealers and track the flow of defaulted bonds from sellers, through the dealer network, and to buyers. As demonstrated within the chapter, the OTC trading microstructure adjusts endogenously once a bond defaults. The causal impact of trading with primary dealers on the recovery rate is identified and quantified. That is, dealers who are familiar with a given bond excel in locating higher-valuation counterparties, providing trade immediacy and liquidity, and committing their own inventory for facilitating a trade. When investors switch to these primary dealers, they can raise recovery rates by 8% vis-à-vis trading with non-primary dealers. This shows a stabilizing market function of primary dealers in OTC markets, and that access to primary dealers lowers investors' credit risk ex-ante.

In Chapter 3, recovery rates are explained from an inter-industry trade network-based perspective. Because trade relationships between industries within the U.S. economy facilitate the transfer of assets across industry borders, corporate bond recovery rates of firms operating in better connected industries benefit from facilitated asset disposal channels. This benefit magnifies when assets are less specialized and can be utilized by potential owners who perform different economic activities than the defaulted firm. Moreover, the chapter demonstrates that economic shocks propagate through value chains and impair recovery rates in closely connected industries. Finally, macroeconomic conditions have a greater effect on the recoveries in central industries than in peripheral industries. Thus, this chapter highlights the importance of the economy network for the emergence of corporate bond recovery rates.

Chapter 4 acknowledges the hybrid nature of defaulted bonds, which share characteristics of both stocks and bonds. As the default event alters a bond's risk-return profile from normal bond-type to equity-like, this chapter introduces new explanatory variables from stock- and bond-markets for explaining recovery rates. Traditional equity and bond pricing factors are employed, and the provided evidence reveals they are significant drivers of corporate bond recovery rates. Furthermore, bond market liquidity, bond market distress and bond market conditions indices, and prevailing equity market valuation levels have a reasonable and significant relationship with recovery rates. Overall, Chapter 4 demonstrates the importance of prevailing financial markets conditions on recovery rates, capturing the pricing implications of both stock and bond markets on defaulted corporate bonds. Thereby, it also provides new explanations for the integration of stock and bond markets.

Finally, Chapter 5 provides an intertemporal setting for predicting the recovery rate. While most previous studies do not account for the time-varying structure of recovery rates and consider information not available at the time of default for estimating the recovery rate, only such modeling approaches that exclusively rely on historical data can be applied in real-world applications. This chapter benchmarks various machine learning methods in out-of-time prediction settings. Within this setting, the newly proposed sparse power expectation propagation approach performs best. Moreover, models that are trained on data from a longer historical time window perform better than models trained on a shorter, rolling time window. The chapter further introduces text-based news measures that reflect investors' expectations about the future which translate into market-based recovery rates. In summary, it provides interesting and important insights into modeling recovery rates when accounting for its time variation, offering accurate prediction models for real-world applications.

Overall, this dissertation provides new and important explanations on the formation of corporate bond recovery rates, and offers new and more accurate recovery rate prediction models for real-world applications. Nevertheless, additional research beyond the scope of this dissertation remains a task for the future. For example, while the importance of primary dealers in the OTC dealer network for the intermediation of defaulted bonds is presented in Chapter 2, an important question remains: How do these primary dealers obtain their knowledge to provide competitive intermediation services to their customers? Bao, O'Hara, and Zhou (2018) show that the post-crisis introduction of the Volcker Rule, which prohibits proprietary trading of banks, led to a decrease in liquidity provision by dealers for downgraded bonds. A better understanding of OTC markets under stress will enable regulators to introduce more adequate and directed policies for preserving market stability, without unintended side effects.

Furthermore, the datasets employed in this dissertation span over a full economic cycle, including the 2007-2008 GFC. However, as more data becomes available, it is worthwhile to investigate how the shocks to various value chains during the COVID-19 pandemic due to regional lockdowns affected recovery rates. While continuous research allows us to improve our understanding of the formation of recovery rates and to provide more accurate estimations for future real-world applications, the peculiarities of different economic regimes may affect recovery rates in still unknown ways. Hence, future research may assess how the explanations and estimation models proposed in this dissertation perform on yet untested data, such as data that includes the COVID-19 pandemic and subsequent events.

While applying machine learning models for determining the recovery rate improves estimation accuracy vis-à-vis traditional models such as OLS regressions, the state-of-the-art machine learning models are often considered *black boxes*. For many of these types of models, it remains unknown what the model has actually learned. However, regulatory bodies demand model transparency from banks in order to understand how they determine the credit risk parameters. Kellner, Nagl, and Rösch (2022) improve recovery rate model interpretability via linear quantile regression with a neural network structure. Nevertheless, their approach is based on determining feature importance of a black box model rather than directly interpreting the model itself. Thus, more research on directly interpretable machine learning models for recovery rate estimation is desired. A summary of the research background, research steps and contribution of this dissertation, and the outlook on future research, is presented in Table 6.1.

Table 6.1: Summary of key research items.

Background	**Research steps and contribution**	**Outlook on future research**
Chapter 2: *Life after Default: Dealer Intermediation and Recovery in Defaulted Corporate Bonds*		
Bond default events initiate the need for an ownership change from buy-and-hold investors to specialized vulture investors, whereas dealers within the opaque OTC bond market serve as intermediaries. However, the role of dealers for bond intermediation under stress was insufficiently understood.	This chapter exposes an endogenous adjustment of the microstructure of the OTC bond market in response to bond default events, and identifies the causal role of primary dealers and associated pricing effects in defaulted bond intermediation through the OTC dealer network. As such, it contributes to the literature on the role of dealers in OTC markets, dealers' endogenous trading skills and expertise, OTC search and bargaining frictions, and OTC capital commitment and liquidity provision, and on explaining recovery rates.	For developing targeted regulations, and in order to preserve market stability, yet a better understanding of the microstructure of OTC markets is needed. E.g., how do primary dealers obtain their expertise as a competitive advantage?
Chapter 3: *Inter-Industry Network and Corporate Bond Recovery Rates*		
Prior to writing this study, inter-industry contagion effects and dependencies were insufficiently considered within the recovery rate literature. Although recovery rates differ by industry, this heterogeneity of recovery rates across industries had received limited attention from researchers.	This chapter highlights the importance of inter-industry trade relationships in explaining corporate bond recovery rates, and shows that bonds of firms in better connected industries recover more, given broader access to potential asset disposal channels across industry borders. Moreover, it demonstrates that industry-wide distress propagates through industry borders and henceforth affects recoveries in adjacent industries. Recoveries of firms in central industries are more closely connected to macroeconomic conditions. The study contributes to the literature of economic linkages in asset pricing and financial connectedness, as well as recovery rates.	The COVID-19 pandemic has demonstrated the vulnerability of the globalized economy due to interconnectedness, however, different value chains were affected differently by regional lockdowns. Future research may employ a comparable network-based approach on a more recent dataset, focusing the response of recovery rates to impaired value chains due to the COVID-19 pandemic.

Table 6.1 *(Continued)*

Background	**Research steps and contribution**	**Outlook on future research**
Chapter 4: *Corporate Bond Recovery Rate and Financial Markets*		
Prior to conducting this study, research on recovery rates neglected the hybrid nature of defaulted bonds, that entail both equity-type and bond-type characteristics. Thus, conditions in equity and bond markets were insufficiently considered for recovery rate estimation.	This study explains corporate bond recovery rates with the prevailing conditions in financial markets. It shows that traditional asset pricing factors from the stock and bond literature are significantly related to recovery outcomes. Moreover, bond market liquidity, bond market distress and conditions indices, and prevailing equity market valuation levels affect recovery rates. By exposing these relationships as important determinants of recovery rates, the study contributes to the market-based recover rate literature, equity and bond pricing literature, and to the research on the integration of stock and bond markets.	Kelly, Palhares, and Pruitt (2023) show that stock and bond markets are more closely integrated than prior research suggested, however, more research is needed to understand the integration of stock and bond markets. For example, common drivers of stock and bond returns may be investigated more closely. The hybrid nature of defaulted bonds exposed in this chapter may therefore serve as a basis.
Chapter 5: *Intertemporal Defaulted Bond Recoveries Prediction via Machine Learning*		
The majority of research focusing on predicting recovery rates neglects the time-varying nature of recovery rates. While out-of-sample predictions are often performed, it is impossible to employ these in real-world applications, as they rely on information not available at the time of default.	This chapter employs out-of-time recovery rate prediction and evaluates prediction performance of a wide range of prediction techniques. The study shows that machine learning models outperform traditional recovery rate estimation models not only out-of-sample, but also out-of-time. Moreover, the newly applied sparse power expectation propagation approach performs best among the evaluated techniques. The study further employs text-based news measures for recovery rate prediction to derive implications on recovery rates. It contributes to the recovery rate literature and offers more accurate recovery models that can be used in real-world applications.	While machine learning models elevate prediction accuracy, they are often *black boxes* that cannot be interpreted. As regulatory bodies demand transparency on how banks determine credit risk parameters, new studies that investigate approaches to interpreting machine learning models for recovery rates are desired.

Bibliography

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). "The network origins of aggregate fluctuations". *Econometrica* 80(5), pp. 1977–2016.

Acharya, V. V., S. T. Bharath, and A. Srinivasan (2007). "Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries". *Journal of Financial Economics* 85(3), pp. 787–821.

Ahern, K. R. (2013). "Network centrality and the cross section of stock returns". Working paper, University of Southern California.

Ahern, K. R. and J. Harford (2014). "The importance of industry links in merger waves". *Journal of Finance* 69(2), pp. 527–576.

Aizerman, A., E. M. Braverman, and L. Rozoner (1964). "Theoretical foundations of the potential function method in pattern recognition learning". *Automation and Remote Control* 25, pp. 821–837.

Altman, E. I. (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy". *Journal of Finance* 23(4), pp. 589–609.

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

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

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

Altmann, A., L. Toloşi, O.Sander, and T. Lengauer (2010). "Permutation importance: a corrected feature importance measure." *Bioinformatics* 26(10), pp. 1340–1347.

Amihud, Y. (2002). “Illiquidity and stock returns: cross-section and time-series effects”. *Journal of Financial Markets* 5(1), pp. 31–56.

Aobdia, D., J. Caskey, and N. B. Ozel (2014). “Inter-industry network structure and the cross-predictability of earnings and stock returns”. *Review of Accounting Studies* 19(3), pp. 1191–1224.

Bai, J., T. G. Bali, and Q. Wen (2019). “RETRACTED: Common risk factors in the cross-section of corporate bond returns”. *Journal of Financial Economics* 131(3), pp. 619–642.

Bajo, E., T. J. Chemmanur, K. Simonyan, and H. Tehranian (2016). “Underwriter networks, investor attention, and initial public offerings”. *Journal of Financial Economics* 122(2), pp. 376–408.

Baker, S. R., N. Bloom, and S. J. Davis (2016). “Measuring economic policy uncertainty”. *Quarterly Journal of Economics* 131(4), pp. 1593–1636.

Bali, T. G., A. Goyal, D. Huang, F. Jiang, and Q. Wen (2022). “Predicting Corporate Bond Returns: Merton Meets Machine Learning”. *Swiss Finance Institute, Research Paper Series 20–110*.

Bao, J., M. O’Hara, and X. Zhou (2018). “The Volcker Rule and corporate bond market making in times of stress”. *Journal of Financial Economics* 130(1), pp. 95–113.

Bartik, T. J. (1991). “Who benefits from state and local economic development policies?” *W.E. Upjohn Institute*.

Bastos, J. A. (2014). “Ensemble predictions of recovery rates”. *Journal of Financial Services Research* 46(2), pp. 177–193.

Baumann, F., A. Kakhbod, D. Livdan, A. Nazemi, and N. Schürhoff (2023). “Life after default: Dealer intermediation and recovery in defaulted corporate bonds”. *Swiss Finance Institute* Working paper.

Baumann, F. and A. Nazemi (2023). “Corporate bond recovery rate and financial markets”. Working paper.

Bavelas, A. (1950). “Communication patterns in task-oriented groups”. *Journal of the Acoustical Society of America* 22(6), pp. 725–730.

Berger, P. G., E. Ofek, and I. Swary (1996). “Investor valuation of the abandonment option”. *Journal of Financial Economics* 42(2), pp. 259–287.

Bernstein, S., E. Colonnelli, and B. Iverson (2018). "Asset allocation in bankruptcy". *Journal of Finance* 74(1), pp. 5–53.

Bessembinder, H., S. Jacobsen, W. Maxwell, and K. Venkataraman (2018). "Capital commitment and illiquidity in corporate bonds". *Journal of Finance* 73(4), pp. 1615–1661.

Bessembinder, H., W. Maxwell, and K. Venkataraman (2006). "Market transparency, liquidity externalities, and institutional trading costs in corporate bonds". *Journal of Financial Economics* 82(2), pp. 251–288.

Bessembinder, H., C. Spatt, and K. Venkataraman (2020). "A survey of the microstructure of fixed-income markets". *Journal of Financial and Quantitative Analysis* 55(1), pp. 1–45.

Betz, J., R. Kellner, and D. Rösch (2021). "Time matters: How default resolution times impact final loss rates". *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 70(3), pp. 619–644.

Bianchi, D., M. Büchner, and A. Tamoni (2020). "Bond risk premiums with machine learning". *Review of Financial Studies* 34(2), pp. 1046–1089.

Bonacich, P. (1972). "Factoring and weighting approaches to status scores and clique identification". *Journal of Mathematical Sociology* 2(1), pp. 113–120.

Boyarchenko, N., R. K. Crump, A. Kovner, and O. Shachar (2022). "Measuring corporate bond market dislocations". *Federal Reserve Bank of New York Staff Report* (957, Rev. December 2022).

Brave, S. and R. Butters (2011). "Monitoring financial stability: A financial conditions index approach". *Federal Reserve Bank of Chicago Economic Perspectives* 35(1), pp. 22–43.

Brave, S. and D. Kelley (2017). "Introducing the Chicago Fed's new adjusted National Financial Conditions Index". *Federal Reserve Bank of Chicago Letter* (386).

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

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

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). "Beyond LATE with a discrete instrument". *Journal of Political Economy* 125(4), pp. 985–1039.

Bris, A., I. Welch, and N. Zhu (2006). "The costs of bankruptcy: Chapter 7 liquidation versus Chapter 11 reorganization". *Journal of Finance* 61(3), pp. 1253–1303.

Bruche, M. and C. Gonzalez-Aguado (2010). "Recovery rates, default probabilities, and the credit cycle". *Journal of Banking & Finance* 34(4), pp. 754–764.

Bui, T. D., J. Yan, and R. E. Turner (2017). "A unifying framework for Gaussian process pseudopoint approximations using power expectation propagation". *Journal of Machine Learning Research* 18(1), pp. 3649–3720.

Burg, G. J. Van Den, P. J. Groenen, and A. Alfons (2017). "Sparsestep: Approximating the counting norm for sparse regularization". *arXiv*.

Calabrese, R. and M. Zenga (2010). "Bank loan recovery rates: Measuring and nonparametric density estimation". *Journal of Banking & Finance* 34(5), pp. 903–911.

Chaderina, M. and V. Glode (2023). "Trading with expert dealers". Working paper.

Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth (2000). "CRISP-DM 1.0: Step-by-step data mining guide".

Chava, S., C. Stefanescu, and S. Turnbull (2011). "Modeling the loss distribution". *Management Science* 57(7), pp. 1267–1287.

Chen, H. (2010). "Macroeconomic conditions and the puzzles of credit spreads and capital structure". *Journal of Finance* 65(6), pp. 2171–2212.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). "Sparse signals in the cross-section of returns". *Journal of Finance* 74(1), pp. 449–492.

Choi, J., S. Hoseinzade, S. S. Shin, and H. Tehranian (2020). "Corporate bond mutual funds and asset fire sales". *Journal of Financial Economics* 138(2), pp. 432–457.

Choi, J. and Y. Kim (2018). "Anomalies and market (dis)integration". *Journal of Monetary Economics* 100, pp. 16–34.

Colliard, J., T. Foucault, and P. Hoffmann (2021). "Inventory management, dealers' connections, and prices in over-the-counter markets". *Journal of Finance* 76(5), pp. 2199–2247.

Cortes, G. S. and M. D. Weidenmier (2019). "Stock volatility and the great depression". *Review of Financial Studies* 32(9), pp. 3544–3570.

Corwin, S. A. and P. Schultz (2012). "A simple way to estimate bid/ask spreads from daily high and low prices". *Journal of Finance* 67(2), pp. 719–760.

Demiroglu, C., J. R. Franks, and R. Lewis (2022). "Do market prices improve the accuracy of court valuations in Chapter 11?" *Journal of Finance* 77(2), pp. 1179–1218.

Di Maggio, M., A. Kermani, and Z. Song (2017). "The value of trading relations in turbulent times". *Journal of Financial Economics* 124(2), pp. 266–284.

Dick-Nielsen, J. and T. K. Poulsen (2019). "How to clean Academic TRACE data". *SSRN Electronic Journal*.

Dick-Nielsen, J. and M. Rossi (2019). "The cost of immediacy for corporate bonds". *Review of Financial Studies* 32(1), pp. 1–41.

Dickerson, A., C. Julliard, and P. Mueller (2023). "The corporate bond factor zoo". *Working paper*.

Dickerson, A., P. Mueller, and C. Robotti (2023). "Priced risk in corporate bonds". *Journal of Financial Economics* 150(2), p. 103707.

Donovan, J., R. M. Frankel, and X. Martin (2015). "Accounting conservatism and creditor recovery rate". *Accounting Review* 90(6), pp. 2267–2303.

Doshi, H., R. Elkamhi, and C. Ornthanalai (2018). "The term structure of expected recovery rates". *Journal of Financial and Quantitative Analysis* 53(6), pp. 2619–2661.

Duffie, D., N. Garleanu, and L. H. Pedersen (2005). "Over-the-counter markets". *Econometrica* 73(6), pp. 1815–1847.

Easley, D. and J. Kleinberg (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.

Edwards, A. K., L. E. Harris, and M. S. Piwowar (2007). "Bond market transaction costs and transparency". *Journal of Finance* 62(3), pp. 1421–1451.

Ellul, A., C. Jotikasthira, and C. T. Lundblad (2011). "Regulatory pressure and fire sales in the corporate bond market". *Journal of Financial Economics* 101(3), pp. 596–620.

Evgeniou, T., J. Peress, T. Vermaelen, and L. Yue (2021). "Network centrality and managerial market-timing ability". *Journal of Financial and Quantitative Analysis* 57(2), pp. 704–760.

Fama, E. F. and K. R. French (1993). "Common risk factors in the returns on stocks and bonds". *Journal of Financial Economics* 33(1), pp. 3–56.

— (2015). "A five-factor asset pricing model". *Journal of Financial Economics* 116(1), pp. 1–22.

Feldhütter, P. (2012). "The same bond at different prices: Identifying search frictions and selling pressures". *Review of Financial Studies* 25(4), pp. 1155–1206.

Feldhütter, P., E. Hotchkiss, and O. Karakaş (2016). "The value of creditor control in corporate bonds". *Journal of Financial Economics* 121(1), pp. 1–27.

Feng, G., S. Giglio, and D. Xiu (2020). "Taming the factor zoo: A test of new factors". *Journal of Finance* 75(3), pp. 1327–1370.

Franks, J. R. and W. N. Torous (1994). "A comparison of financial recontracting in distressed exchanges and chapter 11 reorganizations". *Journal of Financial Economics* 35(3), pp. 349–370.

Freeman, L. C. (1977). "A set of measures of centrality based on betweenness". *Sociometry* 40(1), pp. 35–41.

Friewald, N. and F. Nagler (2019). "Over-the-counter market frictions and yield spread changes". *Journal of Finance* 74(6), pp. 3217–3257.

Frye, J. (2000). "Depressing recoveries". *Risk*, pp. 106–111.

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

Gavazza, A. (2011). "The role of trading frictions in real asset markets". *American Economic Review* 101(4), pp. 1106–1143.

Glode, V. and C. Opp (2016). "Asymmetric information and intermediation chains". *American Economic Review* 106(9), pp. 2699–2721.

— (2019). "Over-the-counter versus limit-order markets: The role of traders' expertise". *Review of Financial Studies* 33(2), pp. 866–915.

Gofman, M., G. Segal, and Y. Wu (2020). "Production networks and stock returns: the role of vertical creative destruction". *Review of Financial Studies* 33(12), pp. 5856–5905.

Goldberg, J. and Y. Nozawa (2021). "Liquidity supply in the corporate bond market". *Journal of Finance* 76(2), pp. 755–796.

Goldstein, I., H. Jiang, and D. T. Ng (2017). "Investor flows and fragility in corporate bond funds". *Journal of Financial Economics* 126(3), pp. 592–613.

Goldstein, M. A. and E. S. Hotchkiss (2020). "Providing liquidity in an illiquid market: Dealer behavior in US corporate bonds". *Journal of Financial Economics* 135(1), pp. 16–40.

Goldstein, M. A., E. S. Hotchkiss, and E. R. Sirri (2007). "Transparency and liquidity: A controlled experiment on corporate bonds". *Review of Financial Studies* 20(2), pp. 235–273.

Gregorutti, B., B. Michel, and P. Saint-Pierre (2015). "Grouped variable importance with random forests and application to multiple functional data analysis". *Computational Statistics & Data Analysis* 90, pp. 15–35.

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

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

Heckman, J. J. (1979). "Sample selection bias as a specification error". *Econometrica* 47(1), pp. 153–161.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). "Understanding instrumental variables in models with essential heterogeneity". *Review of Economics and Statistics* 88(3), pp. 389–432.

Heidenreich, Konstantin (2019). "Machine learning in credit risk management: An empirical analysis for recovery rates". *Karlsruher Institut für Technologie (KIT).*

Hendershott, T., D. Li, D. Livdan, and N. Schürhoff (2020). "Relationship trading in over-the-counter markets". *Journal of Finance* 75(2), pp. 683–734.

— (2022). "True cost of immediacy". *Swiss Finance Institute* Working paper.

Herskovic, B. (2018). "Networks in production: Asset pricing implications". *Journal of Finance* 73(4), pp. 1785–1818.

Hollifield, B., A. Neklyudov, and C. Spatt (2017). "Bid-ask spreads, trading networks, and the pricing of securitizations". *Review of Financial Studies* 30(9), pp. 3048–3085.

Hong, G. and A. Warga (2000). "An empirical study of bond market transactions". *Financial Analysts Journal* 56(2), pp. 32–46.

Hotchkiss, E. S. and R. M. Mooradian (1998). "Acquisitions as a means of restructuring firms in Chapter 11". *Journal of Financial Intermediation* 7(3), pp. 240–262.

Hugonnier, J., B. Lester, and P.-O. Weill (2019). "Frictional intermediation in over-the-counter markets". *Review of Economic Studies* 87(3), pp. 1432–1469.

Hurlin, C., J. Leymarie, and A. Patin (2018). "Loss functions for loss given default model comparison". *European Journal of Operational Research* 268(1), pp. 348–360.

Ivashina, V., B. Iverson, and D. C. Smith (2016). “The ownership and trading of debt claims in Chapter 11 restructurings”. *Journal of Financial Economics* 119(2), pp. 316–335.

Jacobs, M. and A. K. Karagozoglu (2011). “Modeling ultimate loss given default on corporate debt”. *Journal of Fixed Income* 21(1), pp. 6–20.

Jankowitsch, R., F. Nagler, and M. G. Subrahmanyam (2014). “The determinants of recovery rates in the US corporate bond market”. *Journal of Financial Economics* 114(1), pp. 155–177.

Jansen, J., S. R. Das, and F. J. Fabozzi (2018). “Local volatility and the recovery rate of credit default swaps”. *Journal of Economic Dynamics and Control* 92, pp. 1–29.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). “Is there a replication crisis in finance?” *Journal of Finance* 78(5), pp. 2465–2518.

Kalotay, E. A. and E. I. Altman (2017). “Intertemporal forecasts of defaulted bond recoveries and portfolio losses”. *Review of Finance* 21(1), pp. 433–463.

Kellner, R., M. Nagl, and D. Rösch (2022). “Opening the black box – Quantile neural networks for loss given default prediction”. *Journal of Banking & Finance* 134, p. 106334.

Kelly, B., D. Palhares, and S. Pruitt (2023). “Modeling corporate bond returns”. *Journal of Finance* 78(4), pp. 1967–2008.

Kelly, B., S. Pruitt, and Y. Su (2021). “Instrumented principal components analysis”. *Technical report, Arizona State University, Johns Hopkins University, and Yale University*.

Kermani, A. and Y. Ma (2023). “Asset specificity of nonfinancial firms”. *Quarterly Journal of Economics* 138(1), pp. 205–264.

Khieu, H. D., D. J. Mullineaux, and H. C. Yi (2012). “The determinants of bank loan recovery rates”. *Journal of Banking & Finance* 36, pp. 923–933.

Kim, H. and H. Kung (2016). “The asset redeployability channel: How uncertainty affects corporate investment”. *Review of Financial Studies* 30(1), pp. 245–280.

Koijen, R. S. J. and M. Yogo (2023). “Understanding the ownership structure of corporate bonds”. *American Economic Review: Insights* 5(1), pp. 73–92.

Krüger, S. and D. Rösch (2017). “Downturn LGD modeling using quantile regression”. *Journal of Banking & Finance* 79, pp. 42–56.

Kutzias, D., C. Dukino, F. Kötter, and H. Kett (2023). "Comparative analysis of process models for data science projects". *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023)* 3, pp. 1052–1062.

Li, D. and N. Schürhoff (2019). "Dealer networks". *Journal of Finance* 74(1), pp. 91–144.

Manela, A. and A. Moreira (2017). "News implied volatility and disaster concerns". *Journal of Financial Economics* 123(1), pp. 137–162.

May, A. D. (2010). "The impact of bond rating changes on corporate bond prices: New evidence from the over-the-counter market". *Journal of Banking & Finance* 34(11), pp. 2822–2836.

Meinshausen, N. and P. Bühlmann (2010). "Stability selection". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(4), pp. 417–473.

Metz, A. and S. Sorensen (2012). "Trading prices as predictors of ultimate corporate recovery rates". *Moody's Investor Service, Special Comment.*

Mora, N. (2015). "Creditor recovery: The macroeconomic dependence of industry equilibrium". *Journal of Financial Stability* 18, pp. 172–186.

Müller, M., M. Reichenbacher, P. Schuster, and M. Uhrig-Homburg (2023). "Expected bond liquidity". Working paper.

Nazemi, A., F. Baumann, and F. J. Fabozzi (2022). "Intertemporal defaulted bond recoveries prediction via machine learning". *European Journal of Operational Research* 297(3), pp. 1162–1177.

— (2023). "Inter-industry network and corporate bond recovery rates". Working paper.

Nazemi, A. and F. J. Fabozzi (2018). "Macroeconomic variable selection for creditor recovery rates". *Journal of Banking & Finance* 89, pp. 14–25.

Nazemi, A., F. Fatemi Pour, K. Heidenreich, and F. J. Fabozzi (2017). "Fuzzy decision fusion approach for loss-given-default modeling". *European Journal of Operational Research* 262(2), pp. 780–791.

Nazemi, A., K. Heidenreich, and F. J. Fabozzi (2018). "Improving corporate bond recovery rate prediction using multi-factor support vector regressions". *European Journal of Operational Research* 271(2), pp. 664–675.

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

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). "International stock return predictability: What is the role of the United States?" *Journal of Finance* 68(4), pp. 1633–1662.

Reichenbacher, M. and P. Schuster (2022). "Size-adapted bond liquidity measures and their asset pricing implications". *Journal of Financial Economics* 146(2), pp. 425–443.

Roll, R. (1984). "A simple implicit measure of the effective bid-ask spread in an efficient market". *Journal of Finance* 39(4), pp. 1127–1139.

Rossi, A. G., D. Blake, A. Timmermann, I. Tonks, and R. Wermers (2018). "Network centrality and delegated investment performance". *Journal of Financial Economics* 128(1), pp. 183–206.

Sambalaibat, B. (2022). "Endogenous specialization and dealer networks". Working paper.

Schestag, R., P. Schuster, and M. Uhrig-Homburg (2016). "Measuring liquidity in bond markets". *Review of Financial Studies* 29(5), pp. 1170–1219.

Schläfer, T. and M. Uhrig-Homburg (2014). "Is recovery risk priced?" *Journal of Banking & Finance* 40, pp. 257–270.

Schultz, P. (2001). "Corporate bond trading costs: A peek behind the curtain". *Journal of Finance* 56(2), pp. 677–698.

Shleifer, A. and R. W. Vishny (1992). "Liquidation values and debt capacity: A market equilibrium approach". *Journal of Finance* 47(4), pp. 1343–1366.

Sopitpongstorn, N., P. Silvapulle, J. Gao, and J.-P. Fenech (2021). "Local logit regression for loan recovery rate". *Journal of Banking & Finance* 126, p. 106093.

Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis (2008). "Conditional variable importance for random forests". *BMC Bioinformatics* 9(1), p. 307.

Strömberg, P. (2000). "Conflicts of interest and market illiquidity in bankruptcy auctions: Theory and tests". *Journal of Finance* 55(6), pp. 2641–2692.

Suykens, J. A. K. and J. Vandewalle (1999). "Least squares support vector machine classifiers". *Neural Processing Letters* 9(3), pp. 293–300.

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

Varma, P. and R. Cantor (2005). “Determinants of recovery rates on defaulted bonds and loans for North American corporate issuers: 1983-2003”. *Journal of Fixed Income* 14(4), pp. 29–44.

White, H. (1980). “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity”. *Econometrica* 48(4), pp. 817–838.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. The MIT Press, Cambridge, Massachusetts.

Yao, X., J. Crook, and G. Andreeva (2015). “Support vector regression for loss given default modelling”. *European Journal of Operational Research* 240(2), pp. 528–538.

— (2017). “Enhancing two-stage modelling methodology for loss given default with support vector machines.” *European Journal of Operational Research* 263(2), pp. 679–689.

Zhang, C. H. (2010). “Nearly unbiased variable selection under minimax concave penalty”. *Annals of Statistics* 38(2), pp. 894–942.

Appendix A

Appendix to Chapter 2

A.1 Sample Construction and Methodology

Explanatory variables. The empirical studies on dealer intermediation in defaulted bonds and recovery rates incorporate various explanatory variables. We add information from FISD that is directly associated with the bond issue, such as offering amount, days to maturity at default, coupon rate, covenant information, and bond ratings one year prior to default. We encode ratings as integers, starting with AAA=1, AA+ =2, and so forth. The availability of CDS contracts is retrieved from S&P Capital IQ. From S&P Capital IQ, we also collect point-in-time firm information that represents issuers' characteristics and financials most recently available prior to default. This includes equity value, the number of employees, and both short and long-term debt in order to replicate the default barrier as employed by Jankowitsch, Nagler, and Subrahmanyam (2014) as a proxy for structural credit risk.[1]

We furthermore collect information on pre-default bond ownership from eMaxx data. We retrieve GDP and the slope of the interest yield curve from the Federal Reserve Economic Database of the Federal Reserve Bank of St. Louis (FRED), and we construct the 90-day corporate bond default rate derived from our defaulted bond data and the Transaction Reporting and Compliance Engine (TRACE). We collect industry-specific data about stock indices growth and industry-wide sales growth from S&P Capital IQ for creating industry distress measures similar to those employed by Acharya, Bharath, and Srinivasan (2007). Post-default bond

[1] We consider a firm's market value of equity when available, and book value of equity reported in the most recent company filings prior to default in cases where the market value of equity is not available.

liquidity, similar to that used by Jankowitsch, Nagler, and Subrahmanyam (2014), is calculated with bond transaction data from TRACE.

Dealer network and transaction data. The academic version of TRACE data that we utilize covers actual bond transactions that were executed during the years 2004 through 2016.[2] This data includes comprehensive transaction information, including time stamps, the transactions' par amounts, executed prices, unique CUSIP identifiers for each bond, and, most notably, unique masked identifiers for all dealers involved in the transactions. The dealers' clients are uniformly labeled 'C' without further information about their (masked) identities. The aforementioned characteristics of the data allow us to precisely trace individual bonds as they circulate from clients to dealers, between dealers, and from dealers to clients. Before incorporating data from TRACE into the sample construction, we preprocess the data in order to eliminate known flaws in the data by implementing the standard data cleaning methodologies described by Dick-Nielsen and Poulsen (2019). We first apply a basic transaction filter which removes transactions from TRACE data where a trading sequence of multiple transactions with identical execution prices was reported and which represents an introducing dealer interacting with the executing dealer as an agent on behalf of a client. Note that we also apply a filter to remove erroneous transaction reports from TRACE as suggested by Dick-Nielsen and Poulsen (2019), and further follow Jankowitsch, Nagler, and Subrahmanyam (2014) in applying a price filter to remove potentially falsely reported prices for recovery rate calculation. The cleaned data set contains 114,584,837 reported transactions, involving 107,088 distinct instruments in transactions between 2004 and 2016. We match the cleaned TRACE data to FISD based on the instruments' unique CUSIP identifiers and drop all transactions which involve instruments that are not covered by FISD.[3]

After this step, our data includes 108,895,440 reported transactions of 88,156 distinct debt instruments. Given the masked dealer identifiers linked to the transactions recorded in TRACE, we are able to identify unique dealers and track inter-dealer trade relationships within the data. In total, 3,407 unique dealers intermediate bonds during the period 2004 to 2016. 40.5% of the

[2] The TRACE data contain a few thousand observations of transactions executed prior to 2004, which may be the result of lagged transaction reports to TRACE. We drop all of these transactions which were not executed within the 2004-2016 time period.

[3] Cross-checking with S&P Capital IQ reveals that the majority of the dropped data refers to instruments issued by foreign entities.

Table A.1: Summary statistics of dealer centralities in the corporate bond dealer network 2004–2016. The network is constructed with transaction information reported to the Transaction Reporting and Compliance Engine (TRACE) and represents 44,065,910 inter-dealer transactions that occurred between 2004 and 2016. In the network, 3,383 bond dealers that have transaction relationships with other dealers are represented as nodes. Panel A shows summary statistics of dealer centrality measures derived from an equal-weighted variant of the network in which links have binary weights that indicate if two dealers traded with each other. Panel B shows summary statistics of centrality measures derived from a trades-weighted variant of the network in which links are weighted by the number of transactions between two dealers. Both Panel A and Panel B formally describe a core-periphery network structure in which few dealers are located centrally in the network, while most dealers are located in the network's periphery.

	SD	***q25***	***q50***	***Mean***	***q75***	***q95***	***Max***
Panel A: Equal-weighted network							
Degree	0.04	0.00	0.00	0.02	0.01	0.09	0.40
In-degree	0.04	0.00	0.00	0.01	0.01	0.08	0.36
Out-degree	0.04	0.00	0.00	0.01	0.01	0.08	0.37
Eigenvector	0.17	0.01	0.02	0.10	0.10	0.50	1.00
Betweenness	0.00	0.00	0.00	0.00	0.00	0.00	0.09
Closeness	0.05	0.37	0.40	0.41	0.45	0.51	0.62
In-closeness	0.13	0.35	0.38	0.36	0.43	0.49	0.60
Out-closeness	0.07	0.34	0.36	0.36	0.40	0.46	0.57
Panel B: Trades-weighted network							
Degree	55.11	0.01	0.05	7.70	0.50	14.09	1,446.46
In-degree	29.88	0.00	0.02	3.85	0.25	7.01	1,008.83
Out-degree	26.59	0.00	0.03	3.85	0.23	7.50	544.56
Eigenvector	0.02	0.00	0.00	0.00	0.00	0.00	1.00
Betweenness	0.00	0.00	0.00	0.00	0.00	0.00	0.05
Closeness	0.00	0.02	0.03	0.02	0.03	0.03	0.03
In-closeness	0.01	0.03	0.03	0.02	0.03	0.03	0.03
Out-closeness	0.02	0.07	0.07	0.07	0.08	0.08	0.08

transactions represent inter-dealer transactions whereas 25.3% of the transactions are client-to-dealer transactions and 34.2% are dealer-to-client transactions. Of the 3,407 dealers, 80% are directly interacting with clients, and 20% are solely intermediating bonds between other dealers. 3,383 of the dealers interact with other dealers, whereas 24 dealers only interact with clients but not with other dealers. We remove these dealers for creating the dealer network as they are not connected to it and they represent only a negligibly small number of transactions. We then follow the methodology outlined by Li and Schürhoff (2019) in creating two alternative dealer network representations. The equal-weighted dealer network solely indicates the existence of a transaction relationship between two dealers and the trades-weighted variant weighs links by the number of transactions executed between dealers.

From the dealer network representations, we compute dealer centrality measures. The descriptive statistics for degree, in-degree, out-degree, eigenvector (Bonacich (1972)), betweenness (Freeman (1977)), closeness, as well as in-closeness and out-closeness (Bavelas (1950)) central-

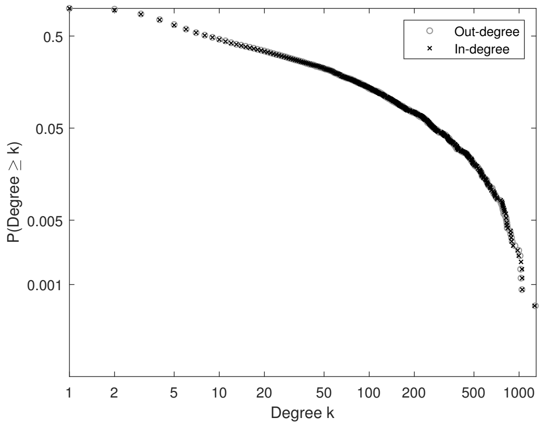


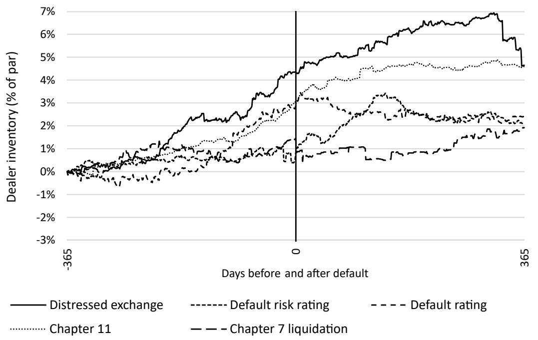
Figure A.1: The figure illustrates the inverse in- and out-degree centrality distribution of the 3,383 bond dealers that form the dealer network 2004–2016. Out-degree centrality is represented by circles and in-degree centrality by cross-marks. The figure is log-scaled and shows that the centrality distribution is right-skewed, with a large number of dealers that maintain only few trade relationships, and a small number of dealers that maintain many trade relationships.

ity are shown in Table A.1. Di Maggio, Kermani, and Song (2017) report a core-periphery structure of the corporate bond dealer network, which we confirm and refer to Figure A.1 for an illustration of the non-randomness in dealer connectedness based on in- and out-degree centrality.

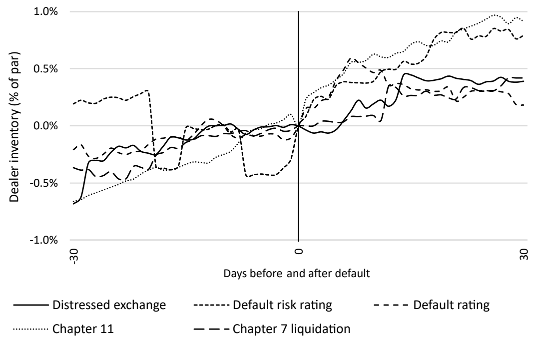
For the purpose of our empirical analysis, we use 1-year monthly trailing dealer networks to determine dealers' centralities. As the data covers the period 2004–2016, we drop all 124 bonds that defaulted before 2005 for which we don't have a complete year of trading data to create the dealer network prior to default. Furthermore, for comparing bond-level characteristics between pre- and post-default trading periods, we only consider bonds for which we have trading data in both periods. We also only consider those transactions in which dealers act as buyers, yielding a data set that comprises 2,446 bonds for comparing dealer centralities. During the year prior to default, each of these bonds is bought by an average (median) of 53 (39) unique dealers, and during the 30-day period after default by an average (median) of only about 20 (11) dealers, indicating a concentration of trading activity in recently defaulted bonds on fewer dealers. This observation suggests that recently defaulted bonds are intermediated by a smaller group of dealers than before the default event, likely because investors switch to more expert dealers, such as primary dealers, after default.

Determining agency trades. For each bond transaction reported in TRACE data, an indicator denotes agency trades in which dealers prearrange the trades in a broker role without taking inventory risk. These agency trades represent about 8% of all transactions reported to TRACE over the years 2004–2016. However, as a standard convention in the literature, principal trades that are offset within one minute after the dealer purchased the bond are also considered prearranged riskless trades, as it is likely that these client-to-dealer trades are only executed after a dealer searched and found a trade counterparty to immediately offset the position. Hence, we follow this convention and denote all client-to-dealer trades that are offset by consecutive dealer-to-dealer or dealer-to-client trades of the same par amount within one minute as prearranged agency trades, in line with Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), Bao, O'Hara, and Zhou (2018), and Li and Schürhoff (2019). It may occur that a dealer splits the trade after the purchase, selling the bonds to several buyers. We account for splits of up to three separate offsetting trades. Using this broader definition to identify dealers' riskless trades, 36% of the trades in 2004–2016 TRACE data are trades in which dealers act as brokers, and 64% of the trades are principal trades. This differs from Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), who report about 90% of trades in TRACE between 2006 and 2016 are principal trades. However, they only consider the top 10–12 dealers that correspond to about 70% of total trading volume in TRACE, whereas we include all dealers in our analysis. When we only consider trades performed by the top 12 dealers, we find that these account for about 67% of the total USD trading volume, for which only 11% of the client-to-dealer trades are agency trades, comparable to the characteristics of the transaction data sample employed by Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018). In our empirical analysis, we consider this definition of agency trades in order to distinguish the role of dealers in intermediating defaulted bonds.

Determining dealer inventory. We create inventory measures that reflect dealers' collective inventory additions and subtractions in defaulted bonds from a normalized reference point. We offset all client-to-dealer trades with all dealer-to-client trades on each day for a given bond. Alternatively, we consider the date when the bond's outstanding amount in FISD is set to zero as the date that the bond ceases to exist. This happens in only a few cases shortly after the default event, and we offset the whole inventory for a given bond to zero in these cases. We



(A) Collective dealer inventory, indexed at zero one year before the default event.



(B) Collective dealer inventory, indexed at zero on the day of the default event.

Figure A.2: Collective average dealer inventory in defaulted firms' bonds, distinguished by default event type. The dealer inventory is calculated as the average par value of all bonds of a firm that dealers hold on their balance sheet and is fixed at 0 one year before default in Panel A and after default including the default day itself in Panel B. After removing default events of the years 2004 and 2016, and 7 outlier firms, 629 firm-default observations involving 2,338 bonds that defaulted between January 2005 and December 2015 are considered.

define the residual as the dealers' collective net inventory change in a given bond. In order to not distort variations in inventory due to price fluctuations, we consider trade volume in par amount for accumulating and offsetting positions. In dealer-to-dealer trades, the inventory of the buyer-dealer will increase by the same amount that the inventory of the seller-dealer decreases, hence, a net effect of zero on the dealers' collective inventory will be recorded in dealer-to-dealer trades. We compute dealer inventory on a daily basis over the pre- and post-default periods for each bond, that is the year before default until 30 days after default. As no starting inventory is known, we may index the collective dealers' inventory for each bond at 0 on a reference date. The daily inventory measure thus reflects deviations from this starting inventory. Figure A.2 illustrates the collective dealer inventory in defaulted bonds by default event type with reference dates one year prior to default (Panel A) and on the default day (Panel B), respectively.

A.2 Empirical Model for Bond Recovery

Let RR_0 (RR_1) be the recovery rate at a non-primary (primary) dealer, and let $PrimaryDealer$ indicate the investor's choice of dealer with $PrimaryDealer = 1$ for a primary dealer and $PrimaryDealer = 0$ for a non-primary dealer. The difference in potential recovery rates RR_j between primary and non-primary dealers equals

$$RR_1 = RR_0 + \delta \, PrimaryDealer. \tag{A.1}$$

An important empirical question is whether the causal effect of the primary dealer $\delta \equiv RR_1 - RR_0$ is positive or negative. We specify the potential recoveries as $RR_j = \mu_j(x) + \epsilon_j$, $j = 0, 1$, and, for simplicity, we assume linearity:

$$RR_j = X\beta + \delta \, PrimaryDealer + \epsilon_j, j = 0, 1. \tag{A.2}$$

To account for the endogenous investor-dealer matching during the bond default, we make the following assumption: The investor's net benefit I of trading with a primary dealer depends on observed determinants $Z = (X, W)$ including the dealer's experience and expertise, connections, expected trade delays, and other intermediation services provided by the dealer and an unobserved component U: $I = \mu(Z) - U$, where $\mu(\cdot)$ is an unspecified function and U is a continuous random variable with a strictly increasing distribution function. W is an instrument affecting the investors' choice satisfying the standard exclusion restriction.

The investor's dealer choice can be expressed as $PrimaryDealer = 1 \Leftrightarrow \mu(Z) > V \Leftrightarrow F_V(\mu(Z)) > F_V(V) \Leftrightarrow P(Z) > U$. Written in this way, $P(Z)$ denotes the propensity score that captures the selection probability of a primary dealer while u is a uniformly distributed random variable between 0 and 1 representing the propensity to trade with a non-primary dealer. Assuming linearity, the propensity to trade with a primary dealer can be estimated using:

$$\Pr(PrimaryDealer = 1|Z) = P(X\gamma + W\theta). \tag{A.3}$$

Table 2.1 reports the results of different specifications for (A.3). Let $K_j(p)$ be the selection corrections for the expected recovery surprises for primary and non-primary dealers, respectively,

$K_1(p) = \mathbb{E}[\epsilon_1 | U \leq p]$, $K_0(p) = -\frac{p}{1-p}\mathbb{E}[\epsilon_0 | U \leq p]$, and define

$$K(p) = pK_1(p) + (1-p)K_0(p). \tag{A.4}$$

We can estimate the determinants of recovery rates and δ in the sample of trades with primary and non-primary dealers by jointly estimating the regressions

$$RR_j = X\beta + \delta\, PrimaryDealer + K_j(p) + \varepsilon_j, \tag{A.5}$$

with mean-zero errors ε_j, $j = 0, 1$, and $K_j(p)$ is an unspecified function of the propensity score with $\lim_{p \to 0} K_1(p) = 0$, $\lim_{p \to 1} K_0(p) = 0$ (Brinch, Mogstad, and Wiswall, 2017).[4] Observed recoveries in the pooled sample can be decomposed as

$$RR = X\beta + \delta\, PrimaryDealer \times p + K(p) + \varepsilon, \tag{A.6}$$

with mean-zero error ε and $K(p)$ is the Mills term defined in (A.4) and/or treated as an unspecified function of the propensity of trading with a primary dealer during a bond default. $K(p) \leq 0$ satisfies $\lim_{p \to 0} K(p) = \lim_{p \to 1} K(p) = 0$ and can be semi-parametrically estimated by polynomials.

[4] Under normality assumptions, $K_j(p)$ is proportional to the respective inverse Mills ratio.

Appendix B

Appendix to Chapter 3

B.1 Network Representation of the U.S. Economy

Since 1947 the U.S. BEA has provided detailed tables, depicting input-output (IO) accounts, which track dollar payments between all producers and purchasers of industrial goods and services within the U.S. economy. In these accounts, producers summarize companies within industrial and service sectors, including government enterprises. Producers may also purchase goods or services from adjacent industries as input factors. Hence, producers also appear as purchasers in the accounts. However, not all purchaser industries are also producers themselves, as they include households, the government, and the foreign sector as further economic agents.

The BEA provides IO tables for different levels of industry detail. Detailed IO tables are available only for every five years since 1947 and are published with a five-year delay. Summary IO tables provide yearly data. With regard to our dataset of defaulted bonds covering the period 2001-2016, we consider 1997 detailed IO tables, and annual summary IO tables for the years 2000–2015, so that we only consider the network structure prior to default.

For grouping firms into producer and purchaser industries, the BEA utilizes adjusted NAICS industry definitions. In 1997, the BEA switched from using the 1937 Standard Industrial Classification (SIC) codes to using the North American Industry Classification System (NAICS) codes. While SIC codes were originally designed to classify mainly manufacturing industries, NAICS codes further include a variety of defined service industries, and therefore represent today's economy better than the SIC codes. For the IO tables, the BEA adjusts the NAICS codes and provides concordance tables for translating NAICS codes into adjusted NAICS codes.

We use summary IO tables in our main analysis to maintain consistency with the level of industry detail of the available control variables used for interaction analysis, and detailed IO tables for robustness checks. The summary IO tables are used in order to create yearly network representations of the U.S. economy which track bilateral transactions between all industrial sectors on an annual basis. In addition to industry sectors, we further include a foreign agent for imports and exports, the government as a redistributive agent and households as purchasers of goods and services and providers of labor. This complete network model of an economy is known as a social accounting matrix (SAM). To create it, we closely follow the methodology of Ahern (2013) and Ahern and Harford (2014).

The IO reports track all inputs and outputs of the industries in the U.S. economy. Commodities serve as inputs and outputs and are traded between industries in exchange for monetary compensation. Values of commodity volumes in these transactions are denoted in producer prices. Hence, it is possible to allocate the volumes of produced and purchased commodities to producers and purchasers. Therefore, the IO reports provide two tables: MAKE tables (number of industries $\times$ number of commodities) which exhibit production outputs, and USE tables (number of commodities $\times$ number of industries) which exhibit purchases of inputs. Both commodities and industries of producers or purchasers are defined by NAICS codes. The IO reports distinguish between two government agents: a redistributive agent and government enterprises. Furthermore, we introduce artificial producer industries in order to account for outputs that are not covered by MAKE tables but need to be included in a complete social accounting matrix (e.g., compensation of employees, taxes, and gross operating surplus).

For each year we create a matrix that records dollar flows between all industries by multiplying a modified version of the MAKE table, representing the market share of each producer per output commodity with the USE table. This yields an IO matrix (number of industries $\times$ number of industries) that shows the dollar flows between all industries. Rows represent the industries' receipts and columns represent purchases. This approach relies on the assumption that purchasers do not choose from which producer they receive their input commodities, but that receipts are equally distributed among all producers of a commodity in relation to their market shares in the production of the according commodity.

Next, we aggregate different industries (e.g., households and labor are combined to a single household sector and imports and exports are combined to a single foreign sector). To fully include the government as a redistributive agent that collects taxes and triggers consumption and investments expenditures, we use data from the BEA's National Income and Product Accounts (NIPA) tables that are also provided on a yearly basis.

We introduce several tax components that are not covered by the IO tables. Assuming that each industry contributes tax in proportion to its total value added, we proportionally assign the total value of tax collections from the NIPA tables to each industry. Further, these adjustments cover tax payments from households and the capital sector. Finally, differences between total amounts of receipts and purchases of each industry are recognized in the capital sector so each industry's inputs and outputs equate. This yields a SAM where all the rows and columns are balanced such that an industry's purchases neither exceed nor fall below receipts. In this network representation, nodes define industries, and links between nodes represent dollar flows of inter-industry trade. A condensed representation of the SAM can be found in Table B.1.

Comparable to Evgeniou, Peress, Vermaelen, and Yue (2021), we remove specific sectors such as households, government, capital, and foreign sectors, and only consider important trade relationships, that is, if the trade volume of a supplier industry with a customer industry represents at least 1% of its total trade volume. An illustration of the inter-industry network after these adjustments can be found for the year 2001 in Figure B.2. Based on this network, we create the centrality measures for our empirical analysis.

B.2 Concepts from Network Theory

We use concepts from social and network theory to evaluate interrelationships within the network representation of the U.S. economy, basing our analysis upon the inter-industry networks that capture interconnections between all private-sector industries within the U.S. economy. The centrality of a node quantifies the importance of a node for the network. Several different centrality measures exist and capture various properties of single nodes and the network structure. This section introduces those measures which we apply in our analysis.

Degree centrality: The most intuitive and simple measure for the centrality of a node is degree centrality. Degree centrality DC_i of node i is calculated by dividing the degree deg_i of a node by the number of all possible links that it could potentially have to other nodes in the network. Hence, it follows

$$DC_i = \frac{deg_i}{n-1} \tag{B.1}$$

where n is the number of nodes which are part of the network. Degree centrality considers a node that has links to all other nodes in the network as the most central. Degree centrality depends on the size of the network (i.e. an increasing network size leads to potentially larger degrees of centrality for the nodes).

Closeness centrality: Degree centrality and variations of it do not include more information on the structure of a network other than the directly connected neighbors of a node and the total number of nodes within the network. It is intuitive that a node which serves as the only connection between two parts of a network is important, independent of its degree centrality. The network might be separated and change its structure substantially, if that node, which is called a bridge, is removed from the network (Easley and Kleinberg (2010)). Closeness centrality takes up this issue by measuring the lengths of the shortest paths from a node to all other nodes in the network and taking its inverse.

$$CC_i = \frac{1}{\sum_j^n d_{ij}} \tag{B.2}$$

is the closeness centrality of node i while d_{ij} is the length of the shortest path from node i to node j and n represents the total number of nodes in the network. If the links are weighted, one has to consider whether the weights represent cost (e.g., distance) or tie strength, which has the opposite character of cost. If the weights represent tie strength, one would have to inverse the weights before calculating shortest paths. Hence,

$$CC_i^w = \frac{1}{\sum_j^n d_{ij}^w} \tag{B.3}$$

while

$$d_{ij}^w = min\left(\frac{1}{g_{ih}} + ... + \frac{1}{g_{hj}}\right) \tag{B.4}$$

with adjacency matrix g takes weights into account. However, both methods factor in either the number of intermediary nodes or the links' weights.

Betweenness centrality: Betweenness centrality, as in Freeman (1977), proposes the idea that bridges in networks are central. To measure the importance of a node in this sense, betweenness centrality of a node in the network considers all the shortest paths between all nodes of the network passing through that node. For node i, $P_i(kj)$ is the number of shortest paths between nodes k and j on which node i lays. With $P(kj)$ being the total number of shortest paths between nodes k and j, $\frac{P_i(kj)}{P(kj)}$ then illustrates the importance of node i for connecting nodes k and j. When this fraction equals 1, i is part of all of the shortest paths between nodes k and j, and as it approaches 0, i becomes less relevant for the connection of nodes k and j. The betweenness centrality of node i is then defined by calculating the previously introduced ratio for all pairs of nodes and then averaging it:

$$BC_i = \sum_{k \neq j : i \notin \{k,j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2} \tag{B.5}$$

For weighted networks, the same equation applies. However, $P_i(kj)$ and $P(kj)$ depend on either d_{ij}, d_{ij}^w or $d_{ij}^{w\alpha}$.

Eigenvector centrality: A node that has a large degree centrality but is mostly connected to other nodes that themselves have small degree centralities is overestimated compared to a node that is connected to other highly connected nodes. Bonacich (1972) introduces eigenvector centrality, which captures this property. Eigenvector centrality is a centrality measure that considers node importance within the economy network by determining the importance of each of its neighbors. A node is considered central if it is connected with other nodes that are central themselves. Thus, eigenvector centrality EC_i of node i is defined as the sum of the centralities of its neighbors:

$$EC_i = \frac{1}{\lambda} \sum_{j}^{n} g_{ij} EC_j \tag{B.6}$$

or

$$\lambda \; EC = EC \times g \tag{B.7}$$

where $\lambda \neq 0$ is a constant, n is the number of all nodes within the economy, and g is the adjacency matrix of the given network. We choose λ as the largest eigenvalue in the absolute value of matrix g, following the Perron-Frobenius theorem, as this guarantees a unique and positive eigenvector solution.

B.3 Stability Selection Technique

Variable selection has become an increasingly important issue in finance as a result of the increasing availability of high-dimensional financial data. The model interpretability, prediction accuracy and computational efficiency are desirable characteristics of financial models which could be achieved if we utilize variable selection techniques in the analysis of high-dimensional data. Rapach, Strauss, and Zhou (2013) state that the least absolute shrinkage and selection operator (LASSO) and adaptive LASSO are more robust than the traditional statistical approach, such as stepwise regression for variable selection. Motivated by these advantages of a model containing selected explanatory variables, we use variable selection technique for selecting variables from an extensive set of macroeconomic variables.

LASSO is the most common method used in selecting variables in the finance literature.

LASSO has a l_1 penalty defined by $\sum_{j=1}^{k}|\beta_j| = \|\beta\|_1$. The LASSO coefficient is estimated by

$$\sum_{i=1}^{N}(y_i - \alpha - \sum_{j=1}^{k}\beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^{k}|\beta_j| \tag{B.8}$$

Some of the estimated LASSO coefficients of variables are exactly equal to zero. This is an advantage provided by LASSO compared to a ridge regression, as the ridge coefficients will never equal zero. Therefore, LASSO can be used for variable selection and shrinkage. Nazemi and Fabozzi (2018) apply LASSO as variable selection techniques to select explanatory variables from many macroeconomic predictors.

The LASSO-selected macroeconomic variables can be highly unstable as a result of a small perturbation to the dataset. Meinshausen and Bühlmann (2010) suggest a stability selection technique for improving the stability of the variable selection process. The stability selection technique for selecting variables involves generating numerous bootstrapped samples of the entire dataset and counting the proportion of times that each variable is selected. Finally, the stability selection selects the set of variables that are selected more often than a particular threshold. The main advantage of this algorithm compared to LASSO or adaptive LASSO are stability of results and less risk of incorrectly selected variables. Meinshausen and Bühlmann (2010) mention that a stability selection algorithm significantly improves variable selection and structure estimation compared to selection models such as LASSO or adaptive LASSO. Moreover, a stability selection model has less dependency on the first regularization value. In recent years, stability selection models have been widely used in bioinformatics where these studies report that this method improves the selection accuracy.

B.4 Support Vector Regression

We use three different support vector regression techniques. Yao, Crook, and Andreeva (2015) find that support vector regression techniques outperform linear regression for recovery rate prediction, considering both in-sample and out-of-sample predictions. We rely on least squares support vector regression (LS–SVR) as proposed by Suykens and Vandewalle (1999) and two improved support vector regressions that are based on LS–SVR.

Aizerman, Braverman, and Rozoner (1964) show that Mercer's theorem allows for computationally efficient calculation of kernelized problems, so that an appropriate kernel must be selected. We use radial basis function (RBF) kernels for all three support vector regression techniques. The RBF kernels are defined by

$$K(x_i, x_j) = exp\left(-\frac{\| x_i - x_j\|^2}{2\sigma^2}\right) \tag{B.9}$$

which only satisfies the requirements that a kernel be positive semi-definite and represent a similarity measure between pairs of input samples.

Least squares support vector regression: We define the LS–SVR with quadratic loss function as:

$$\begin{aligned} min\ J(\Omega; u_i) &= \frac{1}{2}\|\Omega\|^2 + \frac{C}{2}\sum_{i=1}^{N} u_i^2 \\ s.t. \quad y_i &= \Omega^T\phi(X_i) + b + u_i, \quad i = 1, ..., N \end{aligned} \tag{B.10}$$

with independent variables vector w and intercept b for N observations. The error terms u_i^2 are scaled by a regularized parameter C and $\phi(x_i)$ represents the kernel function, which projects the original data to a higher dimensional space.

It is possible to solve the problem by solving its dual form problem. Therefore, we obtain the Lagrangian function:

$$L(\Omega, b, u_i, \alpha_i) = J(\Omega, u_i) - \sum_{i=1}^{N} \alpha_i(\Omega^T\phi(X_i) + b + u_i - y_i), \tag{B.11}$$

with α_i as the Lagrangian multiplier. Based on the Karush-Kuhn-Tucker condition, solving the dual-form problem is equivalent to solving the following linear equations system:

$$\begin{pmatrix} 0 & \mathrm{e}^T \\ \mathrm{e} & \overline{K} \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \tag{B.12}$$

with $\mathrm{e} = (1, ...1)^T$ as an $1 \times N$ unit vector, $y = (y_1, ..., y_N)^T$ as the dependent variables vector, $\alpha = (\alpha_1, ..., \alpha_N)^T$ as the vector of Lagrangian multipliers and $\overline{K} = K + \frac{1}{C}I$, where K is the kernel matrix as defined in Equation B.9 and I is the identity matrix.

The final estimated regression model is represented by

$$g(X) = \sum_i \alpha_i^* \; K(X_i, X) + b^* \tag{B.13}$$

Least squares support vector regression with different intercepts for seniorities: We also utilize this technique in order to account for unobserved homogeneity within seniority classes that can represent different intercepts. The LS–SVR with different intercepts is defined by:

$$\begin{aligned} min \; J(\Omega, b_k; u_{kj}) &= \frac{1}{2}\|\Omega\|^2 + \frac{1}{2}\sum_{k=1}^{M} b_k^2 + \frac{C}{2}\sum_{k=1}^{M}\sum_{j=1}^{p_k} u_{kj}^2 \\ s.t. \quad y_{kj} = \Omega^T \phi(X_{kj}) &+ b_k + u_{kj}, \quad k = 1, ..., M, \quad j = 1, ..., p_k \end{aligned} \tag{B.14}$$

for M as the number of seniorities and p_k as the number of defaulted bonds within each seniority, so that $p_1 + p_2 + ... + p_M = N$ equals the total number of defaulted bonds. The Lagrangian function is then:

$$L(\Omega, b_k, u_{kj}; \alpha_{kj}) = J(\Omega, b_k; u_{kj}) - \sum_{k=1}^{M}\sum_{j=1}^{p_k} \alpha_{kj}(w^T \phi(x_k j) + b_k + u_{kj} - y_{kj}) \tag{B.15}$$

and the dual-form problem is:

$$min \; \frac{1}{2}\alpha^T K \alpha + \frac{1}{2}\alpha^T W \alpha + \frac{1}{2C}\alpha^T \alpha - y^T \alpha, \tag{B.16}$$

with W as a block diagonal matrix.

Semi-parametric least squares support vector regression: Semi-parametric LS-SVR as described by Yao et al. (2015) incorporates dummy variables to indicate the unobservable heterogeneity of bond seniorities. It is assumed that dummy variables representing seniority

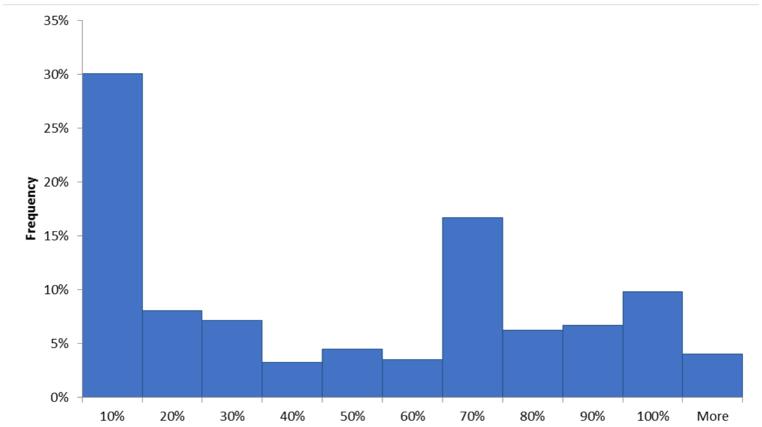


Figure B.1: Histogram of all observed U.S. corporate bond recovery rates from our dataset, covering the years 2001-2016. Recovery rates on the horizontal axis is shown as a percentage of par value.

influence the dependent variable linearly. The model is defined by:

$$\begin{aligned} &min\ J(\Omega, b; u_{kj}) = \frac{1}{2}\|\ \Omega\|^2 + \frac{1}{2}\beta^T\beta + \frac{1}{2}b^2 + \frac{C}{2}\sum_{k=1}^{M}\sum_{j=1}^{p_k} u_{kj}^2 \\ s.t. \quad & y_{kj} = \Omega^T\phi(X_{kj}) + \beta^T z_{kj} b + u_{kj}, \quad k = 1, ..., M, \quad j = 1, ..., p_k \end{aligned} \tag{B.17}$$

with z_{kj}, which contains the dummy variables for seniority, and β as a fixed effects the vector of the corresponding parameters with respect to the seniority-class specific variables. We obtain the Lagrangian function as follows:

$$L(\Omega, b, u_kj; \alpha_{kj}) = J(\Omega, b_k; u_{kj}) - \sum_{k=1}^{M}\sum_{j=1}^{p_k} \alpha_{kj}(w^T\phi(x_kj) + \beta^T z_{kj} + b + u_{kj} - y_{kj}) \tag{B.18}$$

and yield the dual form problem

$$min\ \frac{1}{2}\alpha^T K\alpha + \frac{1}{2}\alpha^T Z\alpha + \frac{1}{2}\alpha^T V\alpha + \frac{1}{2C}\alpha^T\alpha - y^T\alpha \tag{B.19}$$

with $Z_{ij} = z_{ki}^T z_{kj}$ and V as an $N \times N$ matrix with all elements equal to 1.

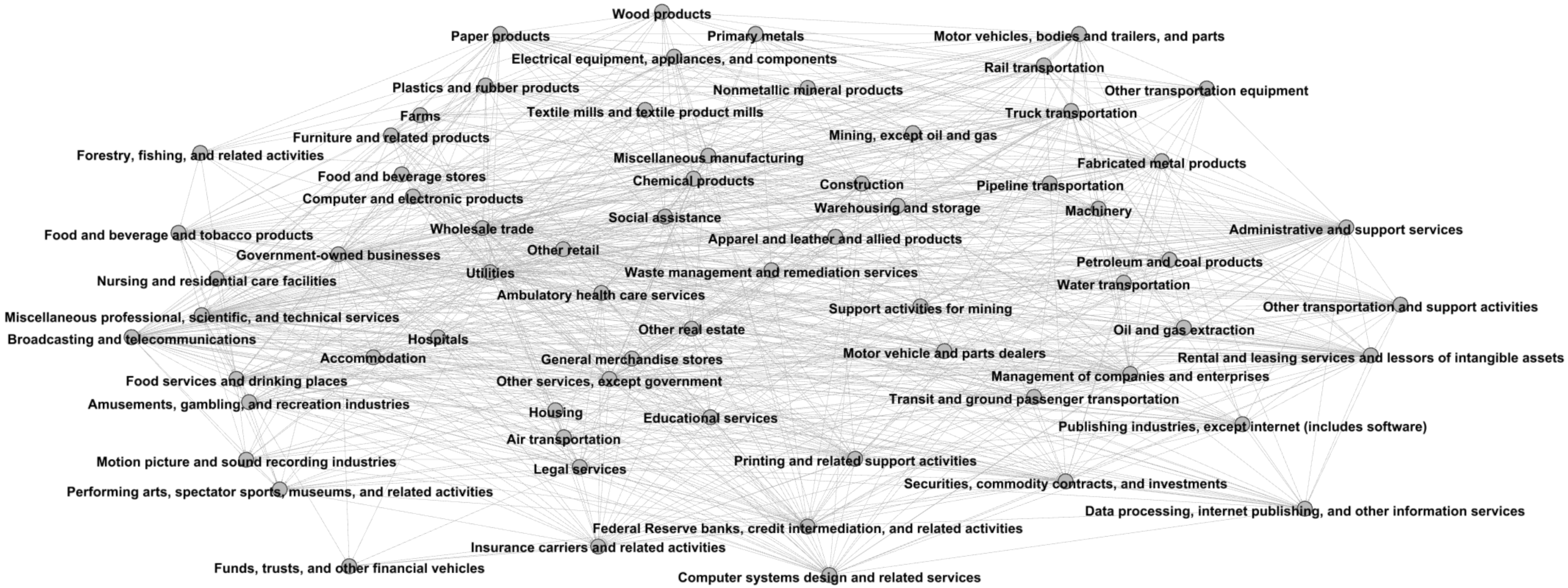


Figure B.2: Illustration of the network representation of the U.S. economy for the year 2001. Nodes represent 67 distinct industries. Links represent the existence of important trade relationships between industries, that is, if a customer industry represents at least 1% of its supplier's total trade volume. The illustration is created by simulating repelling forces between unconnected nodes, and attractive forces between connected nodes.

Table B.1: 2001 Social Accounting Matrix (aggregated to 19 industries for illustrative purpose). Rows represent suppliers, columns represent customers, figures represent payments from customers to suppliers in millions of U.S. dollars. Suppliers and customers include households, the capital sector, the foreign sector, and the government as a redistributive agent. Other codes are according to the definition of the BEA (adjusted NAICS codes): 11: Agriculture, forestry, fishing, and hunting; 21: Mining; 22: Utilities; 23: Construction; 31G: Manufacturing; 42: Wholesale trade; 44RT: Retail trade; 48TW: Transportation and warehousing; 51: Information; FIRE: Finance, insurance, real estate, rental, and leasing, PROF: Professional and business services; 6: Educational services, health care, and social assistance; 7: Arts, entertainment, recreation, accommodation, and food services; 81&F030: Other services, except government & Change in private inventories; Gov. Ent.: Government enterprises; Househ: Households; Capital: Capital sector; Foreign: Foreign sector; Gov.: Government as a redistributive agent.

Customer/ Supplier	**11**	**21**	**22**	**23**	**31G**	**42**	**44RT**	**48TW**	**51**	**FIRE**	**PROF**	**6**	**7**	**81&F030**	**Gov. Ent.**	**Househ.**	**Capital**	**Foreign**	**Gov.**
11	54,055	56	16	1,029	140,247	1,566	1,686	39	33	71	723	317	3,906	1,051	1,109	48,335	267	20,451	14,606
21	1,257	21,077	62,668	9,683	153,100	202	219	2,081	678	2,602	944	699	847	2,670	9,152	6,100	41,562	7,473	1,579
22	8,444	8,175	3,899	5,386	81,049	6,248	16,902	4,023	8,904	56,507	10,742	23,887	18,406	4,602	23,992	148,744	1,312	2,071	1,216
23	1,330	2,789	4,667	583	12,259	726	2,177	2,763	4,321	43,804	3,912	1,855	2,332	2,654	32,760	51	643,690	97	192,417
31G	41,284	21,379	27,209	248,330	1,397,367	37,581	38,642	59,394	68,089	49,969	89,595	119,617	91,527	0	204,793	1,272,947	605,752	533,273	86,008
42	14,132	3,615	9,256	31,949	182,499	22,809	11,243	12,834	13,229	13,779	14,478	21,651	12,718	4,398	26,680	279,754	93,413	84,266	9,939
44RT	511	482	2,405	54,512	17,674	2,129	4,844	4,095	3,203	11,669	7,259	3,893	6,366	5,864	6,382	821,637	36,590	1,852	880
48TW	7,041	4,078	21,502	14,297	84,740	28,633	23,445	55,409	15,432	15,922	29,872	12,001	7,028	4,598	35,667	158,304	19,216	54,211	2,020
51	556	1,858	8,103	12,363	42,969	14,962	15,125	10,106	202,500	53,036	71,912	24,580	10,814	8,398	72,030	301,960	128,976	40,573	15,618
FIRE	16,115	19,892	47,778	29,237	100,836	41,542	74,741	47,897	56,909	571,067	143,869	127,481	48,498	57,861	66,379	1,680,509	87,223	59,414	5,324
PROF	2,486	12,004	42,665	38,670	229,149	62,923	57,581	35,639	103,820	193,180	223,359	86,243	48,614	22,703	127,782	143,022	309,517	41,803	92,090
6	166	149	969	939	2,844	1,177	5,114	623	1,639	2,989	3,775	18,517	1,188	1,095	13,107	1,152,892	4,810	1,866	1,304
7	179	181	7,914	2,886	10,308	3,409	3,841	9,483	16,944	27,045	39,904	10,531	19,235	2,981	12,036	506,377	3,770	2,621	552
81&F030	647	292	2,232	11,233	18,146	5,625	5,707	2,343	9,312	19,234	18,288	10,193	5,316	13,688	22,521	296,479	1,043	199	274
Gov. Ent.	3,003	3,054	6,740	6,073	49,930	15,187	14,623	28,015	21,702	48,805	29,105	17,934	13,843	3,980	34,374	285,994	0	98,657	1,562,038
Househ.	31,373	38,611	53,904	325,403	881,901	330,201	416,679	207,964	247,207	526,842	851,402	604,494	241,079	177,285	1,119,403	0	1,752,500	0	1,145,800
Capital	69,003	59,132	83,632	130,472	455,852	125,925	134,082	72,458	177,903	1,211,814	232,385	91,354	85,022	102,006	251,989	188,293	0	388,683	0
Foreign	23,801	100,355	2,512	10	991,155	0	2,841	10,910	8,513	16,207	24,052	2,135	308	1,487	133,224	0	0	0	27,700
Gov.	14,178	27,419	46,438	32,133	140,730	161,797	162,752	27,343	76,102	418,032	77,673	37,781	63,150	25,448	49,675	1,660,650	130,366	7,700	0
Total	289,563	324,596	434,507	955,187	4,992,755	862,641	992,245	593,418	1,036,440	3,282,574	1,873,247	1,215,163	680,196	442,771	2,243,053	8,952,048	3,860,005	1,345,211	3,159,366

Table B.2: List of the most and least central industries. Most central and least central industries of the 2001 inter-industry network. Centrality is defined as log(eigenvector centrality) for the 67 nodes in the network of inter-industry trade.

Most central industries	Least central industries
Government-owned businesses (most central)	Food and beverage stores
Wholesale trade	Farms
Miscellaneous professional, scientific, and technical services	Performing arts, spectator sports, museums, and related activities
Other services, except government	Support activities for mining
Utilities	General merchandise stores
Administrative and support services	Warehousing and storage
Chemical products	Nursing and residential care facilities
Management of companies and enterprises	Motion picture and sound recording industries
Other real estate	Apparel and leather and allied products
Credit intermediation, and related activities	Oil and gas extraction
Broadcasting and telecommunications	Water transportation
Construction	Pipeline transportation
Rental and leasing services and lessors of intangible assets	Forestry, fishing, and related activities
Fabricated metal products	Housing
Plastics and rubber products	Funds, trusts, and other financial vehicles (least central)

Table B.3: Descriptive statistics of centrality and other network-derived variables. Centrality and network-derived variables are based on involving the trade network of 67 unique industries in columns (1)–(3), the detailed network of trade involving 471 unique industries in column (4), and networks based on inter-industry non-bankruptcy mergers and acquisitions (M&A) (column (5)) and bankruptcy mergers M&A (column (6)).

	Trade network			**Trade network - Detail**	**M&A network**	**Bankruptcy M&A network**
	Log(eigenvector centrality)	**Log(neighbor industry distress)**	**Labor's fraction of inputs (in %)**	**Log(eigenvector centrality)**	**Log(eigenvector centrality)**	**Log(eigenvector centrality)**
Mean	-2.2680	-1.2768	0.2696	-3.7168	-7.0883	-7.1559
Std. Dev.	0.5886	0.6327	0.1459	0.5681	2.1025	2.5850
Skewness	0.5189	0.1059	0.1830	1.4926	0.0954	-1.6203
Kurtosis	0.8685	-1.1709	-0.9142	3.7492	2.0812	7.0026
Min	-3.7006	-2.3513	0.0042	-4.8944	-13.8521	-19.4138
5th Percentile	-3.2241	-2.1489	0.0463	-4.4526	-10.6998	-11.6315
25th	-2.7005	-1.9019	0.1595	-4.0472	-8.4822	-8.0427
Median	-2.2987	-1.3192	0.2629	-3.7958	-7.0883	-6.8106
75th	-1.8908	-0.7501	0.3846	-3.5363	-5.8621	-5.8785
95th Percentile	-1.3536	-0.2281	0.5185	-2.6093	-3.6934	-4.6348
Maximum	-0.2055	-0.0195	0.5801	-0.8993	-0.0025	-0.0022

Table B.4: List of 179 macroeconomic variables from which the bold variables were selected for recovery rate modeling by applying a stability selection technique.

Series ID	Short name	Description
1	Loans & leases	Loans and Leases in Bank Credit, All Commercial Banks
2	RE loans	Real Estate Loans, All Commercial Banks
3	Federal debt	Federal Debt: Total Public Debt
4	Cons credit	Total Consumer Credit Owned and Securitized, Outstanding
5	Reserves	Excess Reserves of Depository Institutions
6	C&I loans	Commercial and Industrial Loans, All Commercial Banks
7	Borrowings	Total Borrowings of Depository Institutions from the Federal Reserve
8	Bank credit	Bank Credit of All Commercial Banks
9	Hh debt servs	Household Debt Service Payments as a Percent of Disposable Personal Income
10	Hh fin obl	Household Financial Obligations as a percent of Disposable Personal Income
11	Loans & leases 2	Loans and Leases in Bank Credit, All Commercial Banks
12	Nonp. Loans	Nonperforming Loans (past due 90+ days) to Total Loans
13	Nonp. loans: 100 - 300m	Nonperforming Loans (past due 90+ days, for banks with USD 100M-300M average asset size)
14	Loan losses	Net Loan Losses to Average Total Loans for all U.S. Banks
15	Loan charge-offs	Total Net Loan Charge-offs to Total Loans for Banks
16	Equity return	Return on Average Equity for all U.S. Banks
17	Loan loss reserve	Loan Loss Reserve to Total Loans for all U.S. Banks
18	Nonp com loans	Nonperforming Commercial Loans (past due 90+ days plus nonaccrual) to Commercial Loans
19	GDP	Real Gross Domestic Product
20	ISM PMI	ISM Manufacturing: PMI Composite Index
21	IP: total	Industrial Production Index
22	Cons sent	University of Michigan: Consumer Sentiment
23	Fixed inv	Private Nonresidential Fixed Investment
24	Disp income	Real Disposable Personal Income
25	NI	National income
26	PI	Personal Income
27	Output: mfg	Manufacturing Sector: Real Output
28	Consumption	Real Personal Consumption Expenditures
29	IP: mfg (NAICS)	Industrial Production: Manufacturing (NAICS)
30	PCE dble	Personal Consumption Expenditures: Durable Goods
31	GCE	Government Consumption Expenditures & Gross Investment
32	Gross inv	Gross Private Domestic Investment
33	U: all	Civilian Unemployment Rate
34	Cont claims	Continued Claims (Insured Unemployment)
35	Avg hrs: mfg	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing
36	Emp	Civilian Employment
37	Emp rate	Civilian Employment-Population Ratio
38	U 15+ wks rate	Persons unemployed 15 weeks or longer, as a percent of the civilian labor force
39	Orders: dble gds	Manufacturers' New Orders: Durable Goods
40	Final sales	Real Final Sales of Domestic Product
41	Orders: cap gds	Manufacturers' New Orders: Nondefense Capital Goods Excluding Aircraft
42	Invent/sales	Total Business Inventories to Sales Ratio
43	Cap util: mfg	Capacity Utilization: Manufacturing
44	**Invent change**	**Change in Private Inventories**
45	Cap util: total	Capacity Utilization: Total Industry
46	Inventories	Total Business Inventories
47	Vehicle sales	Light Weight Vehicle Sales: Autos & Light Trucks
48	Starts	Housing Starts: Total: New Privately Owned Housing Units Started

Table B.4 (Continued)

Series ID	Short name	Description
49	Starts_adj	Housing Starts: Total: New Privately Owned Housing Units Started (Seasonally adjusted)
50	Houses sold	New One Family Houses Sold: United States
51	BP: total	New Private Housing Units Authorized by Building Permits
52	Sales / dom purch	Final Sales to Domestic Purchasers
53	M2	M2 Money Stock
54	CPI-U: ex food	Consumer Price Index for All Urban Consumers: All Items Less Food
55	**Inflation expect**	**University of Michigan Inflation Expectation**
56	CPI-U: energy	Consumer Price Index for All Urban Consumers: Energy
57	Saving	Personal Saving
58	**Saving rate**	**Personal Saving Rate**
59	Gross saving	Gross Saving
60	GDP defl	Gross Domestic Product: Implicit Price Deflator
61	CP	Corporate Profits After Tax (without IVA and CCAdj)
62	CP: adj	Corporate Profits After Tax with Inventory Valuation and Capital Consumption Adjustments
63	CP: adj div	Corporate Profits after tax with IVA and CCAdj: Net Dividends
64	CNCF	Corporate Net Cash Flow with IVA
65	TWI: US broad	Real Trade Weighted U.S. Dollar Index: Broad
66	TWI: US major	Trade Weighted U.S. Dollar Index: Major Currencies
67	Acc balance	Balance on Current Accounts for the United States
68	Exports	Real Exports of Goods & Services
69	Merch trade balance	Merchandise Trade as percentage of GDP
70	Imports	Real imports of goods and services
71	Labor cost: mfg	Manufacturing Sector: Unit Labor Cost
72	Labor cost: bus	Nonfarm Business Sector: Unit Labor Cost
73	Compensation: wages	Compensation of employees: Wages and salaries
74	Compensation: mfg nondble	Compensation of employees: Domestic private industries: Manufacturing: Nondurable goods: Food and beverage and tobacco products
75	ECI: mgmt	Full time employment: Wage and salary workers, 16 years and over
76	Compensation: mfg dble	Manufacturing Durable Goods Sector: Compensation
77	ECI: benefits	Employment Cost Index: Benefits: Private Industry Workers
78	ECI: total comp	Employment Cost Index: Total compensation for All Civilian workers in all industries and occupations
79	ECI: wages&salaries	Employment Cost Index: Wages & Salaries: Private Industry Workers
80	1 mo CP	1-Month AA Nonfinancial Commercial Paper Rate
81	10 yr T-bond	10-Year Treasury Constant Maturity Rate
82	3 mo CP	3-Month AA Nonfinancial Commercial Paper Rate
83	Term Structure	TermStructure
84	Fed Funds	Effective Federal Funds Rate
85	Baa -10 yr T-bond spread	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity
86	Aaa yield	Moody's Seasoned Aaa Corporate Bond Yield
87	Mortg 30 yr	30-Year Fixed Rate Mortgage Average
88	Baa yield	Moody's Seasoned Baa Corporate Bond Yield
89	Loan rate	Bank Prime Loan Rate
90	PPI: all	Producer Price Index for All Commodities
91	PPI: industrial	Producer Price Index by Commodity Industrial Commodities
92	PPI: int energy	Producer Price Index by Commodity Intermediate Energy Goods
93	Crude oil price: WTI	West Texas Intermediate Price per Barrel
94	PPI: cons gds	Producer Price Index by Commodity for Finished Consumer Goods
95	PPI: int matls	Producer Price Index by Commodity Intermediate Materials: Supplies & Components
96	S&P 500	S&P 500 Index
97	**S&P 500 vol**	**S&P 500 Volatility**

Table B.4 (Continued)

Series ID	Short name	Description
98	DowJones vol	CBOE DJIA Volatility Index
99	Nasdaq	NASDAQ 100 Index
100	Nasdaq vol	NASDAQ 100 Volatility
101	Russel2000	Russell 2000 Price Index
102	Russel2000 vol	Russell 2000 Volalitlity 1m
103	Wilshire	Wilshire US Small-Cap Price Index
104	Wilshire vol	Wilshire Small Cap Vol
105	Orders: cons gds	Value of Manufacturers' New Orders for Consumer Goods Industries
106	Unf orders: dble	Value of Manufacturers' Unfilled Orders for Durable Goods Industries
107	Overtime: mfg	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing
108	MB	Monetary Base
109	AHE: goods	Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing, Dollars per Hour
110	AHE: const	Average Hourly Earnings of Production and Nonsupervisory Employees: Construction, Dollars per Hour
111	AHE: mfg	Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing, Dollars per Hour
112	M&T Sales	Real Manufacturing and Trade Industries SalesAdjusted
113	CPI-U: apparel	Consumer Price Index for All Urban Consumers: Apparel
114	CPI-U: all	Consumer Price Index for All Urban Consumers: All Items
115	CPI-U: medical	Consumer Price Index for All Urban Consumers: Medical Care
116	CPI-U: transp	Consumer Price Index for All Urban Consumers: Transportation
117	CPI-U: ex shelter	Consumer Price Index for All Urban Consumers: All items less shelter
118	CPI-U: ex med	Consumer Price Index for All Urban Consumers: All items less medical care
119	CPI-U: dbles	Consumer Price Index for All Urban Consumers: Durables
120	CPI-U: services	Consumer Price Index for All Urban Consumers: Services
121	CPI-U: comm.	Consumer Price Index for All Urban Consumers: Commodities
122	PCE defl: dlbes	Personal consumption expenditures: Durable goods (implicit price deflator)
123	Ex rate: Canada	Canada / U.S. Foreign Exchange Rate, Canadian Dollars to One U.S. Dollar
124	Ex rate: Japan	Japan / U.S. Foreign Exchange Rate, Japanese Yen to One U.S. Dollar
125	Ex rate: Switz	Switzerland / U.S. Foreign Exchange Rate, Swiss Francs to One U.S. Dollar
126	**Ex rate: UK**	**U.S. / U.K. Foreign Exchange Rate, U.S. Dollars to One British Pound**
127	1 yr T-bond	1-Year Treasury Constant Maturity Rate
128	5 yr T-bond	5-Year Treasury Constant Maturity Rate
129	PCE defl: nondble	Personal consumption expenditures: Nondurable goods (implicit price deflator)
130	PCE defl	Personal consumption expenditures (implicit price deflator)
131	6 mo T-bill	6-Month Treasury Bill: Secondary Market Rate
132	Starts: MW	Housing Starts in Midwest Census Region
133	**Starts: NE**	**Housing Starts in Northeast Census Region**
134	Starts: South	Housing Starts in South Census Region
135	Starts: West	Housing Starts in West Census Region
136	UI claims	Initial Unemployment Claims
137	M&T invent	Real Manufacturing and Trade Inventories
138	IP: buseqpt	Industrial Production: Business Equipment
139	IP: cons gds	Industrial Production: Consumer Goods
140	IP: cons dble	Industrial Production: Durable Consumer Goods
141	IP: dble matls	Industrial Production: Durable Materials
142	IP: final prod	Industrial Production: Final Products (Market Group)
143	IP: fuels	Industrial Production: Fuels
144	IP: mfg	Industrial Production: Manufacturing (SIC)
145	IP: matls	Industrial Production: Materials
146	IP: cons nondble	Industrial Production: Nondurable Consumer Goods
147	IP: nondble matls	Industrial Production: Nondurable Materials
148	M1	M1 Money Stock

Table B.4 (Continued)

Series ID	Short name	Description
149	M3	M3 for the United States
150	M&T invent/sales	Manufacturers: Inventories to Sales Ratio
151	Emp: const	Nonfarm Private Construction Payroll Employment
152	Emp: FIRE	Nonfarm Private Financial Activities Payroll Employment
153	Emp: gds prod	Nonfarm Private Goods - Producing Payroll Employment
154	Emp: mfg	Nonfarm Private Manufacturing Payroll Employment
155	Emp: services	Nonfarm Private Service - Providing Payroll Employment
156	Emp: total	Total Nonfarm Private Payroll Employment
157	Emp: TTU	Nonfarm Private Trade, Transportation, and Utilities Payroll Employment
158	BP: MW	New Private Housing Units Authorized by Building Permits in the Midwest Census Region
159	BP: NE	New Private Housing Units Authorized by Building Permits in the Northeast Census Region
160	BP: South	New Private Housing Units Authorized by Building Permits in the South Census Region
161	BP: West	New Private Housing Units Authorized by Building Permits in the West Census Region
162	Ex broad: US	Real Broad Effective Exchange Rate for United States
163	3 mo T-bill	3-Month Treasury Bill: Secondary Market Rate
164	U 5-14 wks	Number of Civilians Unemployed for 5 to 14 Weeks
165	U 15+ wks	Number of Civilians Unemployed for 15 Weeks and Over
166	U 15-26 wks	Number of Civilians Unemployed for 15 to 26 Weeks
167	U 27+ wks	Number of Civilians Unemployed for 27 Weeks and Over
168	**U <5 wks**	**Number of Civilians Unemployed for Less Than 5 Weeks**
169	U: mean duration	Average (Mean) Duration of Unemployment
170	PPI: fin gds	Producer Price Index by Commodity for Final Demand: Finished Goods
171	**Aaa-FF spread**	**Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate**
172	Baa-FF spread	Moody's Seasoned Baa Corporate Bond Minus Federal Funds Rate
173	3 mo CP-FF spread	3-Month Commercial Paper Minus Federal Funds Rate
174	Consumer opinion	Consumer Opinion Surveys: Confidence Indicators: Composite Indicators: OECD Indicator for the United States
175	HPI	All-Transactions House Price Index for the United States
176	Aaa-Baa spread	Moody's Seasoned Aaa Bbb Spread
177	High yield market	Size of High Yield Market in U.S. Dollars
178	**High yield DR**	**High Yield Default Rate, Trailing 12-month**
179	Industry Default Rate	Bond defaults within the industry (in percent)

Table B.5: Descriptive statistics of 179 macroeconomic variables from which the bold variables were selected for recovery rate modeling by applying a stability selection technique.

Series ID	Short name	Maximum	Minimum	Average	Median
1	Loans & leases (USD bn)	9,146.1	3,860.0	6,308.8	6,754.4
2	RE loans (USD bn)	4,104.8	1,638.6	3,148.6	3,527.4
3	Federal debt (USD mn)	19,573,445.0	5,726,815.0	11,727,877.1	10,699,805.0
4	Cons credit (USD bn)	3,754.4	1,729.9	2,602.8	2,572.1
5	Reserves (USD mn)	2,699,968.0	888.0	883,723.5	642,072.0
6	C&I loans (USD bn)	2,101.7	863.0	1,317.7	1,244.5
7	Borrowings (USD bn)	437.5	0.0	14.2	0.1
8	Bank credit (USD bn)	12,464.7	5,048.4	8,512.9	9,013.4
9	Hh debt servs (%)	13.2	9.9	11.6	12.2
10	Hh fin obl (%)	18.1	15.0	16.6	16.9
11	Loans & leases 2 (USD bn)	9,122.6	3,869.5	6,296.1	6,752.3
12	Nonp. Loans (%)	5.6	0.7	2.3	1.6
13	Nonp. loans: 100 - 300m (%)	3.3	0.7	1.6	1.2
14	Loan losses (%)	3.1	0.4	1.0	0.8
15	Loan charge-offs (%)	3.1	0.4	1.0	0.8
16	Equity return (%)	15.5	-1.0	10.1	9.4
17	Loan loss reserve (%)	3.7	1.2	1.9	1.8
18	Nonp com loans (%)	3.6	0.5	1.5	1.2
19	GDP (USD bn, chained 2009)	16,778.1	12,643.3	14,689.9	14,745.9
20	ISM PMI (Index)	61.4	33.1	52.0	52.4
21	IP: total (Index 2007=100)	106.6	87.1	98.7	99.7
22	Cons sent (Index Q1 1966=100)	107.6	55.3	82.4	84.5
23	Fixed inv (USD bn)	2,336.2	1,348.9	1,779.8	1,776.3
24	Disp income (USD bn, chained 2009)	12,663.5	9,054.5	10,893.4	10,916.0
25	NI (USD bn)	15,739.6	9,184.6	12,308.6	12,321.4
26	PI (USD bn)	16,043.4	8,923.2	12,250.4	12,221.4
27	Output: mfg (Index 2009=100)	129.0	97.6	116.0	118.1
28	Consumption (USD bn, chained 2009)	11,698.0	8,314.4	9,946.3	10,001.3
29	IP: mfg (NAICS) (NAICS)	108.2	86.6	98.7	100.0
30	PCE dble (USD bn)	1,420.2	919.5	1,140.4	1,139.8
31	GCE (USD bn)	3,274.6	1,911.9	2,800.1	3,049.7
32	Gross inv (USD bn)	3,115.7	1,786.4	2,422.7	2,469.5
33	U: all (%)	10.0	4.2	6.4	5.7
34	Cont claims (Number)	6,635,000	2,010,000	3,213,932	2,990,000
35	Avg hrs: mfg (Hours)	42.2	39.3	41.1	41.1
36	Emp (Thousands of persons)	152,048.0	135,701.0	142,436.8	142,206.0
37	Emp rate (%)	64.4	58.2	60.8	61.0
38	U 15+ wks rate (%)	6.3	1.0	2.9	2.2
39	Orders: dble gds (USD mn)	290,709.0	143,769.0	203,276.1	210,229.0
40	Final sales (USD bn, chained 2009)	16,741.1	12,681.9	14,655.4	14,703.3
41	Orders: cap gds (USD mn)	70,343.0	46,355.0	60,043.2	61,262.0
42	Invent/sales (%)	1.5	1.2	1.3	1.3
43	Cap util: mfg (%)	79.2	64.0	74.6	74.9
44	**Invent change (ΔUSD bn)**	**148.6**	**-205.9**	**-40.0**	**-11.6**
45	Cap util: total (%)	81.0	66.7	76.8	77.0
46	Inventories (USD mn)	1,836,476.0	1,106,874.0	1,444,889.0	1,424,619.0
47	Vehicle sales (Units in mn)	21.7	9.0	15.4	16.2
48	Starts (Units in thousands)	197.9	31.9	104.2	95.0
49	Starts_2 (Units in thousands)	2,273.0	478.0	1,249.3	1,138.0
50	Houses sold (Units in thousands)	1,389.0	270.0	691.8	536.0
51	BP: total (Units in thousands)	2,263.0	513.0	1,293.1	1,178.0
52	Sales / dom purch (USD bn)	19,205.5	10,930.4	15,049.3	15,217.5
53	M2 (USD bn)	13,187.7	4,938.4	8,397.6	8,206.7
54	CPI-U: ex food (Index 1982-84=100)	241.4	176.6	211.2	214.3

Table B.5 (Continued)

Series ID	Short name	Maximum	Minimum	Average	Median
55	**Inflation expect (%)**	**5.2**	**0.4**	**3.0**	**3.0**
56	CPI-U: energy (Index 1982-84=100)	271.1	113.5	195.7	201.0
57	Saving (USD bn)	1,425.7	182.3	548.3	577.2
58	**Saving rate (%)**	**11.0**	**1.9**	**5.1**	**5.4**
59	Gross saving (%)	3,539.9	1,941.7	2,540.8	2,353.8
60	GDP defl (Index 2009=100)	111.6	83.1	98.2	99.8
61	CP (USD bn)	1,801.8	458.9	1,261.1	1,350.2
62	CP: adj (USD bn)	1,741.4	519.1	1,196.0	1,166.7
63	CP: adj div (USD bn)	1,062.0	367.1	704.6	718.3
64	CNCF (USD bn)	2,209.6	1,022.5	1,717.5	1,607.8
65	TWI: US broad (Index March 1973=100)	112.8	80.3	94.0	94.3
66	TWI: US major (Index March 1973=100)	112.6	68.1	84.1	81.7
67	Acc balance (USD bn)	-320.7	-858.7	-525.2	-474.6
68	Exports (USD bn, chained 2009)	2,150.8	1,118.0	1,663.7	1,682.9
69	Merch trade (% of GDP)	24.2	17.2	21.0	21.2
70	Imports (USD bn, chained 2009)	2,702.6	1,639.8	2,227.3	2,310.7
71	Labor cost: mfg (Index 2009=100)	102.1	86.6	93.5	93.1
72	Labor cost: bus (Index 2009=100)	109.0	91.4	99.3	99.9
73	Compensation: wages (USD bn)	8,188.8	4,928.5	6,360.8	6,385.9
74	Compensation: mfg nondble (USD mn)	103,076.0	69,389.0	84,021.0	84,276.0
75	FTE >16y (Thousands of persons)	106.0	80.0	94.6	95.0
76	Compensation: mfg dble (Index 2009=100)	124.1	97.4	109.9	111.1
77	ECI: benefits (Index Dec. 2005=100)	127.0	78.8	106.4	107.9
78	ECI: total comp (Index Dec. 2005=100)	127.4	84.7	107.5	109.6
79	ECI: wages&salaries (Index Dec. 2005=100)	126.6	87.6	107.7	109.5
80	1 mo CP (%)	6.5	0.0	1.5	0.5
81	10 yr T-bond (%)	5.5	1.4	3.5	3.6
82	3 mo CP (%)	6.2	0.1	1.6	0.7
83	Term Structure (%)	5.9	0.0	1.4	0.4
84	Fed Funds (%)	6.0	0.1	1.5	0.4
85	Baa -10 yr T-bond spread (%)	6.0	1.6	2.7	2.8
86	Aaa yield (%)	7.4	3.2	5.1	5.3
87	Mortg 30 yr (%)	7.2	3.3	5.2	5.3
88	Baa yield (%)	9.5	4.2	6.2	6.2
89	Loan rate (%)	9.5	3.3	4.6	3.6
90	PPI: all (Index 1982=100)	208.3	128.1	173.7	178.1
91	PPI: industrial (Index 1982=100)	209.5	129.1	174.8	180.1
92	PPI: int energy (Index 1982=100)	203.4	82.9	149.8	151.9
93	Crude oil price: WTI (USD per bbl)	145.3	17.5	64.5	62.0
94	PPI: cons gds (Index 1982=100)	216.7	137.6	179.5	181.7
95	PPI: int matls (Index 1982=100)	196.4	134.6	170.5	174.6
96	S&P 500 (Index)	2,271.7	676.5	1,372.9	1,275.5
97	**S&P 500 vol (Index)**	**80.9**	**9.9**	**20.2**	**17.8**
98	DowJones vol (Index)	74.6	9.3	18.8	16.4
99	Nasdaq (Index)	4,965.8	804.6	2,296.2	1,855.4
100	Nasdaq vol (Index)	82.5	11.4	26.4	21.0
101	Russel2000 (Index)	3,449.7	812.9	1,876.8	1,777.1
102	Russel2000 vol (Index)	1,038.8	12.9	315.7	316.6
103	Wilshire (Index)	8,954.0	1,878.9	4,790.2	4,400.5
104	Wilshire vol (Index)	2,488.6	30.6	810.4	789.5
105	Orders: cons gds (USD mn)	214,933.0	120,537.0	170,037.4	169,850.0
106	Unf orders: dble (USD mn)	1,175,861.0	478,363.0	832,177.9	879,577.0
107	Overtime: mfg (Hours)	4.6	2.6	4.1	4.2
108	MB (USD mn)	4,075,039.0	595,873.0	1,868,195.0	1,561,699.0
109	AHE: goods (USD per hour)	22.8	15.5	19.3	19.7

Table B.5 (Continued)

Series ID	Short name	Maximum	Minimum	Average	Median
110	AHE: const (USD per hour)	26.3	17.8	21.9	22.4
111	AHE: mfg (USD per hour)	20.6	14.5	17.8	18.0
112	M&T Sales (USD mn, chained 2009)	1,239,361.0	953,116.0	1,086,640.8	1,094,829.0
113	CPI-U: apparel (Index 1982-84=100)	129.6	117.8	122.8	121.1
114	CPI-U: all (Index 1982-84=100)	242.2	175.6	211.3	214.7
115	CPI-U: medical (Index 1982-84=100)	470.2	267.2	367.6	368.7
116	CPI-U: transp (Index 1982-84=100)	223.5	148.7	186.9	190.2
117	CPI-U: ex shelter (Index 1982-84=100)	227.0	168.8	201.2	203.7
118	CPI-U: ex med (Index 1982-84=100)	231.3	170.6	203.5	206.7
119	CPI-U: dbles (Index 1982-84=100)	125.7	106.3	113.6	112.4
120	CPI-U: services (Index 1982-84=100)	303.3	200.6	252.0	258.0
121	CPI-U: comm. (Index 1982-84=100)	190.5	147.8	170.3	171.6
122	PCE defl: dlbes (Index 2009=100)	119.2	88.1	102.1	100.8
123	Ex rate: Canada (CAD/USD)	1.6	0.9	1.2	1.2
124	Ex rate: Japan (JPY/USD)	134.8	75.7	105.8	108.1
125	Ex rate: Switz (CHF/USD)	1.8	0.7	1.1	1.1
126	**Ex rate: UK (USD/GBP)**	**2.0**	**1.2**	**1.6**	**1.6**
127	1 yr T-bond (%)	5.3	0.1	1.6	0.8
128	5 yr T-bond (%)	5.2	0.6	2.7	2.5
129	PCE defl: nondble (Index 2009=100)	113.1	82.5	99.3	101.2
130	PCE defl (Index 2009=100)	111.0	84.4	98.6	99.7
131	6 mo T-bill (%)	5.4	0.0	1.5	0.6
132	Starts: MW (Units in thousands)	446	59	216	173
133	**Starts: NE (Units in thousands)**	**236**	**36**	**126**	**130**
134	Starts: South (Units in thousands)	1,146	230	606	580
135	Starts: West (Units in thousands)	583	79	302	269
136	UI claims (Number)	665,000	239,000	371,416	358,000
137	M&T invent (USD, chained 2009)	1,792,850.0	1,344,697.0	1,519,683.5	1,504,742.0
138	IP: buseqpt (Index 2012=100)	103.8	75.5	90.6	90.6
139	IP: cons gds (Index 2012=100)	114.3	97.2	105.5	104.8
140	IP: cons dble (Index 2012=100)	125.9	79.4	110.4	113.5
141	IP: dble matls (Index 2012=100)	108.2	72.2	93.2	95.8
142	IP: final prod (Index 2012=100)	109.0	90.6	99.9	99.9
143	IP: fuels (Index 2012=100)	108.0	79.0	95.1	97.4
144	IP: mfg (Index 2012=100)	110.0	87.2	99.9	100.2
145	IP: matls (Index 2012=100)	110.4	82.6	95.5	94.8
146	IP: cons nondble (Index 2012=100)	112.3	98.2	104.0	103.4
147	IP: nondble matls (Index 2012=100)	115.7	92.2	103.8	102.0
148	M1 (USD bn)	3,373.4	1,094.2	1,864.7	1,569.3
149	M3 (USD bn)	13,144.8	4,948.8	8,373.2	8,156.8
150	M&T invent/sales (%)	1.5	1.1	1.3	1.3
151	Emp: const (Thousands)	7,726.0	5,427.0	6,558.3	6,701.0
152	Emp: FIRE (Thousands)	8,394.0	7,676.0	8,025.5	8,032.0
153	Emp: gds prod (Thousands)	24,533.0	17,627.0	20,534.1	20,322.0
154	Emp: mfg (Thousands)	17,104.0	11,453.0	13,261.2	12,850.0
155	Emp: services (Thousands)	20,380.0	15,877.0	17,543.4	17,380.0
156	Emp: total (Thousands)	145,170.0	129,733.0	134,933.0	134,053.0
157	Emp: TTU (Thousands)	27,346.0	24,473.0	25,861.3	25,802.0
158	BP: MW (Units in thousands)	402	83	220	181
159	BP: NE (Units in thousands)	306	58	135	132
160	BP: South (Units in thousands)	1,104	257	623	581
161	BP: West (Units in thousands)	619	97	315	280
162	Ex broad: US (Index 2010=100)	129.0	93.0	108.1	108.2
163	3 mo T-bill (%)	5.2	0.0	1.4	0.3
164	U 5-14 wks (Thousands of persons)	4,458	1,764	2,614	2,509

Table B.5 (Continued)

Series ID	**Short name**	**Maximum**	**Minimum**	**Average**	**Median**
165	U 15+ wks (Thousands of persons)	9,130	1,372	4,395	3,344
166	U 15-26 wks (Thousands of persons)	3,488	696	1,533	1,388
167	U 27+ wks (Thousands of persons)	6,800	624	2,862	2,006
168	**U <5 wks (Thousands of persons)**	**3,524**	**2,087**	**2,690**	**2,668**
169	U: mean duration (Weeks)	40.7	12.1	25.1	20.1
170	PPI: fin gds (Index 1982=100)	201.7	137.7	171.8	173.9
171	**Aaa-FF spread (%)**	**6.3**	**0.0**	**4.1**	**4.7**
172	Baa-FF spread (%)	9.3	0.8	4.7	5.0
173	3 mo CP-FF spread (%)	2.9	-0.9	0.1	0.1
174	Consumer opinion (Normal=100)	101.3	96.7	99.5	99.7
175	HPI (Index Q1 1980=100)	382.4	246.3	326.9	328.2
176	Aaa-Baa spread (%)	6.3	-0.2	3.6	3.9
177	High yield market (USD bn)	1,254.6	468.7	833.6	765.8
178	**High yield DR (%)**	**16.4**	**0.5**	**4.9**	**3.1**
179	Industry Default Rate (%)	17.3	0.0	2.6	1.8

Appendix C

Appendix to Chapter 5

C.1 Linear Regression as Benchmark

For comparing the out-of-time and out-of-sample performance of our machine learning techniques with more statistical methods, we include a traditional linear regression model as the benchmark model. Thus, we estimate the following linear regression model for the recovery rate RR_{ijt} of bond i in industry j that defaulted at the time t to serve as benchmark for our machine learning models:

$$\begin{aligned} RR_{ijt} = &\alpha + \beta_c(\text{instrument-specific variables})_i \\ &+\nu(\text{industry distress variables})_{jt} \\ &+\eta(\text{news-based variables})_t \\ &+\zeta(\text{selected macroeconomic variables})_t \\ &+\epsilon_{ijt} \quad \epsilon_{ijt} \sim N(0, \sigma^2) \end{aligned} \tag{C.1}$$

We control for the instrument-specific variables with dummy variables for the industry, seniority, coupon type, and instrument type. The industry distress variables indicate whether the performance of the industry index was worse than -30% and the sales growth was negative in the year preceding the default. The news-based measures can capture uncertainty and disaster risk. The various methodologies benchmarked for selecting the macroeconomic variables are presented in Section C.6.

C.2 Inverse Gaussian Regression

Due to its popularity in recovery rate modeling in studies such as Qi and Zhao (2011) and Kalotay and Altman (2017), we also consider the inverse Gaussian regression. In doing so, the recovery rates are transformed from the interval (0,1) to (-∞, ∞) using the inverse Gaussian cumulative distribution function. These transformed recovery rates are then regressed on the independent variables as described for the case of the ordinary linear regression. Finally, the estimated values are transformed back from (-∞, ∞) to (0,1) using the Gaussian distribution function.

C.3 Regression Tree

One class of machine learning methods that has been found to deliver very good predictive performance as well as an easy-to-understand model is the regression tree. Qi and Zhao (2011), Kalotay and Altman (2017), and Nazemi and Fabozzi (2018) used regression trees successfully for loss-given-default modeling. Two other advantages of the regression tree are that it can be used to model non-linearity and it exhibits a relatively robust behavior against outliers. For these reasons, we apply the classification and regression technique (CART) algorithm as defined by Breiman, Friedman, Stone, and Olshen (1984) for the creation of the regression tree model.

C.4 Random Forest

Breiman (2001) introduces random forest as a model that is more robust and has a better predictive capacity out-of-sample than the regression tree. Random forest is an improvement of bagging which trains a large number of regression trees and then predicts the average of the trees' predictions. Better performance and reduced variance of the predictions are the advantages of bagging compared with regression trees. In a random forest, a random subset of explanatory variables is selected for each regression tree. The random forest has three tuning parameters: The minimum leaf size of the trees, the number of trees and the number of explanatory variables used for each tree. We use one third of all explanatory variables for each tree in accordance with the default value from Breiman (2001). The number of trees and the minimum leaf size are determined by 10-fold cross-validation on the training set.

C.5 Semiparametric Least-squares Support Vector Regression

Suykens and Vandewalle (1999) introduce a least-squares version of the support vector machine classifier. Enticed by the promising results from a study by Nazemi and Fabozzi (2018), we make use of a semiparametric least-squares support vector regression (SP LS-SVR) model which assumes the impact from the S different seniority classes to be linear. The parameter C regularizes the quadratic errors u_{sj}^2 while N denotes the number of defaulted bonds and $\mathbf{W}$ denotes the weight vector of the independent variables. The kernel function for the feature mapping into the higher-dimensional space is defined as $\phi(X_i)$ while the kernel matrix $\mathbf{K}$ is defined as $K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j)$. β is a vector of fixed effects for the seniority of the respective group and the dummy variables for the seniority classes are denoted by z_{sj}.

$$\begin{aligned} &\min\ J(W, b, u_i) = \frac{1}{2}||W||^2 + \frac{1}{2}\beta^T\beta + \frac{1}{2}b^2 + \frac{C}{2}\sum_{s=1}^{S}\sum_{j=1}^{n_s} u_{sj}^2 \\ &\text{s.t. } RR_i = \mathbf{W}^T\phi(X_i) + b + \beta^T z_{sj} + u_{sj}, \qquad j = 1, ..., n_s; s = 1, ..., S \end{aligned} \tag{C.2}$$

The Lagrangian function of this optimization problem evaluates to

$$L(W, b, u_{sj}, \alpha_{sj}) = J(W, b, u_{sj}) - \sum_{s=1}^{S}\sum_{j=1}^{n_s} \alpha_{sj}(W^T\phi(X_{sj}) + b + \beta^T z_{sj} + u_{sj} - RR_{sj}) \tag{C.3}$$

Therefore, with $\mathbf{V}$ denoting a $N \times N$-matrix of ones and $\mathbf{Z}_{ij} = z_{sj}^T z_{sj}$, the dual formulation is

$$\min \frac{1}{2}\alpha^T\mathbf{K}\alpha + \frac{1}{2}\alpha^T\mathbf{Z}\alpha + \frac{1}{2}\alpha^T\mathbf{V}\alpha + \frac{1}{2C}\alpha^T\alpha - RR^T\alpha \tag{C.4}$$

C.6 Selection of Macroeconomic Variables

In our analysis we include 182 macroeconomic variables to account for time variation in the recovery rates due to macroeconomic changes. A broad range of variables from categories such as international competitiveness, stock market conditions, credit market conditions, micro-level conditions, and business cycle conditions are taken into consideration. We compare three methods to select the most informative macroeconomic variables: The stability selection, the SparseStep algorithm, and the MC+ algorithm.

C.6.1 Least absolute shrinkage and selection operator

Tibshirani (1996) introduced LASSO, a regularized least squares method imposing a penalty on the L_1 norm of the regression coefficients. LASSO estimates the regularized coefficients $\hat{B}$ as follows:

$$\{\hat{B}\} = \arg\min_{B} \|RR - XB\|_2^2 + \lambda\|B\|_1 \tag{C.5}$$

where λ denotes the non-negative LASSO regularization coefficient, B denotes the LASSO regularization loadings and X_n denotes the $N \times 1$ vector $X_n = (x_1, ..., x_N)'$ of macroeconomic variables. By selecting a subset of the macroeconomic variables and eliminating the rest of the variables, the resulting model becomes more interpretable and has a higher out-of-sample predictive accuracy than the complete model. In particular, Nazemi and Fabozzi (2018) show that recovery models with macroeconomic variables selected by LASSO outperform models with a few macroeconomic variables.

As shown by Meinshausen and Bühlmann (2010) the variables selected from a LASSO regression can change with a small perturbation of the data. To address this issue, they introduce stability selection which is based on subsampling and evaluating the selection probability of each variable. Using stability selection, variable selection is conducted repeatedly on random samples of the dataset and the number of times each variable is selected during this process is counted. Only the variables that have been selected with a higher relative frequency than the specified counting proportion are ultimately selected by the stability selection method.[1] The main goal of this approach is to model high-dimensional data by a stable selection of macroeconomic variables that capture the most information for recovery rate estimation. We check the robustness of the variables' selection from LASSO by applying stability selection with a counting proportion equal to 0.6.

[1] We make use of the scikit-learn package in *Python* for the stability selection algorithm.

C.6.2 SparseStep

Burg, Groenen, and Alfons (2017) present the SparseStep algorithm. While LASSO penalizes the L_1 norm, the SparseStep algorithm imposes a penalty on the counting norm L_0. Burg, Groenen, and Alfons (2017) apply the following approximation to the counting norm L_0:

$$\|\beta_l\|^0 \approx \frac{\beta_l^2}{\beta_l^2 + \gamma^2} \tag{C.6}$$

where γ denotes a positive constant, β_l denotes the l-th coefficient, and p is the number of independent variables. To arrive at a sparse solution, the approximation to the exact counting norm L_0 is added for regularization:

$$\{\hat{\beta}\} = \arg\min_{\beta} \|RR - X\beta\|_2^2 + \lambda \sum_{l=1}^{p} \frac{\beta_l^2}{\beta_l^2 + \gamma^2} \tag{C.7}$$

While LASSO is a biased estimator, the SparseStep algorithm yields unbiased estimates of the parameter vector. Further, Burg, Groenen, and Alfons (2017) argue that SparseStep often outperforms approaches used in earlier studies such as ridge regression or LASSO in both model fit and prediction accuracy.[2]

C.6.3 MC+ algorithm

Zhang (2010) introduced MC+ for penalized variable selection in high-dimensional linear regression.[3] This method is based on two elements: a minimax concave penalty and a penalized linear unbiased selection algorithm. While LASSO estimates are biased, MC+ provides nearly unbiased estimates. Zhang (2010) outlines the theoretical and empirical advantages of MC+ compared to LASSO. In particular, he illustrates the increased selection accuracy of MC+ in a simulation setting.

[2] For the SparseStep algorithm we make use of package 'sparsestep' in R.
[3] We use the package 'plus' in R for the MC+ algorithm.

C.7 Ranking Variables by Permutation Importance

Altmann, Toloşi, O.Sander, and Lengauer (2010) outline how feature importance derived from random forests is biased towards categorical predictors with a large number of categories. In particular, they show that permutation importance is an importance measure that does not suffer from this bias.[4] Permutation importance is based on the mean decrease of prediction accuracy and is computed as the difference between the baseline R-squared of the model and the R-squared of the model when one variable's or group of variables' values are permuted randomly. In our analysis the permutation importance of each group is scaled such that the importance of the most important group of variables equals 100. Strobl, Boulesteix, Kneib, Augustin, and Zeileis (2008) show that permutation importance suffers from a bias towards correlated variables. Building groups of variables instead of investigating the importance of each variable on its own enables us to generate a ranking that will suffer less from the multicollinearity inherent to our high-dimensional data. Following Gregorutti, Michel, and Saint-Pierre (2015), we adjust the importance of each group by dividing it by the number of variables in the respective group.

[4] We use the implementation of permutation importance from the Python package 'pimp'.

Table C.1: Groups of independent variables

Seniority	
Senior unsecured	Senior secured
Senior subordinated	Subordinated
Industry	
Utility	Financials
Communication	Consumer-cylical
Industrial	IndustryDistress1
IndustryDistress2	
Bond Characteristics	
Zero Coupon	Variable Coupon
Step-up	Convertible
Insured	Retail Note
Corporate medium-term note	
News	
NewsVIX	Government
Intermediation	Natural Disaster
Securities	War
Other	
Financial Conditions	
Loans and Leases in Bank Credit, All Commercial Banks	Real Estate Loans, All Commercial Banks
Federal Debt: Total Public Debt	Total Consumer Credit Owned and Securitized, Outstanding
Excess Reserves of Depository Institutions	Commercial and Industrial Loans, All Commercial Banks
Total Borrowings of Depository Institutions from the Federal Reserve	Bank Credit of All Commercial Banks
Household Debt Service Payments as a Percent of Disposable Personal Income	Household Financial Obligations as a percent of Disposable Personal Income
Loans and Leases in Bank Credit, All Commercial Banks	Nonperforming Total Loans (past due 90+ days plus nonaccrual) to Total Loans
Nonperforming Loans to Total Loans (avg assets betw. USD 100M and 300M)	Net Loan Losses to Average Total Loans for all U.S. Banks
Total Net Loan Charge-offs to Total Loans for Banks	Return on Average Equity for all U.S. Banks
Loan Loss Reserve to Total Loans for all U.S. Banks	Nonperforming Commercial Loans (past due 90+ days plus nonaccrual) to Commercial Loans
Monetary Measures	
M2 Money Stock	Consumer Price Index for All Urban Consumers: All Items Less Food
University of Michigan Inflation Expectation	Consumer Price Index for All Urban Consumers: Energy
Personal Saving	Personal Saving Rate
Gross Saving	Gross Domestic Product: Implicit Price Deflator
Consumer Price Index for All Urban Consumers: Apparel	Consumer Price Index for All Urban Consumers: All Items
Consumer Price Index for All Urban Consumers: Medical Care	Consumer Price Index for All Urban Consumers: Transportation
Consumer Price Index for All Urban Consumers: All items less shelter	Consumer Price Index for All Urban Consumers: All items less medical care
Consumer Price Index for All Urban Consumers: Durables	Consumer Price Index for All Urban Consumers: Services
Consumer Price Index for All Urban Consumers: Commodities	Board of Governors Monetary Base, Adjusted for Changes in Reserve Requirements
M1 Money Stock	M3 for the United States
All-Transactions House Price Index for the United States	
Corporate Measures	
Corporate Profits After Tax (without IVA and CCAdj)	Corporate Profits After Tax with Inventory Valuation and Capital Consumption Adjustments
Corporate Profits after tax with IVA and CCAdj: Net Dividends	Corporate Net Cash Flow with IVA

Table C.1 (Continued)

Business Cycle	
Real Gross Domestic Product	ISM Manufacturing: PMI Composite Index
Industrial Production Index	University of Michigan: Consumer Sentiment
Private Nonresidential Fixed Investment	Real Disposable Personal Income
National income	Personal Income
Manufacturing Sector: Real Output	Real Personal Consumption Expenditures
Industrial Production: Manufacturing (NAICS)	Personal Consumption Expenditures: Durable Goods
Government Consumption Expenditures & Gross Investment	Gross Private Domestic Investment
Civilian Unemployment Rate	Continued Claims (Insured Unemployment)
Average Weekly Hours of Production and Nonsupervisory Employees: Mfg	Civilian Employment
Civilian Employment-Population Ratio	Persons unemployed 15 weeks or longer, as a percent of the civilian labor force
Manufacturers' New Orders: Durable Goods	Real Final Sales of Domestic Product
Manufacturers' New Orders: Nondefense Capital Goods Excluding Aircraft	Total Business: Inventories to Sales Ratio
Capacity Utilization: Manufacturing	Change in Private Inventories
Capacity Utilization: Total Industry	Total Business Inventories
Light Weight Vehicle Sales: Autos & Light Trucks	Housing Starts: Total: New Privately Owned Housing Units Started
Housing Starts: Total: New Privately Owned Housing Units Started	New One Family Houses Sold: United States
New Private Housing Units Authorized by Building Permits	Final Sales to Domestic Purchasers
Value of Manufacturers' New Orders for Consumer Goods Industries	Value of Manufacturers' Unfilled Orders for Durable Goods Industries
Avg Weekly Overtime Hours of Production and Nonsupervisory Employees: Mfg	Avg Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing
Avg Hourly Earnings of Production and Nonsupervisory Employees: Construction	Avg Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing
Real Manufacturing and Trade Industries SalesAdjusted	Personal consumption expenditures: Durable goods (implicit price deflator)
Personal consumption expenditures: Nondurable goods (implicit price deflator)	Personal consumption expenditures (implicit price deflator)
Housing Starts in Midwest Census Region	Housing Starts in Northeast Census Region
Housing Starts in South Census Region	Housing Starts in West Census Region
Initial Unemployment Claims	Real Manufacturing and Trade Inventories
Industrial Production: Business Equipment	Industrial Production: Consumer Goods
Industrial Production: Durable Consumer Goods	Industrial Production: Durable Materials
Industrial Production: Final Products (Market Group)	Industrial Production: Fuels
Industrial Production: Manufacturing (SIC)	Industrial Production: Materials
Industrial Production: Nondurable Consumer Goods	Industrial Production: Nondurable Materials
Manufacturers: Inventories to Sales Ratio	Nonfarm Private Construction Payroll Employment
Nonfarm Private Financial Activities Payroll Employment	Nonfarm Private Goods - Producing Payroll Employment
Nonfarm Private Manufacturing Payroll Employment	Nonfarm Private Service - Providing Payroll Employment
Total Nonfarm Private Payroll Employment	Nonfarm Private Trade, Transportation, and Utilities Payroll Employment
New Private Housing Units Authorized by Building Permits in the Midwest	New Private Housing Units Authorized by Building Permits in the Northeast
New Private Housing Units Authorized by Building Permits in the South	New Private Housing Units Authorized by Building Permits in the West
Number of Civilians Unemployed for 5 to 14 Weeks	Number of Civilians Unemployed for 15 Weeks and Over
Number of Civilians Unemployed for 15 to 26 Weeks	Number of Civilians Unemployed for 27 Weeks and Over
Number of Civilians Unemployed for Less Than 5 Weeks	Average (Mean) Duration of Unemployment
Consumer Opinion Surveys: Confidence Indicators: OECD Indicator for the US	Growth rate of nominal GDP
Growth rate of Nominal Dispoable Income	

Table C.1 (Continued)

Stock Market	
S&P 500 Index return	S&P 500 Volatility 1m
CBOE DJIA Volatility Index	NASDAQ 100 Index return
CBOE NASDAQ 100 Volatility Index	Russell 2000 Price Index return
Russell 2000 Vol 1m	Wilshire US Small-Cap Price Index return
Wilshire Small Cap Vol	
International Competitiveness	
Real Trade Weighted U.S. Dollar Index: Broad	Trade Weighted U.S. Dollar Index: Major Currencies
Total Current Account Balance for the United States	Real Exports of Goods & Services
Balance on Merchandise Trade	Real imports of goods and services
Canada / U.S. Foreign Exchange Rate, Canadian Dollars to One U.S. Dollar	Japan / U.S. Foreign Exchange Rate, Japanese Yen to One U.S. Dollar
Switzerland / U.S. Foreign Exchange Rate, Swiss Francs to One U.S. Dollar	U.S. / U.K. Foreign Exchange Rate, U.S. Dollars to One British Pound
Real Broad Effective Exchange Rate for United States	
Micro-level	
Manufacturing Sector: Unit Labor Cost	Nonfarm Business Sector: Unit Labor Cost
Compensation of employees: Wages and salaries	Compensation of employees: Mfg: Nondurables: Food, beverage and tobacco
Employment Cost Index: Total comp in Management, professional, and related	Manufacturing Durable Goods Sector: Compensation
Employment Cost Index: Benefits: Private Industry Workers	Employment Cost Index: Total comp for civilian workers in all industries and occupations
Employment Cost Index: Wages & Salaries: Private Industry Workers	1-Month AA Nonfinancial Commercial Paper Rate
10-Year Treasury Constant Maturity Rate	3-Month AA Nonfinancial Commercial Paper Rate
TermStructure	Effective Federal Funds Rate
Moody's Seasoned Baa Corporate Yield Relative to Yield on 10-Year Treasury	Moody's Seasoned Aaa Corporate Bond Yield
30-Year Conventional Mortgage Rate	Moody's Seasoned Baa Corporate Bond Yield
Bank Prime Loan Rate	Producer Price Index for All Commodities
Producer Price Index by Commodity Industrial Commodities	Producer Price Index by Commodity Intermediate Energy Goods
Producer Price Index by Commodity for Crude Energy Materials	Producer Price Index by Commodity for Finished Consumer Goods
Producer Price Index by Commodity Intermediate Materials	6-Month Treasury Bill: Secondary Market Rate
1-Year Treasury Constant Maturity Rate	5-Year Treasury Constant Maturity Rate
3-Month Treasury Bill: Secondary Market Rate	3-month Treasury Constant Maturity Rate
Producer Price Index by Commodity for Final Demand: Finished Goods	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate
Moody's Seasoned Baa Corporate Bond Minus Federal Funds Rate	3-Month Commercial Paper Minus Federal Funds Rate
Moody's Seasoned Aaa Bbb Spread	Size of High Yield Market in U.S. Dollars
High Yield Default Rate, Trailing 12-month	Bond defaults within the industry (in percent)