SEMIPARAMETRIC ESTIMATION OF FINANCIAL RISK: CORPORATE DEFAULT, CREDIT RATINGS, AND IMPLIED VOLATILITY

by

YIXIAO JIANG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Roger Klein

And approved by

____

____

____

____

New Brunswick, New Jersey

May, 2019

ABSTRACT OF THE DISSERTATION

Semiparametric Estimation of Financial Risk: Corporate Default, Credit Ratings, and Implied Volatility

By YIXIAO JIANG

Dissertation Director:

Roger Klein

There are contexts in which it is important to estimate a model without overly assuming functional forms and distributions. For this reason, extant empirical work often considers semiparametric single-index models: that is, objects of interest depends on the explanatory vector x through a single linear index $x^t\beta_0$. However, as suggested by economic/financial theories, it is natural to consider models in which covariates interact more freely with each other through multiple indices. This dissertation consists of three chapters featuring the formulation and application of semiparametric, multiple-index methods in finance, spanning corporate default modeling, conflicts of interest in credit ratings, and option implied volatilities.

In the first chapter, I introduce the econometric framework. As the number of indices increases, one technical difficulty that impedes statistical inference is to control bias terms of higher dimensional conditional expectation estimators. To control for this bias, I employ a differencing approach (see, Shen and Klein, 2017) which is known to reduce the bias to any order. However, there is no proof for asymptotic normality for a general multiple-index model and this result is critical for making

inferences. Here, I obtain asymptotic normality (conjectured but not proven in Shen and Klein, 2017) by establishing a novel U-statistic equivalence result that utilizes the theory of empirical process developed by Eddy and Hartigan (1977). I also provide institutional background for the empirical substances of this dissertation and a brief literature review.

The second chapter covers a semiparametric, ordered-response model of credit rating in which ratings are equilibrium outcomes of a stylized cheap-talk game. The proposed model allows the rating probability to be an unknown function of multiple indices permitting flexible interaction, non-monotonicity, and non-linearity in marginal effects. Based on Moody's rating data, I examine credit rating agencies' (CRAs) incentive to bias ratings when the CRA's shareholders invest in bond issuers. I find the degree of Moody's rating bias varies significantly for both rating categories as well as the institutional cross-ownership between Moody's and the bond issuer.

In the third chapter[1], we consider an ordered-response model in which the threshold parameters are random and can correlate with some or all covariates. We use a control function approach to identify the index coefficients and provide a novel identication and estimation strategy for the conditional threshold points up to location and scale. As a leading example, we consider estimation of the so-called "soft adjustment" — adjustments made by CRA based on unobserved and possibly subjective criteria — in the credit rating process. Empirically, we find a significant reduction of Moody's soft adjustment after the Dodd-Frank reform.

Chapter 4 develops a Hausman type specification test for a partially linear model against a semiparametric bi-index alternative which permits interaction effects. Using recent SP 500 index traded options data, we confirm that a partially linear model permitting a flexible "volatility smile" as well as an additive quadratic time effect is a statistically adequate depiction of the implied volatility data.

[1]This chaper is based on a joint work with Zhutong Gu, currently an Assistant Professor at Peking University-HSBC Business School, and Shuyang Yang, an Economist at Amazon. We started this paper in 2016 when all three of us were Ph.D. students at Rutgers University.

Acknowledgements

I was extremely blessed to interact with many amazing individuals who made my decade-long journey at Rutgers, from undergraduate to Ph.D., meaningful. I would like to first thank my advisor, Professor Roger Klein, for his unshaken faith in me. Roger influenced me not only as an econometrician with relentless energy but also as an omniscient yet humble advisor who treats students as equals. I would not be pursuing a career in academia if it were not for him. His influence will remain with me and continue to motivate and empower me in the days to come.

My equal gratitude goes to Professor Bruce Mizrach for being an extraordinary mentor. Bruce introduced me to the world of finance, where the empirical substance of this dissertation stems from, and has been lighting the way for me. I am also indebted to Professors John Landon-lane and Chan Shen for serving in my dissertation committee and providing constructive comments. Improvement of this dissertation also benefits from discussions with Tom Fomby, Simi Kedia, Diana Knyazeva, Yuan Liao, Daniel Millimet, Hilary Sigman, Norm Swanson, Bin Xie, Xiye Yang, and participants of varioius conferences, whom I owe sincere gratitude as well.

I appreciate all economics faculty members for exemplifying professionalism, especially Professors Douglas Blair, Todd Keister, and Richard P. Mclean. Many thanks to my fellow colleagues. I will for sure miss the days and nights talking about econometrics with Zhutong Gu and Shuyang Yang; the happiness and sorrow fighting along side Geoff Clarke, Mingmian Cheng, and Ryuichiro Izumi on the job market; the dusk till dawn in New Jersey Hall with Mark Avery, Sung hoon Choi, Hyungil (Henry) Kye, Jun Liu, and many other young economists. I appreciate the staff members, especially Linda Zullinger, for their assistance and support.

I also would like to acknowledge several groups of individuals that makes my spare time at Rutgers enjoyable. The first batch of gratitude goes to my Scarlet Knights friends from undergraduate for our decade-long friendship, especially Yang Kou, Zheqian Tong, and Ziyi Tang. I also thank my “family” at Piscataway New Jersey, Qiang Chen, Yangeng Lin, Lu Wang, Weinan Yan, Jiaxi Ye, and Wanqi Zhu. Last but not least, I gratefully acknowledge the “pot-love” group, co-led by Han Liu and Zixin Lin, for throwing a cherry on top of my Ph.D. voyage; our road-trip over 6 states in the U.S. (including a march into the Arctic Circle!) has been the highlight.

Finally, I wanted to dedicate this dissertation to my family. Words cannot articulate my gratitude to my parents, grandparents, and other kinsfolk in Beijing for their unconditional love and support, which I felt so dearly ocean apart. I thank my uncle and aunt in New York who treat me like their own son. This last word of acknowledgment I have saved for my wife, Siyu Wang, whom I love and respect *ad infinitum*.

Dedication

To my parents, Zhihui and Bin, and wife, Siyu

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

There are contexts in which it is important to estimate a model without overly assuming functional forms and distributions. For example, economists may not want to assume agents utilities are Cobb-Douglas or the distribution of shocks is known ex-ante. To strike a balance between flexibility and precision-loss in estimation, econometricians combine the two extreme approaches (fully nonparametric vs parametric) and make the model "semiparametric": that is, the estimated model contains a finite-dimensional parameter β_0 as well as an unknown function, $m(\cdot)$. Despite decades of development, much of the empirical work is still based on estimation of semiparametric single-index models (SIMs): that is, objects of interest depends on the explanatory vector x through a single linear combination $x^t\beta_0$. While the SIMs considerably relax distribution and functional form assumptions, they still impose restrictions. For example, ratios of marginal effects are constant when calculated as derivatives for continuous random variables.

In this dissertation, I am concerned with estimation and inference problems in models with a flexible functional form, as well as how to apply such methods in the realm of finance. By "flexible functional form" I mean the Conditional Expectation Function (CEF) may depend on the explanatory variables through an unknown function of a potentially multi-dimensional index vector. To be more precise, consider a model in which the dependent variable Y is driven by an explanatory vector $x =$

$[x_1, x_2 ... x_d] \in R^k$ such that,

$$E[Y|X] = m(x^t\beta_0) \quad (1.1)$$

$$\text{where} \quad x^t\beta_0 \equiv [x_1^t\beta_{10}, x_2^t\beta_{20} ... x_d^t\beta_{d0}] \in R^d \quad (1.2)$$

where $m(\cdot) : R^d \to R$ is the true, but unknown CEF function. Note that in the two extreme cases that (1) when $d = 1$, we have the single index model (SIMs), which has been studied extensively in the literature by many[1], and (2) when $d = k$, we have a pure nonparametric model in which the stochastic relationship between x and y is entirely captured by the CEF function and does not depend on any finite-dimensional parameter β_0.

To estimate the model accurately and efficiently (e.g., the estimator $\hat{\beta}$ converges to the truth β_0 "fast" as the sample size goes to infinity), I must address the bias in estimating the CEF in a high dimensional setting. In this dissertation, I will estimate densities and CEFs using *kernels*. Heuristically, a kernel estimator for $E[Y|x^t\beta_0 = v_0]$ can be taken as a weighted average of Y_i for observations with $v_i \equiv x_i^t\beta_0$ that are close to v_0. This estimator, however, has biases and variances. In fact, variance of this estimator goes to zero at a slower rate as the continuous dimension of the problem increases. The intuition is as $dim(v)$ increases, it is hard to find a group which are comparable to v_0 in very dimension of v. Therefore, the effective sample that are used to estimate the CEF shrinks, inducing a larger variance. This is the so-called "curse-of-dimensionality".

To make the estimator converges fast to the truth, the rate of convergence on the bias needs to be increased, a.k.a, correct the bias, to combat the curse of dimensionality. Higher order kernels (HOKs) (Muller, 1984; Ichimura and Lee, 1991; Lee, 1995) are often used in the literature to correct biases so that the semiparametric estimator

[1] A partial list includes Klein and Spady (1993); Ichimura (1993); Powell *et al.* (1989); Manski (1985); Härdle and Stoker (1989); Horowitz and Härdle (1996); Klein and Sherman (2002)

can be properly located at the true parameter vector of interest. However, HOKs can result problems when the dependent variables are naturally bounded. For example, as confirmed in the empirical exercises in this dissertation as well as other studies, HOKs deliver estimated probability outside of [0,1] rendering estimation results in difficult to interpret. Set against this background, Shen and Klein (2017) provide conditions on bias control to obtain asymptotic normality with regular kernels. The authors conjecture that a U-statistic result holds under their "recursive differencing" strategy. For single-index models, this result clearly holds. However, because of the complex structure of the estimator, a standard U-statistics argument is difficult to employ in higher dimensions.

A critical econometric contribution of this dissertation is that I prove a U-statistic equivalence result. This result, in its core, states that in certain circumstances discussed below, it is permissible to replace the recursive differencing estimator with a regular Nadaraya-Watson estimator in the limit. By leveraging this equivalence result, I show that asymptotic normality of index parameter estimators can be easily achieved for an arbitrary number of indices, which formally verifies Shen and Klein (2017)'s conjecture. This result also applies to estimating and testing a large class of semiparametric models.

In this dissertation, I will estimate a variety of empirical models in finance, spanning from default risk, credit ratings, and implied volatilities. These models are essentially variants of the general econometric framework described in Eq. 1.1 and 1.2. Relaxing the functional form restrictions has tremendous advantages in evaluation of financial models. For example, in the first two chapters, I evaluate how Credit Rating Agencies (CRAs) assess default risk and assign credit ratings. It is difficult in this context to even have a rough idea of what information is relevant or how the CRA utilize information to produce ratings. Consequently, when it comes to estimation, econometricians have little knowledge on the CEF that maps the default risk predictors to the rating probabilities. In Chapter 3, I estimate the implied volatility surface, which is a two-dimensional function that maps time-to-maturity and moneyness to option implied volatility. While inverting the Black-Scholes formula gives a closed-form solution, such a functional form is only valid for European style options and when the set of Black-Scholes assumptions are met. In both estimating credit rating and implied volatility models, lifting the restriction on CEFs clearly makes econometric analysis more flexible as well as robust.

To proceed, I first preview a U-equivalence result in Section 1.1, which is the foundation for making inference in models studied in the remaining chapters. In Section 1.2 and 1.3, I describe the empirical substances of this dissertation, namely the conflicts of interest in credit ratings and estimation of implied volatility. I then review the relevant finance literature on those subjects in Section 1.4.

1.1 A U-Statistic Equivalence Result

Let E denote the CEF function $E[Y|X = x] \equiv E(v, \theta_0)$ in the model under a multiple-index restriction, v the index vector, τ a trimming function that removes small denominator when estimating E. For reasons that will be discussed later, I employ the "recursive differencing" estimator for $E(v, \theta_0)$ proposed by Shen and Klein (2017):

$$\widehat{E^*_{t+1}(v,\theta)} := \frac{\sum_i (Y_i - \widehat{\Delta^t_{iv}}) K_h(V_i - v)}{\sum_i K_h(V_i - v)} \tag{1.3}$$

in which the kernel function $K_h(\cdot)$ is the normal probability density. The component $\widehat{\Delta^t_{iv}} \equiv \widehat{E^*_{t-1}(V_i,\theta)} - \widehat{E^*_{t-1}(v,\theta)}$ here, which will be estimated "iteratively", plays a role in reducing the bias[2]. However, since Δ itself is a difference of ratios, the large sample distribution of estimators involving $\widehat{E^*_{t+1}}$ is very difficult to study.

One econometric contribution of this dissertation is to prove the asymptotic

[2]When $\Delta = 0$, this is the standard Nadaraya-Watson (NW) estimator. Starting with the NW estimator, Shen and Klein (2017) shows that one can iteratively compute $\widehat{\Delta^t}$ and then $\widehat{E^*_{t+1}}$. Effectively, this method can be applied for $t = 1, 2, \ldots$ to control the bias in estimating E to any order.

normality results of the estimator for θ when the index dimension is $d > 1$. This result will be formally proved in its entirety in Chapter 2. Here I sketch the key steps. To be concrete, I take the normalized index parameter, denoted by θ, in Model (1.1) as an example. After a standard Taylor expansion at θ_0, $\sqrt{N}(\hat{\theta} - \theta_0)$ depends on the score (or gradient):

$$\sqrt{N}\hat{G}(\theta_0) = \mathcal{A}(E, \hat{w}) + \mathcal{B}ias(\hat{E}^*, \hat{w})$$

where the estimated weight function $\hat{w}_i = \alpha_i \nabla_\theta \widehat{E}(v_i, \theta_0)$ and

$$\begin{aligned}
\mathcal{A}(E, \hat{w}) &\equiv N^{-1/2} \sum_i \tau_i (Y_i - E_i)\hat{w} = \overbrace{\mathcal{A}(E, w)}^{apply\ CLT} + o_p(1) \\
\mathcal{B}ias(\hat{E}^*, \hat{w}) &\equiv N^{-1/2} \sum_i \tau_i (\widehat{E_i^*} - E_i)\hat{w} \\
&= \underbrace{N^{-1/2} \sum_i \tau_i (\widehat{E_i^*} - E_i) w \frac{\hat{g}}{g}}_{\mathcal{B}ias(\hat{E}^*, w)} + o_p(1)
\end{aligned}$$

g here is the joint density of the index vector v. The recursive structure of $\widehat{E^*(v, \theta)}$ makes $\mathcal{B}ias(\hat{E}^*, w)$ difficult to analyze, so I replace it with another object that is easier to study. The following U-statistic equivalence result permits this replacement.

In particular, I show that

$$\mathcal{B}ias(\hat{E}^*, w) - \mathcal{B}ias(\hat{E}, w) = o_p(1)$$

where $\hat{E} \equiv \frac{\sum_i Y_i^K K_h(V_i - v)}{\sum_i K_h(V_i - v)}$ is the regular NW kernel estimator for conditional expectations. Importantly, I use a "residual property" of semiparametric derivatives $E[w|v_i(\theta)]_{\theta=\theta_0} = 0$ due to Witney Newey[3] to turn the above difference into the difference of two empirical processes. This is otherwise impossible because $E\{\frac{\hat{g}}{g}(\widehat{E_i^*} - \widehat{E_i})]\}$ is not zero since $\widehat{E_i^*}$ has lower bias. By leveraging the uniform convergence of

[3] A formal proof of this residual result can be found at Klein and Shen (2010). The authors thank Witney Newey for mentioning in a private communication

empirical processes established by Eddy and Hartigan (1977) and Nadaraya (1964), the claimed result follows. I refer the above result as "U-statistic equivalence" because after the replacement, $\mathcal{B}ias(\hat{E}, w)$ is a degenerate $U-$statistic which is $o_p(1)$.

1.2 Conflict of Interest in Credit Ratings

While the Credit Rating Agencie's (CRA) profits exploded with the growth of structured finance, the collapse of these highly rated securities in the last financial crisis has led to suspicions that ratings were indeed "too optimistic" during the boom years. One prevailing and plausible explanation for rating inflation is the conflicts of interest faced by the CRA. A long-standing conflict stems from the "issuer-paid" model, whereby CRAs are paid by the issuers seeking ratings and hence are incentivized to issue inflated ratings.

When issuers are allowed to approach different CRAs and pick whichever agency that provides the highest rating, the CRAs are adversely incentivized to knock out their competitors by providing exaggerated ratings. Once the aforementioned strategy is employed by all CRAs, the credit rating system will fall into a "bad equilibrium" in which ratings are bid up and no longer accurately inform true credit worthiness of the security. More devastatingly, after purchasing investments that are more likely to default than anticipated, it is the vast investors who pay the price for this morbid rating system – losing their savings, their homes and their jobs. Even though reality may not be as bad as the general equilibrium model predicts, we have witnessed that in the recent financial crisis, the consequences brought by the collapse of many AAA rated structured products are quite serious.

In addition to the pressure of losing business from their competitors, the increasingly public ownership of rating agencies might induce other conflicts of interests. Of the two biggest agencies Moody's became a public firm in 2001, while Standard& Poor's is part of the publicly traded McGraw-Hill Companies.

Being a publicly traded firm not only intensifies the pressure to grow and increase profits (Bogle, 2005), but also motivates the CRAs to be biased towards their own shareholders. Warren Bufett, a major investor in Moodys, had to answer questions in front of the Financial Crisis Inquiry Commission in 2010[4] because media reports alleged that Moodys' has been slow to downgrade Wells Fargo, an investee of Berkshire Hathaway[5]. Since large shareholders may extract private benefit through their govenance power or threat of exit, CRAs' rating decisions might possibly be affected by the economic interest of their large shareholders as well (Kedia *et al.*, 2017). This ownership relation between the CRA and Wall Street firms makes it even harder for the CRA to perform its role as an independent arbiter of risk.

In the aftermath of the 2008 crisis, there has been a heated debate about how to effectively regulate the financial environment, and there are heated debate among them. In the famous Dodd-Frank Wall Street Reform and Consumer Protection Act (Pub.L. 111203[6], H.R. 4173[7]; commonly referred to and henceforce as "Dodd-Frank"), an entire section aims to improve the regulation of credit rating agencies. This law required the SEC to establish clear guidelines for determining which credit rating agencies qualify as Nationally Recognized Statistical Rating Organizations (NRSROs). It also gave the SEC the power to regulate NRSRO internal processes regarding record-keeping and how they guard against conflicts of interest[8]. Instead of merely enforcing and strengthening oversight, other regulators advocate overhauling the issuer-pay model. The Franken-Wicker amendment to the Dodd-Frank financial reform law[9] would use a governmental entity to assign securities to qualified ratings agencies based

[4]http://www.philstockworld.com/2011/03/14/transcript-of-warren-buffetts-testimony-in-front-of-the-fcic/

[5]http://www.forbes.com/sites/halahtouryalai/2012/02/16/missing-from-moodys-downgrade-list-warren-buffetts-favorite-bank/

[6]https://www.gpo.gov/fdsys/pkg/PLAW-111publ203/html/PLAW-111publ203.htm

[7]https://www.congress.gov/bill/111th-congress/house-bill/4173

[8]https://www.sec.gov/spotlight/dodd-frank/creditratingagencies.shtml

[9]https://www.sec.gov/comments/4-629/4629-28.pdf

on capacity and expertise. Bernie Sanders[10] also aims to change the business model used by the credit ratings agencies to a nonprofit model, keeping it independent of Wall Street.

Is a separate and nonprofit credit rating system necessary for the purpose of regulation? To answer this question, one needs to understand in what circumstances would the issuer-CRA relation turn from profit to peril for the financial system as a whole. The first two chapters model address these points from different lenses. In Chapter 2, I construct an empirical measurement of the institutional cross-ownership relation and explore the its heterogeneous impact on the ratings. In Chapter 3, I formulate a semi-structural econometric framework to think about the impact of private information on ratings and how the CRA's discretion envolves before and after the Dodd-Frank Act.

1.3 From Credit Rating to Options: Does Volatility Risk Matter?

Apart from conflicts of interest, which potentially bias the rating, another critique on the CRA is that credit ratings are lagged indicators of firm default risk. For example, at the moment that Moody's announced a downgrade, this negative news is likely to be priced in other firm-specific assets already.

One reason leads to the untimeliness of credit rating is the trade-off between ratings accuracy and stability. According to Cantor and Mann (2006), investors, issuers, as well as regulators want ratings to reflect enduring changes in credit risk because rating changes have real consequences — primarily due to ratings-based portfolio governance rules— that are costly to reverse. Instead of simply tracking market-based measures of credit risk, ratings should reflect independent analytical judgments that provide counterpoint to often volatile market-based assessments. Should the CRA do its duty rightfully and diligently, one shall expect consistency between credit ratings

[10] https://berniesanders.com/statement-by-senator-bernie-sanders-on-wall-street-and-the-economy/

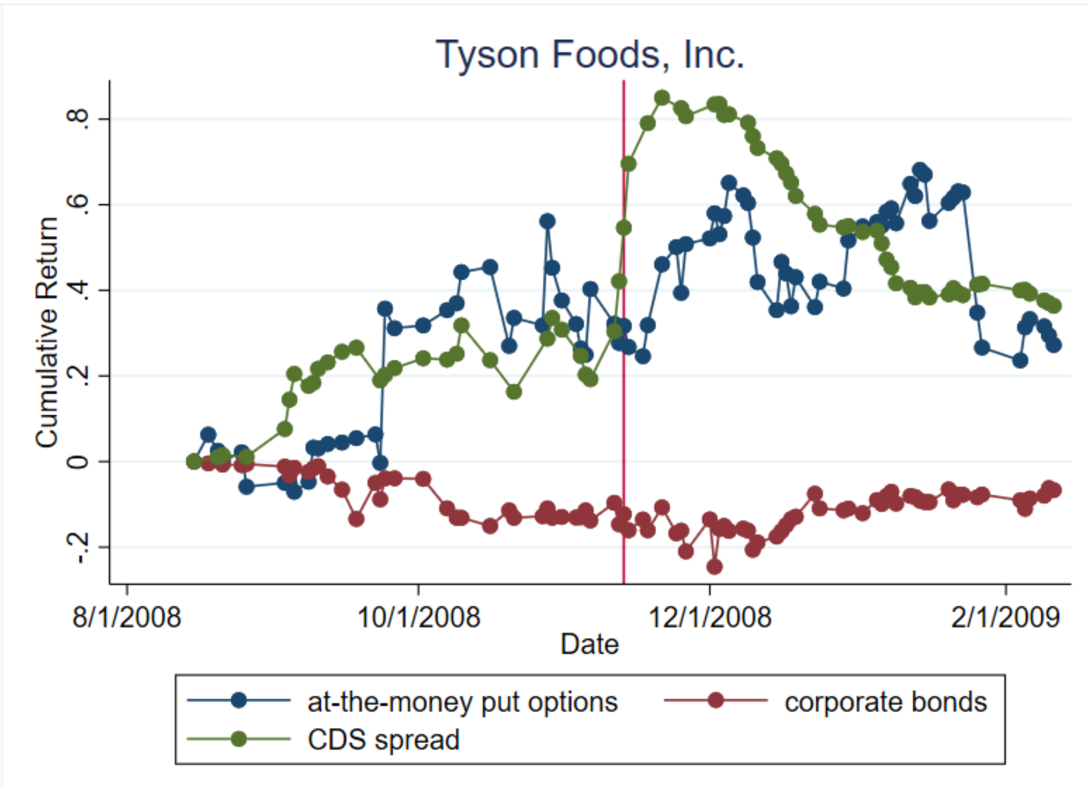
and market-based assessments on related securities, though the ratings often lagged the others.

Taken this view into account, I want to examine whether those market-based assessments capture additional component of risk that is not contained in credit ratings. Stock, bond, CDS, and options are all claims on the same firm assets. In the literature on asset pricing, numerous studies document that the price movement of these assets are closely correlated. The intuition for the *empirical* lead-lag relation relies on the trading motives of market participants. For example, bond investors who receive negative material nonpublic information about the firm have several options to reduce the downside credit exposure. They can either short the underlying bonds/stocks or long the put options/CDS. Such trading behaviors would lead to abnormal price movements in related securities; however, the content and speed of information discovery depends critically on the structure and trading environment in these markets as well as the payoff structure of the instruments. The relative rates of price discovery is illustrated in Figure 1.1, where we plot the responses of single-name asset returns (bond, option, and CDS) to the rating downgrade of Tyson Foods, Inc. on November 13, 2008. This figure shows both the put options and CDS spread rallied well ahead of the rating downgrade.

What drives the significant put option reactions before rating events? Any public information, such as the concurrently trading stocks and bond prices, cannot be the source of marginal option market information. As noted in Cremers *et al.* (2008), jump and volatility risk represent the additional information that traded individual options captured, which is not already captured by equity and riskless debt. On the assumption that corporate debt is a combination of riskless debt and a short potion in a credit put option struck at-the face value of the debt (Merton, 1974), bond price is linked to equity volatility in a structural way.

To more concretely understand this type of volatility, or "jump", risk, I need to turn to the estimation of option implied volatility. Option implied volatility is an

Figure 1.1: Bond, option, and CDS reactions to the downgrade of Tyson Foods, Inc. on 11/13/2008



Note: This figure plots the time-series of CDS spread, at-the-money put options, and corporate bond returns around a downgrade event on November 13, 2008 (The red vertical line). The returns are cumulative in the sense that the return at day t (each dot) is defined as the log-difference between the option or bond price at day t and the day that the series start (90 days before the event). CDS spread is provided by Bloomburg Terminal. Bond return data is from Trace. Option return data is from OptionMetrics. A detailed description of data collection can be found later in this paper.

important component in the Black-Scholes model and therefore affects the observed option prices. As shown in Aıt-Sahalia, Bickle and Stoker (2001, ABS henceforth), inverting the B-S formula with respect to the volatility parameter would give to the following model for Implied Volatility (σ_{iv}),

$$\sigma_{iv} = m(K/F_t, T) + \epsilon \quad \text{with} \quad E[\epsilon | K/F_t, T] = 0 \tag{1.4}$$

where K/F_t is the "moneyness" of an option and ϵ summarizes potential sources of noise, e.g., bid-ask spread. The unknown transformation $m(\cdot,\cdot)$ captures the dependency of IV on K/F_t and T. If options are indeed quoted based on B-S formula, one should expect a constant IV which does not vary across moneyness or time-to-expiration. However, it is found that out-of-the-money (OTM) put options, i.e., put options with $K/F_t < 1$, are traded at higher implied volatility than at-the-money (ATM) options and OTM calls, also known as "volatility smiles".

ABS further show that a semiparametric model permitting a flexible "volatility smile" as well as an additive quadratic time effect, i.e., $m(K/F_t, T) = g(K/F_t) + \theta_1 T + \theta_2 T^2$, is a statistically adequate depiction of the IV data. The above partially linear specification, however, rules out potential interaction effect between moneyness and time-to-expiration. That is, if one plots IV against moneyness and time-to-expiration, which gives the so-called *Implied Volatility surface*, a partially linear structure implies that the term structures of IV across different moneyness values should roughly have the same shape and only differ by a level shift, which is argued by many in the literature (See, for example Fengler (2006)).

My pointed interest in the third chapter in this dissertation is to rigorously test whether a partially linear model is a statistically adequate simplification of the general nonparametric model above. Taking the partially linear structure into account would increase the estimation efficiency of implied volatility models. The specification test that I proposed is a Hausman-type test (Hausman, 1978).

1.4 Literature Review

Much of academic debate has focused on the conflict of interest inherent in the issuer-pay model. This model and its equilibruim structure are studied by Bolton *et al.* (2012) and Sangiorgi *et al.* (2009). The conflict of interests can be characterized as a "trade-off" between providing accurate (and hence unflattering) ratings versus exaggerating on the investment but risking a potential loss on their reputation. In the empirical literature, researchers (Jiang *et al.*, 2012; Cornaggia and Cornaggia, 2013; Kraft, 2015) have also focused on compromised ratings on account of issuer-pay model.

When modeling the rating process, I follow the previous literature on bond ratings to select firm/bond characteristics that determine the ratings (e.g, Pinches and Mingo (1973); Kaplan and Urwitz (1979); Blume *et al.* (1998); Campbell and Taksler (2003); Jiang *et al.* (2012)). Researchers have estimated the rating process in various parametric forms, with the majority focusing on the linear probability and ordered probit models. In contrast, the two semiparametric models that we proposed allow the explanatory variable affect the rating in a much more flexible manner.

This dissertation also relates to the vast literature on the role of large shareholders. Barclay and Holderness (1989) and Admati and Pfleiderer (2009) found that large shareholders may extract private benefit through their govenance power or threat of exit. After their IPO, CRAs' rating decisions might possibly be affected by the economic interest of their large shareholders as well. The paper that motivates this dissertation the most is Kedia *et al.* (2017), in which the authors find Moodys assigned more favorable ratings to firms that are related to its large shareholders, and the favorable treatment cannot be explained by its private information.

1.5 Outline for this Dissertation

The remaining part of this disseration is organized as follow: In Chaper 2, I propose a semiparametric model for credit ratings and formally prove the aforementioned U-statistic equivalence result. I use this model to study the heterogeneous marginal effect of conflict. The asymptotic distribution of the proposed estimators are proved. In Chapter 3, we provide the first identification and estimation results for an ordered-model with heterogeneous thresholds and potentially endogenous regressors. We apply such methods to study the evolution of rating agency's discretion, namely how much of the rating dispersion is explained by public information vs private information, before and after the Dodd-Frank Act. In Chapter 4, I estimate a flexible model for option implied volatility and provide a specification test for it.

Chapter 2

Semiparametric Estimation of a Credit Rating Model

2.1 Introduction

While the Credit Rating Agencie's (CRA) profits exploded with the growth of structured finance, the collapse of these highly rated securities in the last financial crisis has led to suspicions that ratings were indeed "too optimistic" during the boom years. One prevailing and plausible explanation for rating inflation is due to the conflicts of interest faced by the CRA. A long-standing conflict stems from the "issuer-paid" model, whereby CRAs are paid by the issuers seeking ratings and hence are incentivized to issue inflated ratings.[1] In the past two decades, rating agencies are increasingly owned by large financial institutions, which induces a conflict of interest that is less obvious: CRAs can inflate ratings to benefit issuers that are controlled by their shareholders to cater to the economic interest of those shareholders. While much of the extant literature focused on issuer-paid models, this paper examines the empirical relationship between rating inflation and this often-neglected source of conflicts of interest—what I call *shared-ownership*—within a novel econometric framework.

Partially guided by a stylized "cheap-talk" framework, our econometric model allows a bond's latent default risk to be an unknown and potentially non-separable function of multiple indices and an error term. With each index being an unknown

[1] For theoretical studies on the issuer-paid model and rating shopping, see Bolton *et al.* (2012); Sangiorgi *et al.* (2009); Skreta and Veldkamp (2009) and some empirical evidence (Mathis *et al.*, 2009; Jiang *et al.*, 2012; He *et al.*, 2015)

linear combination of covariates, the three indices depend on firm characteristics, bond characteristics, and the Moody-firm-ownership-index (MFOI), which is a shared-ownership index that I introduce later in this paper. Consideration of a non-separable function is essential because this non-separable structure is implied by the equilibrium outcome of a structural framework devised to study the strategic interaction between the CRA and a representative shared owner. As the model is estimated semiparametrically, it is not necessary to know how CRAs use information, both public and private, at their disposal *a priori*. Estimates are robust to a wide class of utility functions assuming some regularity conditions. Because of the permitted interaction among indices, the marginal effects of one component of X, for example, X_1, can vary across subpopulations defined by the index values without constraints.

Our paper contributes to the empirical literature on the modeling of credit rating decisions and the econometric theory of bias controls. Turning to the empirical literature, one approach, employed by Kraft (2015); Campbell and Taksler (2003); Jiang *et al.* (2012), is to estimate a linear probability model for which the rating outcome is a linear combination of covariates and error terms. Constrained by its functional form, the model can only capture the average marginal effects and not the heterogeneity of the marginal effects. Another class of models (Kaplan and Urwitz, 1979; Horrigan, 1966; Blume *et al.*, 1998; West, 1970) defines a latent variable of theoretical interest (i.e., default risk) and specifies a parametric link function between covariates to the conditional choice probability. However, as found below, the functional forms underlying parametric models may not be correct and conflict with the prediction from an underlying behavioral model. Neither of the described approaches allow for a non-separable functional form and can be restrictive in many ways. Therefore, to avoid misspecification, it is important to have a flexible model specification.

Extensive literature addresses semiparametric models and the estimation of semiparametric single index models (SIMs) including Klein and Spady (1993);

Ichimura (1993); Powell *et al.* (1989); Manski (1985); Härdle and Stoker (1989); Horowitz and Härdle (1996); Klein and Sherman (2002). However, there are fewer results available on the estimation of multiple-index regression models. The identification of index coefficients in multiple-index models of this sort has been studied by Ichimura and Lee (1991); Lee (1995) and Ahn *et al.* (2017). However, this paper, to the best of our knowledge, is the first to consider estimating ordered models in a multiple-index context. To establish large sample results for the index parameter estimator, which are necessary for inferences, I must address the bias in estimating the conditional choice probability[2]. Shen and Klein (2017) provides conditions on bias control to obtain asymptotic normality with regular kernels. The authors conjecture that a U-statistic result holds under their "recursive differencing" strategy. For single-index models, this result clearly holds. However, because of the complex structure of the estimator, a standard U-statistics argument is difficult to employ in higher dimensions. In this paper, I verify their conjecture by proving a U-statistic equivalence result that holds for an arbitrary number of indices. This result applies to a large class of semiparametric models.

Using the Mergent's Fixed Income Securities Database(FISD) for the years 2001 to 2007, I estimate the aforementioned model and characterize marginal effects of the shared-ownership index $MFOI$. Since marginal effects in general will not be constant in nonlinear models, the contribution of this application to the pertaining empirical literature is to explore the heterogeneity of rating bias due to institutional cross-ownership. The empirical findings are twofold. First, I find that investment-grade bonds related to large shareholders of Moody's, particularly A-rated bonds, are most vulnerable to conflicts of interest. This result aligns with the observation that large shareholders may use their governance power and/or threat of exit to extract private

[2] Higher order kernels (Muller, 1984; Ichimura and Lee, 1991; Lee, 1995) are often used in the literature to correct biases so that the semiparametric estimator can be properly located at the true parameter vector of interest. However, as confirmed in our empirical exercises, higher order kernels can deliver estimated probability outside of [0,1] rendering estimation results difficult to interpret.

benefit (Admati and Pfleiderer, 2009; Edmans, 2009). The empirical results also imply that employing a flexible estimation framework is important: for A-grade bonds, the magnitude of rating bias is twice that of comparable parametric models.

Second, contrary to the common belief that bonds at the investment-grade/high-yield boundary are likely to benefit the most, I find Moody's does not assign favorable ratings to high-yield bonds regardless of the issuer's shared-ownership relation with Moody's. The second empirical finding is relatively original in the literature. One possible but speculative explanation is related to the "reputation capital" view (White, 2002; Becker and Milbourn, 2011; Bolton *et al.*, 2012): low quality bonds are more likely to default implying a higher probability of triggering *reputation loss*[3]. To protect its reputation, the CRA might be more conservative and self-disciplined when rating low quality bonds.

The rest of the paper is as follows. The next section presents a stylized cheap-talk model that guides our empirical investigations. Section 3 describes the rating data and how I use institutional shareholding data to measure conflicts of interest. Section 4 describes an econometric model for credit rating. Section 5 presents the main empirical findings, including estimates of index coefficients and heterogeneous marginal effects. Section 6 concludes. A more detailed description of the cheap-talk model is provided in Appendix A. Technical details/preliminaries concerning the econometric inference procedure are provided in Appendix B, followed by the formal asymptotic theorems in Appendix C.

2.2 Theoretical Motivation

To guide the empirical investigation, I study a stylized version of the "cheap-talk" model proposed by Crawford and Sobel (1982). This adapted version was devised to study the strategic interaction between a CRA and an informed shared owner — often

[3]This implies that the CRA will be "punished" once a highly rated investment results in default. See Bolton *et al.* (2012) for a discussion.

a large financial institution that owns both the CRA and the bond issuer. To streamline the discussion, I focus on the key prediction of the model and its empirical implication. A full description of the model including players' payoff functions and strategy are provided in the Appendix.

Let y^* denote a bond's latent default risk, which the CRA should estimate for the purpose of assigning ratings. I show that in equilibrium, y^* is driven by three components in a *non-separable* form:

$$y^* = y^*(V, m, b) \tag{2.1}$$

in which (i) V is a potentially multi-dimensional vector representing firm and bond characteristics that the CRA can observe such as a firm asset, leverage ratio, and subordination status; (ii) m represents the level of *soft information* that will be explained below, and (iii) b is a measure of the degree of conflict of interest between the CRA and its shareholder(s).

Here I make three observations about estimating the above model with empirical data.

1. The exact form of $y^*(\cdot,\cdot,\cdot)$ is generally unknown, necessitating a flexible estimation procedure. For this type of models, Crawford and Sobel (1982) shows that an equilibrium solution exists under quite general conditions, with some smoothness and shape restriction on utility functions. However, the exact formula of $y^*(\cdot,\cdot,\cdot)$ is often hard to compute analytically. In the Appendix I give a closed-form solution for the "uniform-quadratic" case in which players have quadratic utility functions and m is uniformly distributed. In this representative case, $y^*(\cdot,\cdot,\cdot)$ is a non-separable function with respect to m and b. Presumably y^* can take a very different functional form when players have non-quadratic utility functions. Therefore for estimation, it is essential to have a flexible model that can, at least, allow for non-separability.

2. On the substantive end, the above formulation in (A.3) reflects that credit risk is driven by both *hard information* and *soft information* in a potentially non-separable form. According to Petersen (2004), soft information represents factors that drive credit risk but cannot be completely summarized in numerical scores, such as a manager's abilities. For estimation, I treat the level of soft information, represented by m, as a regression error term.

3. To empirically study the model in (2.1), I identify variables to measure the hard information vector V and the conflicts of interest measure b. I assume that V is an unknown function of firm and bond characteristics, for example, $V \equiv G(F, B)$. The choice of firm characteristics F and bond characteristics B is discussed in Section 2.3.1. Using the institutional shareholding data (13f) from Thomson Reuters, I construct a variable, termed *Moody-Firm-Ownership-Index (MFOI)*, to characterize the institutional cross-ownership between the bond issuer and Moody's. This variable is formally defined in Section 2.3.2. Similarly, I assume that the conflict of interest b is an unknown function of $MFOI$, for example, $b = C(MFOI)$. Put it differently, I assume the degree of conflicts of interest is associated with the "liaison" between the bond issuer and Moody's.

2.3 Dataset and Variables

The data are derived from multiple sources. First, I obtain initial ratings on corporate bonds issued by firms from either CRSP or Compustat from Mergent's Fixed Income Securities Database (FISD). The sampling period begins in 2001, when Moody's went public, and ends in 2007 to prevent any confounding effect of the financial crisis and other subsequent regulation acts. I then obtain a number of firm characteristics from CRSP-Compustat to match the rating data. After combining data from multiple sources, the final dataset is composed of 4,967 bonds issued by 986 firms.

2.3.1 Firm and bond characteristics

Table 2.1 shows that a number of firm and bond characteristics (termed F_i, B_i, respectively) are selected as additional controls based on bond rating literature (Horrigan, 1966; West, 1970; Kaplan and Urwitz, 1979; Jiang *et al.*, 2012; Blume *et al.*, 1998). The explanatory variables are: (1) Firm leverage, defined as the ratio of long-term debt to total assets (LEVERAGE). (2) Operating performance, defined as operating income before depreciation divided by sales (PROFIT). (3) Issue size, defined as the par value of the bond issue (AMT). (4) Issuer size, defined as the value of the firm's total assets (ASSET), and (5) Subordination status, a 0-1 dummy variable that is equal to one if the bond is a senior bond (SENIORITY). (6) Stability variable (STABILITY), defined as the variance of the firm's total assets in the last 16 quarters. Firm-level variables are computed using a five-year arithmetic average of the annual ratios because Kaplan and Urwitz (1979) note that bond raters might look beyond a single year's data to avoid temporary anomalies.

Table 2.1: Firm and Bond Characteristics

Variable	Description	Mean	Std. Dev.	Min	Max
ASSET	log(asset) of the issuer	9.643	2.280	4.360	14.324
STABILITY	Variance of asset	0.230	0.169	0.003	1.416
LEVERAGE	Firm leverage ratio	0.264	0.178	0.002	1.212
PROFIT	Operating performance	0.026	0.058	-0.739	0.436
AMT	log(issuing amount)	12.224	1.681	2.708	19.337
SENIORITY	a bond's subordination status	0.809	0.393	0.000	1.000

2.3.2 Conflicts of interest

As noted above, conflicts of interest is measured by institutional cross-ownership between Moody's and a bond issuer. To characterize the degree of cross-ownership, I first obtain the list of Moody's shareholders and calculate their ownership stake in Moody's (the percentage of Moody's stock that they hold) for each quarter in the sampling period. Next, I access each shareholders investment portfolio to find

Table 2.2: Moody's large shareholders from 2001-2016

Shareholder	T	Mean	Max	Min
HARRIS ASSOCIATES L.P.	21	2.42%	5.02%	0.00%
CHILDREN'S INV MGMT (UK) LLP	20	2.29%	5.31%	0.01%
SANDS CAPITAL MANAGEMENT, INC.	28	3.01%	5.59%	0.40%
T. ROWE PRICE ASSOCIATES, INC.	64	1.47%	5.94%	0.18%
BARCLAYS BANK PLC	55	2.52%	6.32%	0.03%
GOLDMAN SACHS & COMPANY	63	1.94%	7.24%	0.01%
VALUEACT CAPITAL MGMT, L.P.	13	5.19%	7.77%	0.93%
VANGUARD GROUP, INC.	64	3.79%	7.98%	1.64%
MSDW & COMPANY	57	2.20%	8.14%	0.22%
DAVIS SELECTED ADVISERS, L.P.	51	5.56%	8.14%	0.10%
FIDELITY MANAGEMENT & RESEARCH	64	1.99%	9.08%	0.00%
CAPITAL RESEARCH GBL INVESTORS	13	4.80%	11.31%	0.07%
CAPITAL WORLD INVESTORS	35	6.07%	12.60%	0.66%
BERKSHIRE HATHAWAY INC.	64	14.87%	20.43%	11.33%

Note: In the second column (T), I report the number of quarters in the 16 year period (out of 64) that the shareholder in Column 1 has liaison with Moody's. In column 3 (4,5), I report that shareholder's average (min, max) ownership stake in Moody's (The percentage of Moody's stock owned).

out which bond issuers have the same shareholders as investors. The shareholder's manager type code (MGRNO) and the firm's Committee on Uniform Securities Identification Procedures (CUSIP) number are used to match the shareholding data with the 986 bond issuers. A list of Moody's large shareholders (shareholders who have owned more than 5% of Moody's from 2001-2016) is presented in Table 2.2.

To summarily characterize the shared-ownership relation between each bond issuer and Moody's from this large dataset, I propose the following aggregate measure. Suppose Moody's has $j = 1, 2, \cdots, M$ shareholders in a given quarter[4], and any subset of those shareholders can invest in an issuing firm. The key variable of interest, the MFOI, is defined as follows:

$$MFOI = \sum_{j=1}^{M} p_j \lambda_j \tag{2.2}$$

[4]Since all of the variable are time-specific, I drop the time t subscript for notation simplicity

where λ_j denotes shareholder j's ownership take in Moody's (the percentage of the CRA owned by the shared owner j), and p_j denotes issuing firm i's weight in shareholder j's investment portfolio (the percentage of the shareholder's portfolio accounted for by the issuing firm). I choose a product form because there are no conflicts of interest associated with shareholder j if either portion is zero.

Since institutional investors hold diverse portfolios, most p_j and λ_j take on small values[5], resulting in an extremely skewed to zero distribution: approximately 20% of the bonds in our sample are issued by firms that are not affiliated with Moody's at all. Most of the bonds come from firms whose investors are Moody's small shareholders. Only the top 5% of bonds are issued by firms with extremely large MFOI (those who are likely to be related to Moody's large shareholders).

2.4 Econometric Strategy

2.4.1 Model

Let $X_i \equiv (F_i, B_i, MFOI_i)$ be a vector composed of firm characteristics, bond characteristics, and the described shared-ownership relation proxy $MFOI$. Denote y_i^* as the latent default risk associated with a corporate bond. Based on the economic model for y^* described in Section 2, I estimate an ordered-response model in which the CRA assigns each bond with an ordinal rating $Y_i = 1, 2, 3 \ldots L$ based on y_i^* and a

[5] p = 0.25%, λ =0.07% are the 75 percentile cutoffs

series of cutoff points c_j[6] between rating categories:

$$\begin{aligned} Y_i &= \sum_{j=1}^{L} j\mathbf{1}\{c_{j-1} < y_i^* < c_j\}, \qquad (2.3) \\ y_i^* &= y^*(X_i, U_i) \\ \mathbf{1}\{E\} &: \text{ an indicator function of the event } E \end{aligned}$$

where U_i is a potentially multidimensional disturbance term representing the soft information. Motivated by the theoretical framework in (2.1), the function $y^*(\cdot,\cdot)$ that links default risk with hard/soft information is left unspecified and may be fully non-separable. Such a flexible non-separable structure, however, is precluded in ordered-probit/logit models in which y^* is assumed to be linear in X_i and U_i.

For the model defined above, a key component of estimation interest is $Prob(Y_i = j|X)$, which is the probability that a bond will be rated in category j given the set of explanatory variables. In a more general nonparametric formulation,

$$Pr(Y_i = j|X_i) = P_j(X_i), \text{ for } j = 1, 2, \cdots, L \qquad (2.4)$$

This specication imposes few restrictions on the form of the joint distribution of the data. Therefore, there is little room for misspecification, and the consistency of the estimator is established under more general conditions than is the case under parametric modeling (Powell, 1994). However, when the dimension of X is large, the resulting estimator will have considerable variance due to the "curse of dimensionality." To estimate the above probability with a moderately sized sample, I propose estimating this probability based on the following index assumption and making the model semiparametric:

---

[6]Credit ratings are coded as follow: $Aaa = 1, Aa = 2, \cdots, C = 7$. The cutoff points $c_j's$ may be fixed points, as in the case of ordered-probit/logit models. Alternatively, these cutoff points may be random variables from different distributions that are independent of the explanatory variables allowing the rating criteria to vary with issuers. The estimator employed in this paper allows for either possibility.

Assumption 1. *There exists a firm aggregator (or index) $V_F \equiv F_i\beta_0^F$, a bond aggregator $V_B \equiv B_i\beta_0^B$ and for a differentiable function $H(\cdot,\cdot,\cdot)$ such that for all category j:*

$$Prob(Y_i = j|X_i) = Prob(Y_i = j|F_i\beta_0^F, B_i\beta_0^B, MFOI_i) \equiv H_j(V_F, V_B, MFOI_i) \tag{2.5}$$

The above assumption states that X_i influences the ratings through three channels: a firm index $V_F = F_i\beta_0^F$, a bond index $V_B = B_i\beta_0^B$ and, most importantly, the institutional cross-ownership measure $MFOI_i$. Because it is nonparametric, the mapping $H_j(\cdot,\cdot,\cdot)$ allows the rating probability to be a flexible function permitting non-monotonicity and interactions in its arguments. Note also that the function H_j may vary by category; thus, the model allows the rating agency to have different criteria for each rating category. This type of "multiple-index" model, first proposed by Ichimura and Lee (1991), arises naturally in many applications where a single-index model cannot fully capture the underlying economic behaviors[7].

2.4.2 Quantile marginal effect

Recall that this paper aims at studying the effect of shared-ownership, as measured by $MFOI$, on credit ratings. Given the above model, it is convenient to define the impact of $MFOI$ on the rating as the cumulative change in (3.2) from a marginal increase in $MFOI_i$: for example, the probability of obtaining a *better* rating from a counterfactual change in $MFOI$:

$$\begin{aligned} ME(F_i, B_i, m_b; \Delta, K) &\equiv Prob(Y_i < K|F_i, B_i, m_b + \Delta) - Prob(Y_i < K|F_i, B_i, m_b) \\ &= \sum_{j=1}^{K-1} H_j(V_F, V_B, m^b + \Delta) - H_j(V_F, V_B, m^b), \end{aligned} \tag{2.6}$$

[7]Examples include sample selection (Klein *et al.*, 2015), extraneous variables (Stoker, 1986), and decision-making with multiple players (Lührmann and Maurer, 2008).

The second equality follows directly from the index assumption in (3.2). That is, for a bond initially rated in category K, I take a partial sum of the H_j- differentials from the highest credit rating category $j = 1$ to $j = K - 1$. Therefore, the marginal effect defined here effectively captures how much more likely it is that a K-rated bond will be rated at least to $K - 1$ when $MFOI$ increases from m^b to $m^b + \Delta$.

Importantly, the marginal impact of $MFOI$ can be sufficiently heterogeneous. As implied by Kedia *et al.* (2017), the impact of $MFOI$ may be significant only when m_b exceeds some threshold: that is, when an issuer is related to "large" shareholders of Moody's. To explore the heterogeneity of the shared-ownership effect acrss subpopulations defined by the value of m_b, denote the "quantile Marginal effects" (QME) as

$$QME(\mathcal{Z}_q; K) \quad \equiv \quad E[ME(F_i, B_i, m_b; \Delta, K)|m_b \in \mathcal{Z}_q] \tag{2.7}$$

That is, the unit-level marginal effects defined in (2.6) are averaged for observations with $MFOI_i$ in a particular quantile of interest $\mathcal{Z}_q$. This measure is best understood as a "local" version of the average marginal effect: instead of measuring the average impact of $MFOI$ for the entire sample, $QME(\mathcal{Z}_q; K)$ addresses how such an impact differs for issuers with different degrees of affiliation with the CRA. To obtain inference and test economic hypotheses, I derive the large sample distribution of the $QME(\mathcal{Z}_q; K)$ estimator.

2.4.3 Estimation

Note that the function H_j in (3.2) is not parametrically specified, it is well-known that identification of the index parameter vector β_0 is up to any multiplicative and additive constant, or the so-called identification is up to location and scale. More formally, I redefine $V_F = F_1 + F'\theta_0^F$ and $V_B = B_1 + B'\theta_0^B$ as functions of the identified parameter vector $\theta_0 \equiv [\theta_0^F, \theta_0^B]$, where $F_1(B_1)$ is the firm (bond) characteristic that is chosen for the normalization and $F'(B')$ is a vector for other firm(bond) covariates.

Estimation of the $QME(\mathcal{Z}_q; K)$ proceeds in two steps. The first step estimates the normalized index parameters $\theta_0 = [\theta^F, \theta^B]$. The second step computes the sample analogue of (2.7) with the (normalized) estimated index: $\widehat{V_{Fi}} \equiv F_1 + F'\hat{\theta}^F$, $\widehat{V_{Bi}} \equiv B_1 + B'\hat{\theta}^B$ and $MFOI_i$.

Step 1: Index parameter

More formally, the estimator is obtained by maximizing the following (log-) "quasi-likelihood:"

$$\hat{\theta} \in argmax\ Q(\theta) \equiv \sum_{i=1}^{N} \tau_i \{\sum_{k=1}^{L} Y_i^k Ln(P_i^k(\theta))\} \tag{2.8}$$

where $Y_i^k = \mathbf{1}\{Y_i = k\}$, $P_i^k(\theta) \equiv Prob(Y_i = k|X_i)$ is the probability that $Y_i = k$ conditional on the three indices, and τ_i is a trimming function that removes observations with poor estimates of $P_i^k(\theta)$. Under an appropriate trimming strategy and a residual property of semiparametric derivatives, asymptotic normality can be obtained with a regular kernel estimator for $P_i^k(\theta)$ for single-index models (Klein and Shen, 2010). However, in higher dimensions, additional bias control mechanisms are required to ensure normality. Therefore, I use the following "recursive differencing" estimator proposed by Shen and Klein (2017) to reduce the bias:

$$\widehat{P^k(\theta)} = \frac{N^{-1}\sum_j (Y_i^k - \delta_j(V_i))K_h(V_j - V_i)}{N^{-1}\sum_j K_h(V_j - V_i)} \tag{2.9}$$

where $K_h(V_j - V_i) \equiv \frac{1}{h^3}K(\frac{V_{Fj}-V_{Fi}}{h})K(\frac{V_{Bj}-V_{Bi}}{h})K(\frac{MFOI_j-MFOI_i}{h})$, h is a bandwidth parameter affecting the bias and variance in estimating P_i^k, and $K_h(x) \equiv \frac{1}{\sqrt{2\pi}}exp(-\frac{x^2}{2})$ is a Gaussian kernel function that downweights observations with V_j far away from V_i.

The exact formula for $\delta_j(V_i)$, termed "localization bias," is determined recursively to reduce the bias. The recursion depends on both the number of dimensions as well as the boundedness of the data. For single-index models, δ_j is zero, and the above estimator reduces to the regular Nadaraya-Watson estimator for conditional expectation (Stage 0). In our model with three indices, I need one additional stage of the recursion to reduce the order of bias to $O(h^4)$. Formal guidance on how to use this recursive differencing estimator to an appropriate stage is given in Shen and Klein (2017).

Step 2: Marginal Effect

After obtaining an estimator for θ and the estimated index $\widehat{V_{Fi}} \equiv F_1 + F'\widehat{\theta}^F, \widehat{V_{Bi}} \equiv B_1 + B'\widehat{\theta}^B$, a second stage "plug-in" estimator for QME_q^K is

$$\widehat{QME}(\mathcal{Z}_q; K, \hat{\theta}) \equiv \frac{\sum_{i=1}^N \hat{t}_{qi}\widehat{ME_i}(F_i, B_i, m_b; \Delta, K, \hat{\theta})}{\sum_{i=1}^N \hat{t}_{qi}} \tag{2.10}$$

where the quantile trimming function $\hat{t}_{qi} = \{MFOI_i \in \mathcal{Z}_q\}$ ensures that the average is taken over observations with $MFOI$ in the quantile of interest $\mathcal{Z}_q$. The unit-level marginal effect is estimated by the difference of predicted probabilities:

$$\widehat{ME_i}(F_i, B_i, m_b; K, \hat{\theta}) = \sum_{k=1}^{K-1} [\widehat{P}_k(\widehat{V_{Fi}}, \widehat{V_{Bi}}, MFOI_i + \Delta; \hat{\theta}) - \widehat{P}_k(\widehat{V_{Fi}}, \widehat{V_{Bi}}, MFOI_i; \hat{\theta})]$$

2.4.4 Inference

I also compute the large sample distribution of both $\hat{\theta}$ and $\widehat{QME(\mathcal{Z}; \mathcal{K}, \hat{\theta})}$. To preserve space, I briefly note a technical contribution—termed the *U-statistics equivalence*—which plays a key role in deriving the asymptotic distribution of $\hat{\theta}$.

From standard results, the asymptotic distribution of $\sqrt{N}(\hat{\theta} - \theta_0)$ depends on $\hat{H}(\theta^+)\sqrt{N}\hat{G}(\theta_0)$, where $\hat{H}(\theta^+)$ is the estimated Hession evaluated at some intermediate point $\theta^+ \in (\theta_0, \hat{\theta})$. In a large class of semiparametric index models, including the model given here, the gradient has the form:

$$\sqrt{N}\hat{G}(\theta_0) = \underbrace{N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{L}\tau_i[Y_i^k - E_i^k(\theta_0)]\nabla_\theta E_i^k(\theta)|_{\theta=\theta_0}\alpha_i}_{\text{A}} \quad (2.11)$$
$$+ \underbrace{N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{L}\tau_i[E_i^k(\theta_0) - \widehat{E_i^k(\theta_0)}]\nabla_\theta E_i^k(\theta)|_{\theta=\theta_0}\alpha_i}_{\text{B}} + o_p(1)$$

where E_i^k is the conditional expectation $E[Y_i^k|X_i]$ under the index assumption, whereas $\widehat{E_i^k(\theta_0)}$ is an estimation of that assumption. In the case of the ordered model, $E_i^k(\theta_0)$ is the conditional probability given in (2.9) and $\alpha_i = 1/E_i^k(\theta_0)$. This class also includes the quasi-maximum-likelihood estimators for semiparametric binary response (Klein and Spady, 1993) with $\alpha_i = 1/E_i(\theta_0)[1 - E_i(\theta_0)]$ and $k = 1$. The multiple-index semiparametric least-squares estimators (see Ichimura and Lee (1991) and Ichimura (1993))are also included, in which $k = 1$ (no categorical-specific conditional expectation), $\alpha_i = 1$.

Referring to the gradient representation given above, component A has no estimated quantities and can be handled by the standard central limited theorem. Shen and Klein (2017) asserted that in semiparametric index models with regular Gaussian kernels, B can be written as higher order degenerate U-statistics so the bias will vanish asymptotically. While this assertion is true for the single-index model, to the best of our knowledge there are no formal theorems proving $B = o_p(1)$ in higher dimensional cases. In Theorem 1 of the Appendix—what I refer to as the U-statistics equivalence result—I show that B is asymptotically equivalent to a degenerate U-statistics that is $o_p(1)$. This result can be applied to a large class of semiparametric models with arbitrary dimensions[8]. In Theorems 2 and 3, I derive the large sample distribution of $\hat{\theta}$ and $\widehat{QME}(\mathcal{Z}_q; K, \hat{\theta})$.

[8]To make the presentation cohesive, I only give proof in the context of an ordered model with three indices. When the dimension increases, I apply the recursive differencing multiple times according to the rules given in Shen and Klein (2017) to reduce the bias to a certain order.

In Theorem 1 below—what I refer as the U-statistics equivalence result—I show that B, the second component in the gradient, is asymptotically equivalent to a degenerate U-statistics that is $o_p(1)$. This result, as noted above, applies to a large class of semiparametric models with arbitrary dimensions[9]. Based on this important result, in Theorem 2,3, I derive the large sample distribution of $\hat{\theta}$ and $\widehat{QME}(\mathcal{Z}_q; K, \hat{\theta})$.

Theorem 2.4.1 (**U-Statistics Equivalence**). *With the window size $1/12 < r < 1/10$ for the case of three indices and the gradient representation given in (2.11), set the iteration of recursion equals 1, for a class of estimators defined in Section 4.4, it can be shown that with $\hat{g}(v,\theta) \equiv N^{-1}\sum_j K_h(V_j - v)$,*

$$\hat{g}(v,\theta)B = o_p(1)$$

where B is the second component in the gradient representation given in (2.11).

Theorem 2.4.2 (**Normality of θ**). *For the 3-index semiparametric ordered model discussed in the main text, with the window size $h = std(v)N^{-r}$, $1/12 < r < 1/10$ and Q_2 the likelihood function defined in (2.8), where the trimming function is based on the estimated index,*

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, H_0^{-1}\Sigma H_0^{-1})$$

where $H_0 \equiv E[\nabla_{\theta'\theta} Q_2(\theta_0)]$ $\Sigma = E\{\sqrt{N}\sum_{g=1}^{N} G_g G_g' \sqrt{N}\}$, $G_g = \nabla_\theta \sum_{i\in g} g_i(Y_i|\theta_0)$ and $g_i(Y_i|\theta_0) \equiv \sum_{k=1}^{L} Y_i^k Ln(P_i^k(\theta_0))$.

Theorem 2.4.3 (**Normality for Quantile Marginal effects**). *Under A.1-A.5, we have*

$$\sqrt{N}(QME_q^K - \widehat{QME_q^K}) \sim N(0, \sum_{k=1}^{K-1} E[\psi_k' \psi_k])$$

[9]To make the presentation cohesive, I only give the poof in the context of an ordered model with three indices. When the dimension grows higher, one should apply the recursive differencing multiple times according to the rules given in Shen and Klein (2017) to reduce the bias to a certain order.

where $\psi_k \equiv \psi_{1j}^k + \psi_{2j}^k + \psi_{3j}^k + \psi_{4j}^k$ *with*

$$\begin{aligned}
\psi_{1j}^k &\equiv \frac{E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z+\delta;\theta_0)] - E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z;\theta_0)]}{E[t_{qj}]} H_0^{-1} G(\theta_0) \\
\psi_{2j}^k &\equiv \{\nabla_q E[t_{qj} m_j(\theta_0)] - \nabla_q E[t_{qj} QME_q^k]\} \frac{\mathcal{B}_j}{E[t_{qj}]} \\
\psi_{3j}^k &\equiv \frac{E[t_{qj}|V_F, V_B, Z+\delta]\epsilon_j^\delta - E[t_{qj}|V_F, V_B, Z]\epsilon_j}{E[t_{qj}]} \\
\psi_{4j}^k &\equiv \frac{t_{qj} m_j(\theta_0) - E[t_{qj}] QME_q^k}{E[t_{qj}]} - \frac{t_{qj} - E[t_{qj}]}{E[t_{qj}]} QME_q^k
\end{aligned} \tag{2.12}$$

2.5 Results

In this application, I estimate the heterogeneous impact of MFOI, the aforementioned shared-ownership index, on credit ratings in the described semiparametric model. Previous estimates reported in the literature are typically constrained to a single number by the functional form of the underlying regression model. For example, Kedia *et al.* (2017) found that the ratings assigned by Moody's are, on average, 0.213 notches better than ratings by S&P's for firms related to Moody's two major shareholders. This number can be understood as the "average treatment effect" of a 0-1 variable capturing whether a bond issuer has a relationship with Moody's shareholders. However, if the benefit of developing a rapport with Moody's shareholders is actually heterogeneous, such an estimate is not informative on the effect that varies across relevant subpopulations and may not even be consistent for the overall population mean (Abrevaya *et al.*, 2015). Using a flexible econometric approach, the application explores the heterogeneity of the shared-ownership effect across subpopulations defined by rating categories and/or possible values of issuer characteristics.

For comparative purposes, and to highlight the importance of employing a more flexible framework, I estimate both the proposed semiparametric model and the ordered-probit model described in the previous section. I compare both the estimated index coefficients as well as the marginal effects in quantiles. Lastly, I compare the two approaches in terms of predicting credit ratings.

2.5.1 Index parameter estimates and average marginal effects

Table 2.3: Index Parameter and Average Marginal Effects

	Index Parameters	*Marginal Effects (percentage point)* AA	A	Baa	Ba	B	C	Average
Semiparametric								
ASSET	1.00	0.11	1.71	5.89	6.13	4.31	0.64	4.42***
STABILITY	-2.71***	-0.23	-4.46	-8.89	-10.80	-8.81	-1.25	-5.74***
LEVERAGE	-4.25***	-0.01	-0.97	-2.08	-2.78	-2.57	-0.48	-1.49***
PROFIT	24.21***	0.49	4.81	17.88	15.52	10.06	1.31	9.91***
AMT	0.41***	0.05	0.07	-0.09	-1.96	-1.23	0.08	-0.49***
SENIORITY	1.00	0.81	0.62	3.36	8.52	4.52	-0.22	3.10***
MFOI		0.51	9.78	9.00	2.12	2.27	0.14	5.86***
Ordered-Probit								
(with Year and Industry Fixed Effects)								
ASSET	1.00	0.98	5.69	9.12	10.05	7.47	1.69	6.77***
STABILITY	-0.51***	-0.67	-4.35	-5.91	-5.83	-4.45	-1.07	-3.71***
LEVERAGE	-5.14***	-0.41	-2.65	-3.59	-3.54	-2.71	-0.65	-2.26***
PROFIT	14.92***	1.83	11.87	16.11	15.90	12.14	2.91	10.13***
AMT	-0.09	-0.02	-0.11	-0.15	-0.18	-0.13	-0.03	-0.06
SENIORITY	1.00	1.08	6.45	8.75	8.63	6.59	1.58	5.48***
MFOI	-71.13***	0.51	3.02	4.84	5.33	3.96	0.89	3.59***

Note: *** represents statistical significance at the 1% level
- In column 1, I report the index parameter estimates. In the semiparametric model, the parameters of asset and seniority are normalized to one. Since MFOI enters the model nonparametrically by itself, there are no parameter estimates for MFOI.
- In columns 2 to 7, I report the average marginal effect of covariates from the semiparametric model (top panel) and ordered probit (lower panel) for each rating category. The marginal effects are computed by increasing the *asset* and *issuing amount (AMT)* by 1- In the last column, the average marginal effects are calculated by taking a weighted average of category-specific marginal effects where the weights are the percentage of a rating category in the entire sample.

In Table 2.3, I report the estimates of index parameter vectors θ and average marginal effects for the parametric (ordered-probit) and semi-parametric models. Statistical significance of the semiparametric model is obtained based on the asymptotic covariance matrix derived in this paper (Theorem 2 in the Appendix)[10].

[10]The standard errors in the ordered probit model are computed from the White (1982) formula.

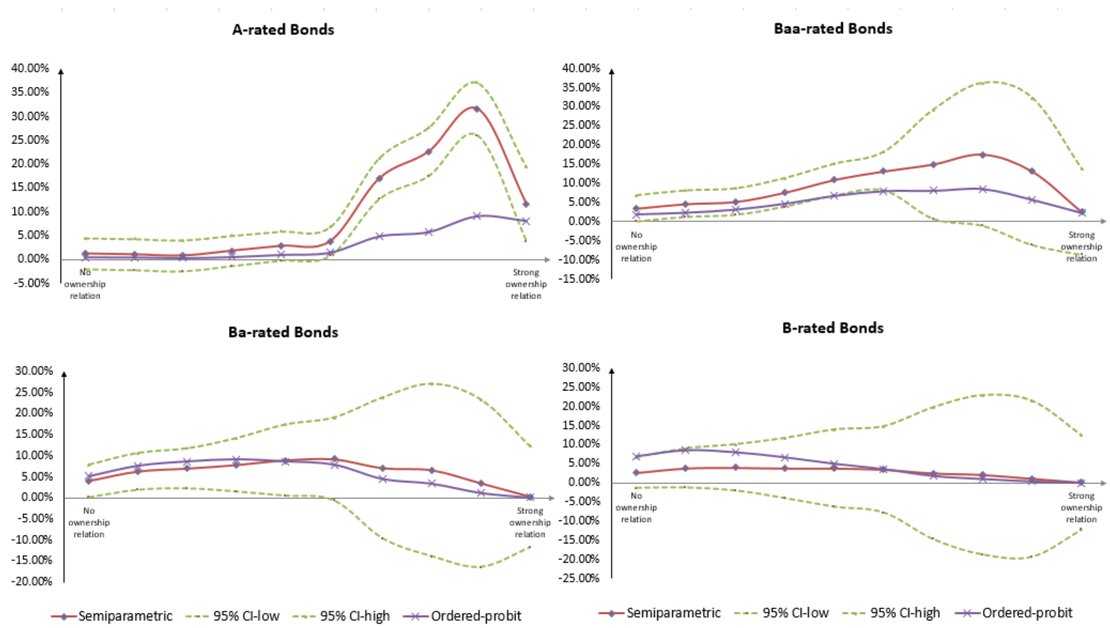
The parameters of SENIORITY and ASSET are normalized to one; both variables belong to the model considering the bond rating literature. The signs and statistical significance of index parameters are consistent across models, except for ASSET.

Next, I compare the average marginal effects in the two models because, in ordered models, there is typically no natural economic interpretation for the index parameters — I know nothing beyond whether a variable belongs to the model. I begin by discussing the marginal effects of MFOI, our main variable of interest. Overall, the semiparametric model yields a much larger effect than the ordered-probit (5.86% vs 3.59%). However, for A and Baa-rated bonds, the estimated effects from semiparametric models triple that of the ordered-probit model: when MFOI increases by one standard deviation[11], A-rated bonds are 9.78% more likely to be rated into a higher category from the semiparametric model, whereas the estimated effect from ordered-probit is only 3%. The more dispersed effect captured by the semiparametric model highlights the potential value of employing a more flexible approach.

The estimated impacts of firm and bond characteristics are consistent across models. In terms of economic magnitude, the most significant impact on ratings comes from PROFIT, which is the ratio of profits to total assets realized from business operations: a 10% increase in PROFIT increases the likelihood of obtaining a higher rating by approximately 10 percentage points. A bond issuer's asset level also has a significant effect on ratings. When ASSETs go up by 1%, bonds are 4.4% and 6.8% more likely to be rated in a higher category in semiparametric and ordered-probit models, respectively. When the bond issuer has a higher leverage ratio or asset volatility, the ratings on its bonds tend to be lower. Subordination status (SENIORITY) has a highly significant effect on rating in both specifications: declaring seniority causes a bond 3.1% (5.5%) more likely to be rated higher in the semiparametric model (ordered-probit).

[11] In our sampling period, $SD_{MFOI} = 0.004$. In terms of economic magnitude, this implies that a bond issuer is related with another shareholder of Moody's who owns 10% of Moody's stock, and the bond issuer accounts for 4% of the shareholder's portfolio.

Figure 2.1: Quantile Marginal Effects of MFOI



Note:

- The horizontal axis in each sub-panel indicates the decile level of MFOI from the lowest (issuers with no shareholding relation with Moody's) to the highest (issuers with a strong shareholding relation with Moody's). The red solid line corresponds to the (average) marginal effects of MFOI in the corresponding decile of interest with the two dashed green lines bounding the 95% confidence interval.

-The marginal effects are computed by increasing MFOI by one standard deviation. For example, in the top panel, the right end-point in the red line should be interpreted in the following way: when the CRA-issuer relation strengthens by one standard deviation as measured by MFOI, the probability that an A-grade bond will be rated higher will increase by 9.5%.

- The confidence intervals are computed using the asymptotic theory derived in this paper.

2.5.2 Quantile effects of MFOI

In Figure 4.1, I plot the estimated quantile marginal effects of MFOI—the average marginal effect of MFOI for observations with MFOI in a particular quantile— from the semiparametric model (red solid line) and ordered-probit model (purple solid line). Note that one conclusion from Table 2.3 is that the effect of MFOI varies significantly by categories. Here, I examine the heterogeneity of marginal effect along the quantile of MFOI. There are two main findings.

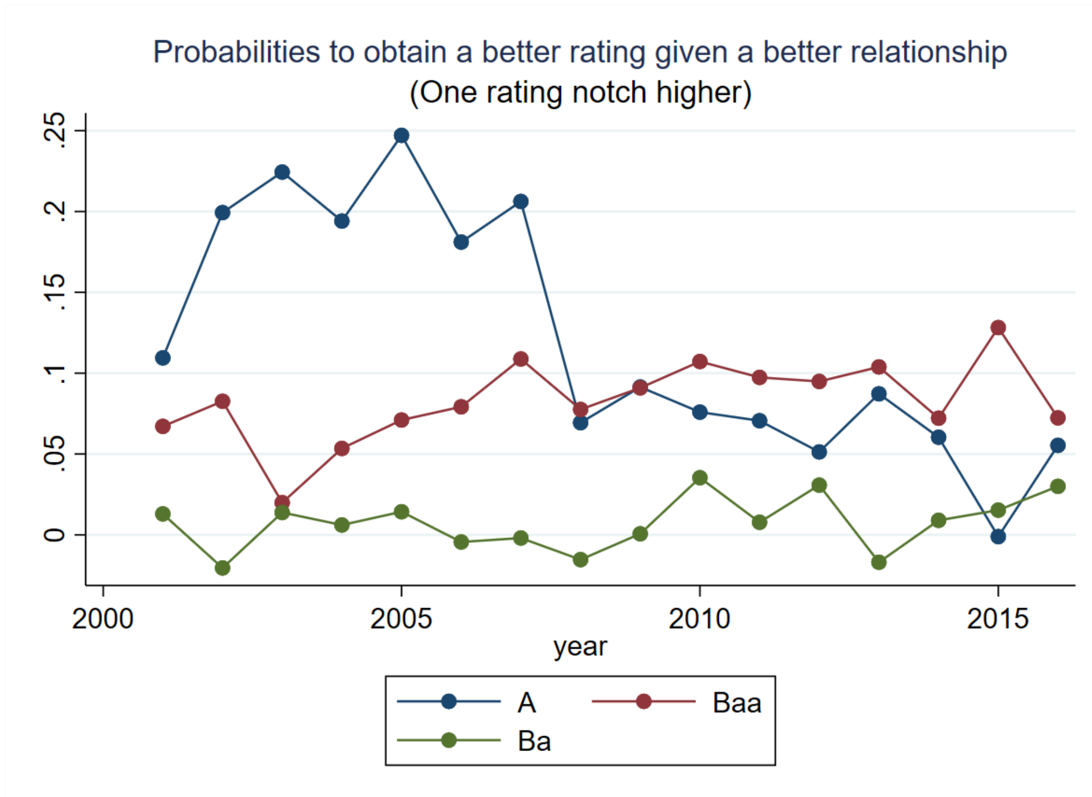
First, I find that rating inflation is unlikely to occur on bond issuers associated with small MFOI. Using the A-grade bonds as an example, the estimated marginal effect of MFOI is approximately 30% for firms that have strong connectedness with Moody (those with MFOI in the ninth decile) implying that roughly one third of A-grade bonds issued by those firms might receive favorable treatment. A strengthening CRA-issuer relation also has a significant positive impact for Baa-grade bonds (as depicted in the top-right panel of Figure 4.1); however, the economic magnitude is much smaller (from 30% to 15%). In contrast, marginal effects are not statistically significant for issuers associated with low-decile MFOIs. Second, the inflation rating is not pronounced for bonds below investment grade regardless of the issuer's shareholding relation with Moody's. As depicted in the lower two panels of Figure 4.1, the probability that a bond is rated into a higher category is at most 6% for Ba-rated bonds and 4% for B-rated bonds; both effects are not statistically significant.

Note that the *average* magnitude of rating bias identified approximate the magnitude found in Kedia *et al.* (2017)[12]. Additionally, by estimating the heterogeneous marginal effect, our model highlights the *distributional pattern* of rating bias. Qualitatively, our main conclusion from our empirical exercise is that the degree of Moody's rating bias varies significantly for both rating categories as well as the bond issuer's affiliation with Moody's shareholders. As can be seen from Figure 2.2, the favorable treatment on A-bonds does seem to decline over time.

Capturing such heterogeneity is difficult in a parametric setting because of the constrained functional form. By comparing the ordered-probit estimates (purple line) and semiparametric estimates (red lines), the ordered-probit estimates are more "homogeneous": they vary between zero and 10 % (whereas the semiparametric estimates can be as large as 30%) and have identical patterns across different rating categories. One possible explanation could be that the ordered-probit model assumes that the rating probabilities for all categories are driven by the same normally

[12] Kedia *et al.* (2017) found that Moody's ratings are a 0.213 category better than S&P's, on average, using a finer rating scale (A1,A2,A3...). This number translates to a 7.1% average marginal effect in our scale assuming that the rating probability is linear. Recall that in the semiparametric model, I find that the average marginal effect of MFOI is 5.86%.

Figure 2.2: Time series variation of marginal effects from 2001-2016



Note: The y-axis is the marginal effect, which is the probability that the bond will be rated one notch higher given a standard deviation increase in MFOI. From top to bottom, the three hairlines describe the time series change in marginal effects for A, Baa, and Ba bonds.

distributed random variable. In contrast, the semiparametric model allows the rating probability function in (3.2) to be category-specific.

2.5.3 Prediction Performance

In this section, I compare our semiparametric model and various parametric models in the literature as for in-sample fitness. Using the dataset on Moody's initial ratings from 2001-2007, I first estimate the semiparametric model proposed in the previous section:

$$\begin{aligned} Pr(Y_i = k|X) &= E(R_i^k|V_F, V_B, MFOI) = P_k(V_F, V_B, MFOI) \\ V_F &= ASSET + \theta_1^F STABILITY + \theta_2^F LEVERAGE + \theta_3^F PROFIT \\ V_B &= AMT + \theta_1^B SENIORITY \end{aligned} \tag{2.13}$$

with the following ordinal information on ratings:

$$Y_i = \begin{cases} 1 & \text{if the bond is rated as Aaa} \\ 2 & \text{if the bond is rated as Aa} \\ ... & \\ 7 & \text{if the bond is rated as Caa} \end{cases}$$

This specification includes our theoretical prediction $y^* = X\beta_0 + H(Z_i, U_i)$ as a special case. Despite the generality with which this framework accounts for the influence of explanatory variables, the complexity of the estimation procedure raises the question of whether these features can be satisfactorily addressed by a simpler model.

In the empirical literature of credit ratings, the most prevailing approach is linear probability models (LPM, see Jiang *et al.* (2012); Campbell and Taksler (2003); Kedia *et al.* (2017) among other works), which assumes the observed ratings Y_i is a linear function on predictor variables X and error term U, e.g., $Y_i = X_i\beta_0 + Z_i\pi_0 + U_i$. This approach ignores the discreteness in ratings and implicitly assumes, for example, Aa-grade bonds ($Y_i = 2$) are twice as likely to default as Aaa-grade bonds ($Y_i = 1$). Such an assumption is logically problematic, as rating scales convey only ordinal rather than quantitative relationship. Moreover, Hausman *et al.* (1992) argues that the linear probability model always deliver conditional distributions of dependent variable that is unimodal and have little weight in the tails. Therefore, in contexts where the focal interest is the conditional probabilities, as it is in our case, researchers switch to more advanced discrete-choice models.

The second frequently employed approach is ordered probit/logit models (Kaplan and Urwitz, 1979; Blume *et al.*, 1998; West, 1970). This class of models assume the default risk

$$y^* = X\beta_0 + Z_i\pi_0 + U \tag{2.14}$$

and the same aforementioned ordinal information. Hausman and his coauthors regarded ordered-probit as "a suitably extended version of LPMs" when the dependent variable is naturally discrete. However, ordered-probit model deliver consistent estimates only when the functional form on y^* and the distributional assumption on U are correctly specified. As derived in the behavioral framework, the CRA's estimates of default risk y^* does not have the "threshold-crossing" property[13] required by ordered-probit. Therefore, the estimates of β_0 as well as conditional probabilities $Pr(Y_i = K|X)$ in ordered-probit models are subject to misspecification bias.

In ordered-response models, usually fitness is measured by the percentage of being correctly predicted. The predicted rating is the category with highest conditional probability Pr[$Y_i^k = 1|X$].

---

[13]This is referred as a threshold crossing model because $Pr(Y_i \leqslant K|X) = Pr(X\beta_0 + U < c_k|X) = Pr(U < c_k - X\beta_0|X)$. That is, a bond will be rated at least category k higher when the error term U "crosses", or lower than, the respective threshold point $c_k - X\beta_0$

Semiparametric vs Ordered-Probit

The ordered-probit model specified above replaces (2.13) with the following parametric structure:

$$Pr(Y_i = k|X) = \begin{cases} \Phi(-ww_i) & \text{if } k = 1 \\ 1 - \Phi(c_{L-1} - ww_i) & \text{if } k = 7 \\ \Phi(c_k^* - ww_i) - \Phi(c_{k-1}^* - ww_i) & \text{otherwise} \end{cases}$$

where

$$ww_i = \alpha_0 X_i + \gamma_0 Z_i + c_1^*, \qquad \alpha_0 = \frac{\beta_0}{\sigma} \qquad \gamma_0 = \frac{\pi_0}{\sigma}$$

The Φ function is the cumulative distribution function for U, and the α's and γ can be consistently estimated through the standard maximum likelihood estimation.

After estimating the semiparametric model and the benchmark ordered probit model, we report the fitness for the two in Table 2.4, which is the traditional format of displaying the results of bond-rating predictions. The upper table corresponds to the semiparametric model with three indices and the lower table corresponds to the linear parametric probit model. Collectively the semiparametric model correctly predicts 68 % of bonds, which is 10 % higher than the standard parametric model with the same explanatory variables. In addition, the semiparametric model performs a better predictive power in all the rating categories, especially the Aaa, A and Ba grades.

Semiparametric vs Other Models in the Literature

I further compare the semiparametric model with previous models in the literature of bond ratings and reports the results in Table 2.5. It is important to note that, for each previous work, the statistics on the percentage of correctly predicted is directly imported from the corresponding paper. Therefore, we view the comparison as suggestive in the sense that the dataset and the explanatory variables being used are different. In terms of percentage of correctly predicted, the semiparametric

Table 2.4: Predictions of New-Issues: 2001-2007

	Predicted Ratings(Semiparametric)						
Actual ratings	Aaa	Aa	A	Baa	Ba	B	Caa
Aaa	13	21	16	3	3	3	
Aa		753	146	17			
A		91	979	374	2		
Baa				1341	46	55	
Ba			16	439	213	116	
B			3	190	101	661	3
Caa				28	6	62	44

	Predicted Ratings(Linear ordered probit)						
Actual ratings	Aaa	Aa	A	Baa	Ba	B	Caa
Aaa	0	34	14	9		2	
Aa		732	165	19			
A		339	682	421	4	2	
Baa		4	191	1293	51	69	
Ba			7	508	85	184	
B			4	224	83	613	34
Caa			1	32	3	90	14

Note: The upper table is the prediction result for the 3-index model and the lower table is the prediction result for the ordered-probit model with specification below
- 3-index model: 4009/5913 = 67.72 % correct
- Probit model: 3419/5913 = 57.82 % correct

Table 2.5: Comparison with other predictive models in the bond rating literature

	% of correct prediction in category							
Study	Aaa	Aa	A	Baa	Ba	B	Caa	%
West(1970)	0.00	0.65	0.76	0.45	0.57	0.67		0.6234
Horrigan(1966)	1.00	1.00	0.71	0.53	0.64	0.4		0.5857
BLM*(1998)	0.26	0.36	0.74	0.54				0.5721
PM**(1975)		0.71	0.83	0.48	0.89	0.74		0.7538
KU(1979)	1.00	0.22	0.92	0.47	0.00	0.00		0.6875
3-index	0.22	0.82	0.68	0.83	0.27	0.69	0.31	0.6772

* - In Blume *et al.* (1998), the authors estimate only the investmnet grade bonds using S%P's rating

** - In Pinches and Mingo (1975), the authors use Multiple Discriminant Analysis (MDA) instead of regular regression model outperforms West (1970), Horrigan (1966) and Blume *et al.* (1998). The semiparametric model has roughly the same predictive power with Kaplan and Urwitz (1979). However, KU made poor prediction on Aa Ba and B bonds, while the semiparametric model shows more robust predictive power across all rating categories.

2.6 Conclusion

This paper contributes to the literature by evaluating rating quality using a semiparametric ordered model. Compared to extant parametric models, the semiparametric model proposed in this paper allows for a richer set of interactions among covariates. I study to what extent Moody's ratings are affected by the economic interests of its shareholders, which is pertinent for the regulation of credit rating agencies.

In summary, I conclude that a strong connection with Moody's shareholders could increase the probability of receiving a higher rating by as much 31 %, or, on average, one out of three bonds issued by firms with a Moody connection received favorable treatment. This effect is twice that of comparable parametric models. In addition, I found that high-yield bonds issued by any firms, regardless their ownership interaction with Moody's, are unlikely to be treated favorably, which seems credible because overrating a subprime bond would incur a greater expected reputation loss than overrating a safe bond.

Chapter 3

Ordered Response Models with Unobserved Correlated Thresholds

3.1 Introduction

Alleviating information asymmetry is central to enhancing the efficiency of financial markets. As one of the most important information intermediaries, Credit Rating Agencies (CRAs) have been dedicated to the production and dissemination of credit ratings to market participants since 1920s. Credit ratings, at its core, are devised to benchmark the default risk of financial instruments. The proper functioning of CRAs reduces information asymmetry between borrowers and lenders, which is crucial to the health of financial markets.

In the aftermath of the recent financial crisis, the reliability of CRAs' rating methodology, however, has been scrutinized. In fact, apart from estimating default risk with financial variables, CRAs have the discretion to adjust rating outcomes subject to their own understanding of qualitative factors. This step is termed "soft adjustment" by Kraft (2014) because adjustments are made based on hidden and subjective factors such as the manager's ability. Given their exposure to a variety of conflicts of interest, it is unclear whether CRAs utilize the soft adjustment to reflect material nonpublic information or simply distort ratings in a subtle way.[1] As a regulatory response,

[1] There has been an extensive literature focusing on the conflicts of interest in credit ratings (Kedia *et al.*, 2017; Becker and Milbourn, 2011; Jiang *et al.*, 2012). Several channels, through which rating inflation can occur, are examined and documented by past studies. For example, public firms are operated under intensive pressure to grow and increase profits (Bogle, 2005), which motivates CRAs to report inflated rating in order to retain repeated customers for rating fees under the current issuer-pay

one chapter of the Dodd-Frank Wall Street Reform and Consumer Protection Act (henceforce “the Dodd-Frank”) is dedicated to rating agency reform through enhancing information transparency and strengthening supervision.

Set against this background, this paper develops an empirical framework to investigate the information production process of CRAs. A variety of equilibrium-based economic models have been developed to help better understand the mechanisms, as well as potential problems, of the information production process of CRAs[2]. However, on the substantive end, extant bond rating models are largely in reduced-forms (Kaplan and Urwitz, 1979; Blume *et al.*, 1998, etc), and therefore fall short to genuinely capture CRA’s strategic behaviors predicted by the theories. Implied by a structural framework, our study contributes to this literature by providing an empirical valuation of CRA’s information. We focus on (i) quantifying and estimating the impact of material non-public information on credit rating assignments, and (ii) examining the time variation of such effects before and after the passage of the Dodd Frank Act.

We examine the rating process for corporate bond ratings in an innovative econometric framework. To be specific, we model the aforementioned soft adjustment as *firm-specific* thresholds in an ordered-response model, in which a bond will be assigned to a rating category if its latent default index is between the corresponding thresholds. The literature on bond ratings often requires the thresholds to be fixed parameters, which leaves the differences in *ex post* ratings completely attributable to the idiosyncratic rating errors for observationally identical bonds. In contrast, firm-specific thresholds echoe the aforementioned “soft” adjustment, because firms with identical fundamentals can wind up with different ratings due to adjustment

---

business model (Cornaggia and Cornaggia, 2013; Jiang *et al.*, 2012).

[2]The early papers of Allen (1990), Ramakrishnan and Thakor (1984) and Millon and Thakor (1985) provide the theoretical foundations for thinking about rating agencies as diversified information producers and sellers. Relatedly, in a cheap-talk framework developed by Goel and Thakor (2015), the authors model a rating agency’s objective in setting ratings is as to balance the divergent goals of the issuing firm and the investors purchasing the issuing securities.

on the thresholds. Of interests are both the latent index coefficients, which capture the relative importance of different risk predictors, and thresholds parameters, which convey information on the degree of soft adjustment.

We note that in this context, soft adjustment may stem from two sources: unobserved heterogeneity and conflicts of interest. Both of them may cause mechanical correlation between the firm-specific thresholds and regressors, leading to the problem of endogeneity. Our paper contributes to the literature of location estimators of ordered response models by allowing endogenous regressors and correlated thresholds (Manski, 1985; Horowitz, 1992; Lewbel, 1997; Klein and Sherman, 2002, etc). In particular, we demonstrate that under an additive separability condition, a bond issuer's "connectedness" with CRA's can be exploited as an efficient control for endogenity. Our identification strategy on index parameters follows the control function approach (Blundell and Powell, 2004; Florens *et al.*, 2008; Imbens, 2007; Imbens and Newey, 2009, etc). To quantify the soft adjustment, we focus on the average threshold conditional on the aforementioned measure of "connectedness". The identification strategy exploits a special property of the rating probability functions termed *conditional shift restrictions*, which is a generalization of Klein and Sherman (2002) to models with endogenous predictors. The regression coefficients are estimated by semiparametric pseudo maximum likelihood, combined with a grid-search algorithm to estimate conditional mean thresholds. Finally, the extracted hidden adjustments are used to evaluate policy changes.

We estimate the proposed model with 11,134 initial bond ratings issued by Moody's from 2000 to 2016, with the sample split by the enactment of the Dodd-Frank Act in 2010. Moody's became a public firm in 2000, with over 300 shareholders every quarter since then. One conflict that resurfaced recently concerns the ownership structure of CRAs: publicly traded CRAs may bias ratings towards issuers that are invested by their own shareholders. This issue was first noted by Kedia *et al.* (2014) and further examined by Kedia *et al.* (2017). However, none of the above papers explicitly considers the relationship between soft adjustment and rating bias. In contrast, we assess the change of soft adjustment after the passage of Dodd-Frank and examine the relationship between shareholding structures and soft adjustments, accounting for the fact that bond issuers may choose bond characteristics (such as issue amount, subordination status) endogenously based on perceived favorable treatment and private soft information.

The estimation results suggest several noticeable changes in terms of Moody's rating methodology before and after the Dodd-Frank. First, issuing amount and profitability have gained more weights in CRAs' discretion of hard information, implying that Moody's has become more stringent towards low-profit and high-debt issuers. Second, there is a significant drop in the dispersion of soft adjustment for all categories after the reform. This provides evidences in support of the effectiveness of the Dodd-Frank Act in the credit rating industry, at least in terms of the effort to reduce information opacity. Third, by examining the pattern of threshold parameters, we find that it has become more difficult for a bond to be rated as investment grade bonds on average, and that Moody's has been more stringent towards issuers that are related with Moody's large shareholders. These findings suggest Moody's has become more conservative and sensitive to conflicts of interest after the Dodd-Frank.

The rest of this paper is structured as follows. In Section 3.2, we consider the stylized rating process for corporate bonds with soft adjustment and discusses identifying strategies of the hidden soft adjustment along with the assumptions needed. In Section 3.3, we propose a two-stage semiparametric index and location estimator to quantify soft information. Section 3.4 provides the background of the U.S. credit rating industry and presents our data. Empirical results are in Section 3.5. Finally, Section 3.6 concludes this chapter.

3.2 Bond Rating Models and Soft Adjustment

Uncovering the "blackbox" of the rating methodology used by CRAs has always been a pursuit of financial regulators, academic researchers and business practitioners. Let $Y_i \in \{0, 1, \cdots, J-1\}$ be a discrete ordinal credit rating for bond i. To fix idea, we characterize the rating process as:

$$Y_i = \sum_{j=0}^{J-1} j \times \mathbf{1}\{T_{j-1,i} < V_{0i} \leqslant T_{ji}\}, \quad j \in \{0, 1, \cdots, J-1\} \tag{3.1}$$

wherein a bond's latent "default risk index" is driven by observed firm and bond characteristics $X_i \equiv (X_i'^F, X_i'^B)'$ as a single-index, i.e. $V_{0i} \equiv X_i\beta_0$; $\boldsymbol{T_i} = (T_{0i}, \cdots, T_{J-1,i})$ is a vector of *bond-specific* thresholds that partitions the risk index into different rating categories. $\mathbf{1}\{\cdot\}$ represents an indicator function.

Importantly, allowing the thresholds to be bond-specific echoes the idea of "soft adjustment" (Kraft, 2014).[3] That is, by perturbing $\boldsymbol{T_i}$, the CRA is able to assign different ratings to bonds that have identical financial characteristics, reflecting certain qualitative adjustments. Extant credit rating models often restrict thresholds to be constant plus a pure random error; this assumption, however, rules out soft adjustment, leaving the differences in *ex post* ratings completely attributed to idiosyncratic errors for otherwise identical bonds.

Of interests are the estimation of the index coefficients β_0 and recovering patterns of the bond-specific threshold $\boldsymbol{T}_i$. However, when $\boldsymbol{T}_i$ is interpreted as soft adjustments, the correlation between $\boldsymbol{T_i}$ and regressors requires new identification strategies to be developed. We believe such correlation may arise from two sources: *unobserved heterogeneity* and the *conflicts of interest*. The unobserved heterogeneity reflects

[3] According to Kraft (2014) and Petersen (2004), credit ratings involve both "hard" and "soft" information. The CRA first constructs a default risk index using publicly available quantitative information from issuers' financial statements. This generates the relatively objective "hard" information. Next, the CRA conducts a subjective "soft" adjustment based on other qualitative factors and then finally release a categorical rating to the public given its internal criteria.

qualitative information about the underlying bond that is otherwise not reflected in X_i and this information is available to both the bond-issuing firm and the CRA but not the researcher.[4] Therefore, the bond-specific threshold should correlate with X_i due to the mechanical correlation between a bond's observed and unobserved characteristics. Therefore, some bond and firm characteristics, X_i, are likely to be endogenously selected in the presence of unobserved heterogeneity.

Other than reflecting CRA's private information, the soft adjustment may also be due to conflicts of interest. In our context, bond issuers and the CRA are connected through common shareholders. CRA may cater to the interest of its shareholders by assigning inflated ratings. Endogeneity arises in this context because the issuer can contemplate on the choices of characteristics by forming an expectation of the soft adjustment given their own private relationship with the CRA. For instance, a better connection with the CRA could induce riskier and more audacious issuance due to the expectation of receiving an upward rating adjustment.[5] Estimates will be generally inconsistent if the induced endogeneity is not appropriately taken into account.

Before discussing our identification strategy in detail, we make several observations. First, the soft adjustment, if identified, might serve the purpose of uncovering the "blackbox" of rating models, increasing the transparency of rating methodology and improving the predictive power of current models. Second, this econometric framework can be utilized to investigate whether the rating process is impaired by conflicts of interest. By estimating the systematic pattern of T_{ji}, one can empirically test whether the CRA has consistent rating criteria towards all bond issuers.

[4]For example, the unobserved firm characteristics could encompass the managerial efficiency, organizational productivity or other qualitative financial risk related information beyond spreadsheet. Moreover, this firm heterogeneity is very likely to be time-varying, so adding company fixed effects might not fully solve the problem.

[5]To illustrate, consider the following example: one bond characteristics that affects credit ratings is *issuing size*, namely how large the debt issuance is. If the bond issuer has prior information about $\boldsymbol{T_i}$, say knowing that the CRA will be more lenient, it may strategically choose to issue more debt since a better credit rating could lower the borrowing interest rate. See online appendices for the example.

3.2.1 Identification via Shareholding Structure

It should also be noted that our model is semiparametric in the sense that no distributional assumption is made about the thresholds T_{ji}. And we consider this necessary especially given the complex structure and meanings of the soft adjustment across bonds and categories. For this reason, the index parameter vector β_0 can be identified only up to location and scale. Suppose $\beta_0 \equiv (\beta_{10}, \beta_{20}, \cdots, \beta_{d0})' \in \mathbb{R}^{d-1}$ is the coefficient vector that is conformable to the $d-$dimensional bond and firm characteristics X_i. We let $\beta_{10} = 1$ for a continuous variable X_{1i}, e.g. the log of total asset in our empirical analysis, and denote the identifiable index by $V_{0i} \equiv X_{1i} + \widetilde{X}_i'\beta_0$. Under this normalization, β_0 becomes the relative contribution of each characteristic with respect to that of X_{1i}. A sufficient condition for identification is $\det(\widetilde{X}_i'\widetilde{X}_i) > 0$ with X_{1i} being a continuous variable and $\widetilde{X}_i \equiv (X_{2i}, \cdots, X_{di})$.

For identification, we rely on a control function approach to handle endogenous variables that are correlated with structural thresholds.[6] To be specific, let R_i be a vector capturing the whole connectedness between the CRA and a bond issuer through common shareholders, e.g. each common shareholder's identity, investment stake in the CRA and the bond-issuing firm i, etc. We impose an additive separable structure on thresholds.

A-I.1 Additive Separability. *For each j, $T_{ji} = \delta_j(R_i) + u_{ij}$, where $u_{ij} \perp (X_i, R_i)$ and u_{ij} is i.i.d. across i and j.*

A-I.1 basically conveys that each threshold can be decomposed into two additive terms, i.e. a category-specific component, $\delta_j(\cdot)$ and an orthogonal random shock u_{ij}. The functional form of $\delta_j(\cdot)$ needs to be flexibly specified. The former component reflects the heterogeneous soft adjustment or category effect, while the disturbance represents calculation errors of the CRA or pure noise. Under A-I.1, it implies that the soft

[6] The control function approach is frequently employed in nonseparable models with endogeneity (Blundell and Powell, 2004; Florens *et al.*, 2008; Imbens and Newey, 2009; Hoderlein and Sherman, 2015, etc).

adjustment is equivalent to the conditional mean of threshold, i.e. $E(T_{ji}|R_i) = \delta_j(R_i)$. It also implies the independence of X_i and the thresholds T_{ji} after conditioning on R_i.[7] To see this, note that

$$\Pr(Y_i \leqslant j|X_i, R_i) = \Pr(-u_i \leqslant \delta_j(R_i) - V_{0i}|X_i, R_i) = \Pr(Y_i \leqslant j|V_{0i}, R_i) \quad (3.2)$$

More importantly, A-I.1 also dictates that the shareholding structure R_i suffices to control for all unobservables that matter in both firms' issuing decisions and the CRA's ratings. Given the sources of the soft adjustment considered earlier, the *conflict-of-interest*, having stronger connections with the CRA may lead to favorable adjustment at each category, a finding in Kedia *et al.* (2017). When thresholds reveal a certain degree of *unobserved heterogeneity* of the firm itself, our compromising assumption is that common shareholders have as much private information as the CRA. If so, any material nonpublic information about the issuer firm is "materialized" in common shareholders' investment decisions, thus already embodied in the shareholding structure R_i. For example, institutional investors are more likely to invest in issuers with "better" unobserved soft quality. In empirical contexts, it is alway the case that both sources of adjustment are present. However, it is unnecessary to separately identify the two sources if our objective is only to consistently estimate the index parameters and back out the *ex post* individual soft adjustment.[8]

Our choice of shareholding structure is substantive admittedly. If there are other hidden factors which drive the rating decision but bear no relation with the investment of common shareholders, our control could be insufficient. Developing a fully structured issuing model that covers every possible pathway of information sharing between the CRA and common shareholders is a daunting task. Instead, we present a simple example in the Appendix to highlight the endogeneity problem and the role of

---

[7] This is also implied by a weaker condition—conditional independence, i.e. $X_i \perp \boldsymbol{T}_i|R_i$.

[8] One shortcoming of mixing together those errors is that counterfactuals in regards to the change of shareholding structure would be confounded and unclear.

R_i.

In order to encompasses the entire connectedness structure between bond issuer i and the CRA through all common shareholders, R_i can be very high-dimensional. To this end, we assume that the information contained in R_i can be sufficiently summarized by a "index" of three key aggregate variables in the empirical analysis such that

$$\Pr(Y_i \leqslant j | X_i, R_i) = \Pr(Y_i \leqslant j | X_i'\beta_0, R_i'\alpha_0) \tag{3.3}$$

Under this simplification, the space of finite-dimensional parameters also expands, e.g. $\theta_0 \equiv (\beta_0, \alpha_0) \in (\mathcal{B} \times \mathcal{A})$. Such models have been studied in **?** and Klein and Vella (2009). Identification of double-index parameters requires the existence of at least one continuous variable in each index and a sufficient condition precluding the composition of same variables in both indices, which our model has already satisfied.[9] To streamline the discussion, we defer the detailed description of these three variables until the empirical section. As our identification of the soft adjustment does not rely on the index structure of R_i, we therefore illustrate it in a general specification.

3.2.2 Soft Adjustment and Conditional Shift Restrictions

Turning to the soft adjustment $\delta_j(R_i)$, the key object of interest in this paper, we show that identification of this object can be achieved by exploiting a "special" property of the rating probabilities. This property is a generalization of the "shift restriction" proposed by Klein and Sherman (2002)[10]. Besides ours, there are other thresholds or location estimators that have been considered in the binary or ordered choice literature

[9] As in other semiparametric models, index parameters are identified up to location and scale. Specifically, let $R_{1i} + \widetilde{R}_i'\alpha_0$, where R_{1i} is continuous and $\widetilde{R}_i \equiv (R_{2i}, R_{3i})$, with $|\det(\widetilde{R}_i\widetilde{R}_i')| > 0$.

[10] In Klein and Sherman (2002), they use those shift restrictions to estimate the relative scaled thresholds in semiparametric ordered response models. In this paper, we generalize the shift restriction technique to allow endogenous regressors with correlated thresholds.

(Manski, 1985; Horowitz, 1992; Lewbel, 1997, 2003; Chen, 1999, 2000, etc). But most of them focus on models with only exogenous variables.

In what follows, assume that (β_0, α_0) is known to us. Recall from A-I.1 that for a representative category j, the soft adjustment can be expressed as $\delta_j(r) = E(T_{ji}|R_i = r)$, e.g., the conditional expectation of the $j's$ threshold given $R_i = r$. Likewise, we define $\Delta_{j,k}(r) \equiv E(T_{ki} - T_{ji}|R_i = r)$ as the conditional mean threshold difference between categories k and j.With the conditional cumulative rating probability function defined as

$$P_j(v, r) \equiv \Pr(Y_i \leqslant j|V_{0i} = v, R_i = r), \quad j \in \{0, 1, \cdots, J-1\}. \tag{3.4}$$

The above function measures the probability of bond i being rated into category j or above given its true risk index and shareholding relationship with the CRA. Identification of the soft adjustment relies critically on the implication of Proposition 3.2.1 below.[11]

Proposition 3.2.1 (Conditional Shift Restriction)**.** *Under Assumption A-I.1, for each* $(v, r) \in \mathbb{R} \times \mathcal{R}$*, then* $P_j(v, r) = P_k(v + \Delta_{j,k}(r), r)$*, for each* $j, k \in \{0, 1, \cdots, J-1\}$ *and*

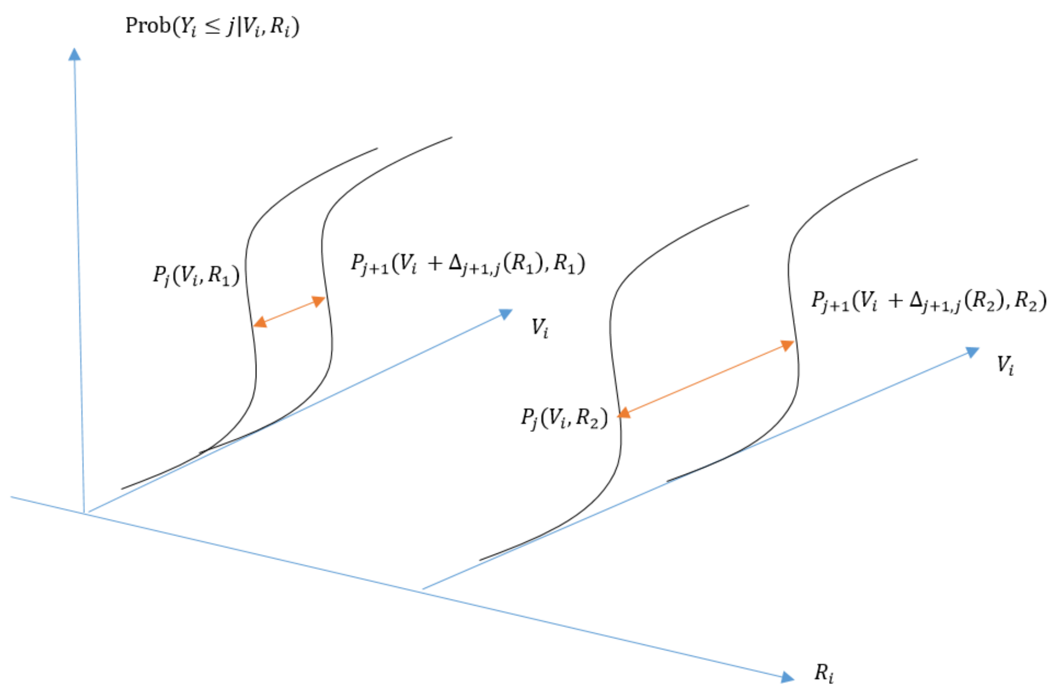
$$\Delta_{j,k}(r) = P_k^{-1}(P_j(v, r), r) - v \tag{3.5}$$

Proposition 3.2.1 introduces the conditional shift restrictions, a generalization of Klein and Sherman (2002)'s. In particular, it reveals the hidden restrictions across categories. Figure 3.1 depicts an example of shifts between jth-conditional probability functions and the $(j+1)$th for single-dimensional R_i. Given the index $V_{0i} = v$ and the conditioning variable $R_i = r$, one can equate $P_{j+1}(v, r)$ to $P_j(v, r)$ by increasing

[11]Note that A-I.1 is sufficient but not necessary for conditional shift restrictions. A weaker set of conditions require that the conditional distributions of u_{ij} be the same for each level j. However, we would lose the interpretation of the $\delta_{j,k}(r)$ being the soft adjustment.

the index v by $\Delta_{j,k}(r)$, which is precisely the conditional mean thresholds differences. Intuitively, identification of $\Delta_{j,k}(r)$ can be achieved by equating the two probability functions $P_{j+1}(v,r)$ to $P_j(v,r)$ for a given r.

Figure 3.1: Conditional Shift Restrictions from $P_j(V_i, R_i)$ and $P_{j+1}(V_i + \Delta, R_i)$



More importantly, $\Delta_{j,k}(r)$ can be identified by inverting $P_k(\cdot, r)$ for each r and $j, k \in \{0, 1, \cdots, J-1\}$. A formal proof of Proposition 3.2.1 is given in the appendix. Under proper support conditions, the unconditional mean of thresholds can be obtained by taking expectation, i.e. $\Delta_{j,k} = E[\Delta_{j,k}(R_i)]$.[12]

Recall that $\Delta_{j,k}(r)$ only measures the average distance between thresholds as a function of the shareholding structure R_i. In practice, the level of thresholds themselves are often of interest as they reflect the CRA's rating criteria. To make comparison of thresholds possible, we adopt the normalization as in A-I.2.

A-I.2 Base level. *There exists a known category j such that $\delta_j(r)$ is constant for any*

[12]In case the large support condition fails, one may compute a set-average expectation: $\Delta_{j,k}(\mathcal{R}^0) = E[\Delta_{j,k}(R)|R \in \mathcal{R}^0]$, where $\mathcal{R}^0 \subset \mathcal{R}$ is a compact set of interest.

$r \in \mathcal{R}$, *i.e.* $\partial\delta_j(r)/\partial r = 0$.

In the empirical estimation, we choose $j = 0$ and assume that there is essentially no heterogeneous soft adjustment for Aaa-rated bonds, i.e. $\delta_0(r) = 0$, for any $r \in \mathcal{R}$. We provide further support for our choice of the normalized category in the empirical section. As a result in Proposition 3.2.2, the soft adjustment at category $j \in \{1, \cdots, J - 2\}$ can be backed out as $\Delta_{0,j}(r) = E(T_{ji} - T_{0i}|R_i = r) = \delta_j(r)$.

Proposition 3.2.2 (Identification of Soft Adjustment)**.** *Under Assumption A-I.1 and A-I.2, $\delta_j(r)$ is identified, for each $r \in \mathcal{R}$ and $j \in \{0, 1, \cdots, J-1\}$.*

3.3 A Two-stage Semiparametric Estimator

In this section, we provide a two-stage semiparametric estimators for $(\theta_0, \mathbf{\Delta}(r))$ for each $r \in \mathcal{R}$, where $\mathbf{\Delta}(r) \equiv (\Delta_{0,1}(r), \Delta_{0,2}(r), \cdots, \Delta_{0,J-2}(r))'$ denotes the identified vector of threshold differences. In the first stage of estimation, we target at the index parameters, $\theta_0 \equiv (\beta_0, \alpha_0)$, up to location and scale by pseudo-maximum likelihood (ML) estimation. Then with the risk index estimator $\widehat{V}_i = X_i'\widehat{\beta}$ and the relationship control index $R_i'\widehat{\alpha}$, we estimate the conditional mean thresholds (or soft adjustment), $\widehat{\Delta}(r)$ at each point $r \in \mathcal{R}$ by a grid search estimator in the second stage. The grid search algorithm is attractive for its fast computing speed, as opposed to GMM and other extremum estimators.[13]

3.3.1 First Stage: Index Estimators

Conditional Probability Function

We begin by introducing the estimator of conditional rating probability function in Eq. (3.4), a basic building block that will be fed into the likelihood function.

[13]The two-stage estimator can be combined in a single-step GMM estimator, which might lead to more efficient estimation. However, doing it in two stages would be much faster in practice when the dimension of parameter space gets large.

Semiparametric index estimators have been extensively studied in the literature. For example, Manski (1985); Klein and Spady (1993); Powell *et al.* (1989); Ai and Chen (2003); Blundell and Powell (2004); Klein and Shen (2010); Hoderlein and Sherman (2015).

For any $(\beta, \alpha) \in \Theta$, define $V_i(\beta) \equiv X_i'\beta$ and $R_i(\alpha) \equiv R_i'\alpha$ and we suppress θ for notational simplicity whenever it is self-evident. We use the local constant kernel estimator to obtain the semiparametric conditional probabilities. In particular, the leave-one-out semiparametric estimator of the conditional probability function for $Y_i \leqslant j$ is used in Eq. (3.6),

$$\widehat{P}_j(i;\theta) \equiv \widehat{P}_j(V_i(\beta), R_i(\alpha)) = \frac{\sum_{l\neq i}^{N} K_h(V_l(\beta) - V_i(\beta))K_h(R_l(\alpha) - R_i(\alpha))\{Y_l \leqslant j\}}{\sum_{l\neq i}^{N} K_h(V_l(\beta) - V_i(\beta))K_h(R_l(\alpha) - R_i(\alpha))} \tag{3.6}$$

To seek a $\sqrt{N}$-consistent parameter estimator, one need resort to the bias reduction techniques to make sure that the asymptotic bias vanishes faster than $\sqrt{N}$ in the limit.[14]

Pseudo-ML Estimator

Note that the double-index parameters can be solved in a semiparametric pseudo-MLE framework similar to Klein and Vella (2009) and Maurer *et al.* (2011). It requires only one-step of optimization like below. Define $\widehat{P}_{-1i}(\theta) = 0$ and $\widehat{P}_{Ji}(\theta) = 1$.

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} N^{-1} \sum_{i=1}^{N} \sum_{j=0}^{J} \widehat{t}_i \{Y_i = j\} \ln\left(\widehat{P}_j(i;\theta) - \widehat{P}_{j-1}(i;\theta)\right) \tag{3.7}$$

where the trimming function estimator $\widehat{t}_i = \Pi_{k=1}^{d_X+d_R}\{\widehat{q}_{Z_k}(\tau_l) < Z_{ki} < \widehat{q}_{Z_k}(\tau_u)\}$ is the product of the indicator functions for each continuous Z_k, with fixed lower and upper quantiles τ_l and τ_u, where $Z_i = (X_i', R_i')'$. $\widehat{q}_{Z_{ik}}(\tau)$ is estimated by the empirical quantile function, $\inf\{z_k : N^{-1}\sum_{i=1}^{N}\{Z_{ki} \leqslant z_k\} \geqslant \tau\}$.

[14] In principle, one may use higher-order kernels, local smoothing and the recursive methods in Klein and Shen (2016). In practice, we only use Silverman's rule-of-thumb bandwidths as we do not find large and significant change to our results with bias-correction techniques.

3.3.2 Second Stage: Conditional Mean Thresholds $\Delta(\cdot)$

The shift restrictions naturally imply an extremum-type estimator by minimizing the distance between $P_j(V_i, r)$ and $P_k(V_i + \Delta_{k,j}(r), r)$ for each r and $j \neq k$. For a J-supported Y_i, there are totally $\binom{J-2}{2}$ possible restrictions to choose from. For parsimonious reason, we only consider the shift conditions of adjacent levels. Additional restrictions could be used to increase efficiency and conduct an overidentification test.

However, in terms of computing time, the optimization needs to be done repeatedly for each value of R_i in the sample or of particular interest. This can take quite long time once the support of Y_i is large. To this end, we choose to estimate it by directly inverting the conditional probability functions following the identification condition in Eq. (3.5). Since the equality holds for each value of v, the final estimator takes the form of averaging over all empirical points of $V_i(\beta)$, for $i \in \{1, 2, \cdots, N\}$.

$$\widehat{\Delta}_{j,j-1}(r) = \frac{1}{N}\sum_{i=1}^{N}\left[\widehat{P}_{j-1}^{-1}\left(\widehat{P}_j(V_i(\widehat{\beta}), r), r\right) - V_i(\widehat{\beta})\right], \quad j \in \{1, \cdots, J-1\} \tag{3.8}$$

We choose to estimate the adjacent levels since the range of overlapping region is the widest. Without redundant information, we are left with $J-2$ restrictions and for each $r \in \mathcal{R}$. To implement it, we use a grid search algorithm which is an extension of the grid search estimator in Klein and Sherman (2002).[15] To be specific, one can repeat the following four steps for each $r \in \mathbb{R}$ and $j \in \{1, \cdots, J-1\}$,

1. Estimate $\widehat{P}_j(\widehat{V}_i(\widehat{\beta}), r)$ nonparametrically for each i.
2. Estimate $\widehat{P}_{j-1}(v, r)$ nonparametrically at each v over a set of grid points, $\mathcal{V}_N^g$.
3. Find the closest v such that $V_i^* = \arg\min_{v \in \mathcal{V}_N^g} |\widehat{P}_j(\widehat{V}_i(\widehat{\beta}), r) - \widehat{P}_{j-1}(v, r)|$ for each i.

[15] The joint estimation of a vector of conditional thresholds using extremum-type estimators could be more efficient than our grid search methods. However, not only the computational time would substantially increase but the estimates can be sensitive to starting values.

4. Compute $\widehat{\Delta}_{j,j-1}(r)$ as $N^{-1}\sum_{i=1}^{N} V_i^* - V_i(\widehat{\beta})$.

For the choices of grid sets $\mathcal{V}_N^g$, it is advised that the adjacent interval distance should be smaller than $O(1/\sqrt{N})$ in order to be negligible in the limit. Given the fast speed of the grid search algorithm, one can pick an even finer grid, though we find the empirical differences are not quite significant. The relative conditional mean threshold of level j with respect to the base level (namely $Y_i = 0$) is readily available by multiplying a lower triangular matrix A with entry equal to 1 below and along the diagonal. Let $\widehat{\Delta}^0(r) = A\widehat{\Delta}(r)$, so $\widehat{\Delta}^0(r) = (\widehat{\Delta}_{1,0}(r), \widehat{\Delta}_{2,0}(r), \cdots, \widehat{\Delta}_{J-2,0}(r))'$.

As suggested in Assumption A-I.2, we normalize the base level $\delta_0(r) = 0$ in order to back out the unobserved soft adjustment starting from $j = 2$ to $j = J - 1$. To do so, we first calculate the empirical control index $R_i(\widehat{\alpha}) = R_i'\widehat{\alpha}$ and then compute the relative thresholds evaluated at each $R_i(\widehat{\alpha})$. By definition, estimates of individual-bond soft adjustment at each category j would be $\widehat{\delta}_{ij} = \widehat{\Delta}_{j,0}[R_i(\widehat{\alpha})]$, a primary measure to be examined after policy change.

Note that the consistency and asymptotic normality of the finite-dimensional index parameter estimators are standard in the semiparametric literature. For index estimators, we apply the asymptotic normality condition in Klein and Sherman (2002) to double-index models. For the soft adjustment estimator in Eq. (3.8), consistency of $\widehat{\delta}_j(r)$ would straightforwardly follow once the consistency of $\widehat{P}_j(\cdot, r)$ and $\widehat{V}_i$ are established.[16] For the inference, we bootstrap the variances starting only from the second stage. To implement it, we repeatedly draw samples with the same number of observations over all possible $(Y_i, \widehat{V}_i, R_i)$ with replacement. Theoretical properties of semiparametric bootstrapped inference is understudied in the literature despite its prevalence in empirical studies (Simar and Wilson, 2007, etc).[17] Our Bootstrapped approach falls into a general class of semiparametric M-estimators studied in Cheng

[16]The point-wise consistency result is given in the online appendices.

[17]Our first-stage index estimator converges at $\sqrt{N}$, a faster rate than the second-stage threshold estimator. Therefore, for computational concerns, we run Bootstrapped inference starting from the second-stage provided that the variance of the index estimator does not contribute in the limit.

et al. (2010) who prove its validity under relatively weak conditions. Fortunately, our computational time is relatively fast, even for some large number of bootstrapped samples.

3.4 Data and Context

3.4.1 Institutional and Regulatory Environment

As the information intermediaries of the financial system, a credit rating agency's primary function is to evaluate a particular debt instrument's credit worthiness. As noticed by Cantor *et al.* (1994) and White (2002), the credit rating industry in the U.S. is highly concentrated: with the"Big Three" credit rating agencies controlling more than 95% of the ratings business. Moody's and Standard & Poor's (S&P) together control 80% of the global market, and Fitch Ratings controls a further 15% (Alessi *et al.*, 2013). Of the two biggest agencies Moody's became a public firm in 2001, while Standard & Poor's is a private division of the McGraw-Hill.

Given the massive defaults of highly-rated securities during the last financial crisis, various reforms have been proposed to regulate the behaviors of CRAs. In the famous Dodd-Frank Wall Street Reform and Consumer Protection Act (Pub.L. 111203[18], H.R. 4173[19]), an entire section is devised to improve the transparency of credit rating agencies, by means of enforcing public disclosure of credit rating methodologies, data, and etc. In subtitle C of Title IX of the amendments, it emphasizes the "improvements to the regulation of credit rating agencies, critical *gatekeeper* in the debt market central to capital formation, investor confidence, and the efficient performance of the United States economy." Subtitle C also cites findings of conflicts of interest and inaccuracies during the recent financial crisis contributed significantly to the mismanagement of risks by financial institutions and investors, which in turn adversely impacted the

[18] https://www.gpo.gov/fdsys/pkg/PLAW-111publ203/html/PLAW-111publ203.htm

[19] https://www.congress.gov/bill/111th-congress/house-bill/4173

health of the United States economy.[20] The Franken-Wicker amendment to the Dodd-Frank financial reform law,[21] taking a somehow more extreme approach, suggests to use a governmental entity to assign securities to qualified ratings agencies based on capacity and expertise. Recall from the earlier discussion that rating agencies make soft adjustment based on private information. By estimating how soft adjustments evolve over time, we aim to assess the effectiveness of Dodd-Frank in enhancing information transparency.

3.4.2 Data and Summary Statistics

Our data derive from multiple sources. The data on the history of credit rating by Moody's is obtained from the Mergent's Fixed Income Securities Database (FISD). Our sampling period spans from 2001, when Moody's went IPO, to 2016, with the enactment of Dodd-Frank in July 2010. We exclude government bonds and retain all initial ratings on bonds issued by firms covered in both Center for Research in Security Prices (CRSP) and Compustat. Institutional shareholding data are obtained from Thomson Reuters 13F database.

Implied by our identification strategy, we choose to focus only on publicly listed firms. Recall that our identification strategy critically assumes that material non-public information can be passed to the CRA through the CRA-issuer connectedness. Being "residual claimants" —agent who have the sole remaining claim on an organization's net cash flows — equity holders, compared with bondholders, have greater incentive to monitor corporate management and acquire inside information (Shleifer and Vishny, 1986). In addition, as required by public disclosure policy, the data on firm

[20]This law required the SEC to establish clear guidelines for determining which credit rating agencies qualify as Nationally Recognized Statistical Rating Organizations (NRSROs) who are required to establish, maintain, enforce and document an effective internal control structure governing the implementation of and adherence to policies, procedures, and methodologies for determining credit ratings. See Partnoy (2009) and White (2002) for the importance of such oversight. https://www.sec.gov/spotlight/dodd-frank/creditratingagencies.shtml

[21]https://www.sec.gov/comments/4-629/4629-28.pdf

fundamentals are easier to obtain and of higher quality for public firms. For both reasons, we examine a sample of public firms whose stocks are held by Moody's shareholders.[22]

It is worth pointing out that we choose to focus exclusively on Moody's rating process because of the agency's unique exposure to conflicts of interest. Among the "big three" rating agencies on the market, Fitch is a private firm; Being a private division under McGraw-Hill, the conflicts-of-interest effect is complicated and indirect for S &P. Focusing exclusively on Moody's ratings, however, might lead to sample selection bias. However, we do not view this as a serious problem in our sample because most of bonds issued by public firms have been rated by at least two firms and their ratings are mostly matched after being converted to the same standard (70% of the ratings assigned by S&P and Moody's differ by at most one notch). This implies that agency heterogeneity does not play quite a fundamental role in the rating methodology to some extent.

Last but not least, we exclusively focus on initial ratings for two reasons. First, the impact of the soft adjustment is most pronounced on initial ratings. Many market participants, including investors, issuers, and regulators, have a strong preference for corporate bond ratings that are not only accurate but also stable. To achieve long-term rating stability, rating agencies are reluctant to change the credit rating unless such effects are believed to be permanent (Altman and Rijken, 2004). Secondly, the proposed estimation and inference framework are designed for cross-sectional applications. Econometrically, modeling a panel of rating changes with upgrades/downgrades, is itself challenging, especially for nonparametric models.

Applying the above restrictions leave us with a final sample of 11,134 initial bonds issued by 1462 firms. Given the aforementioned regulatory changes, we divide the sample into two time periods by the enactment of the Dodd-Frank on July 21, 2010. To

[22]Alternatively, one can also examine firms whose debt are held by Moody's shareholders. We did not go with this approach because we believe that Moody's is more likely to possess soft information on firms whose stocks are externally held by Moody's shareholders.

allow for possible implementation lags, we alternatively define the post Dodd-Frank period starting from the beginning of 2011. By comparing estimates of soft adjustment in both periods, we aim to examine the policy effect of Dodd-Frank Act on the credit rating outcomes[23]. There are 2,540 observations in the crisis period, accounting for 38.4% of the total before the Dodd-Frank.

The distribution of ratings are presented in Table 3.1 by years, with Aaa being the highest credit category and C the lowest. Figure 3.2 compares the rating distribution before and after the Dodd-Frank. Noticeably, Moody's ratings become more centered around Baa grade after the Dodd-Frank. One plausible explanation is that the emergence of a variety of credit derivatives, such as the CDS, makes hedging downgrade risks easier, inducing firms to take more financial leverage. Due to a higher leverage, investment grades bonds are less likely to receive high credit ratings, which might explain why the fraction of Aa rated bonds decreased while the fraction of Baa bonds increased.

Firm and Bond Characteristics

Using data from quarterly Compustat-CRSP merged database and FISD, we construct a sequence of predictors for credit ratings mentioned in the bond rating literature (Pinches and Mingo, 1973; Kaplan and Urwitz, 1979; Blume *et al.*, 1998; Jiang *et al.*, 2012; Campbell and Taksler, 2003, etc). To construct these variables, short-term and long-term debt for each bond issuers are from quarterly Compustat-CRSP merged dataset.[24] The end of quarter stock price data and number of shares outstanding data are also taken from Compustat-CRSP. All financial ratios are computed using a 5-year arithmetic average of the annual ratios, as Kaplan and Urwitz (1979) points out that bond raters might look beyond a single year's data to avoid temporary anomalies.

---

[23] In the pre Dodd-Frank period, a crisis dummy is created to capture the financial crisis effect from 2007 to 2010.

[24] Short-term debt is estimated as the larger of Compustat items 118 ("Debt in current libabilities") and 224 ("Total current liability"). Long-term debt is taken from item 119 ("Total long-term liability").

Table 3.1: Moody's Rating Outcomes by Year

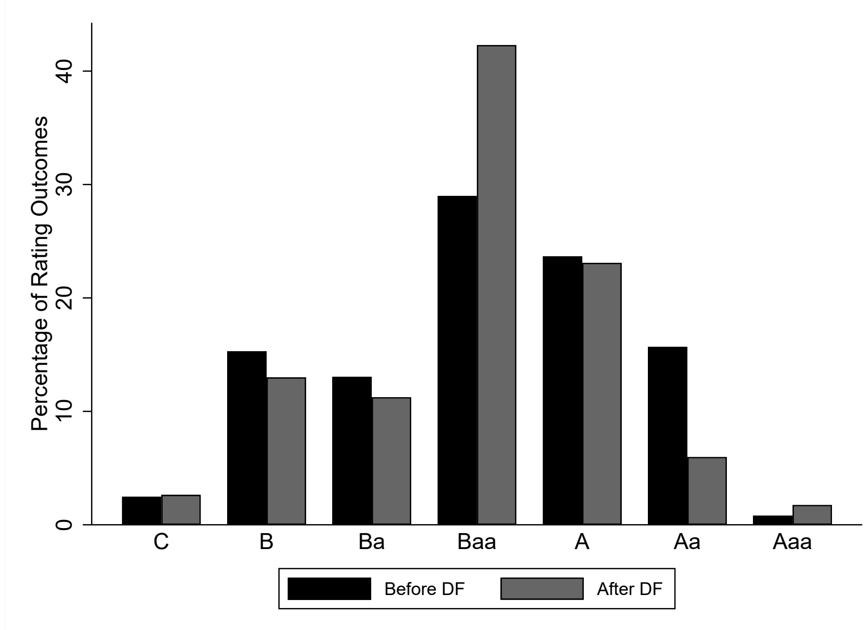
Year	Aaa	Aa	A	Baa	Ba	B	C	Total
2001	10	45	171	221	122	96	11	676
2002	1	78	146	223	85	107	7	647
2003	9	112	155	219	131	174	32	832
2004	3	85	95	177	103	160	18	641
2005	6	118	115	161	94	92	15	601
2006	4	164	163	195	63	68	24	681
2007	9	238	332	167	55	75	13	889
2008	2	110	156	143	29	12	4	456
2009	3	35	129	230	89	104	13	603
2010	7	53	105	183	93	125	26	592
2011	10	38	142	226	41	98	17	572
2012	3	43	166	289	93	134	25	753
2013	12	58	181	325	109	117	36	838
2014	8	37	144	324	94	102	25	734
2015	20	35	218	399	90	68	10	840
2016	26	59	192	347	81	68	6	779
Total	133	1,308	2,610	3,829	1,372	1,600	282	11,134

Note: Subtiers are aggregated together into general tiers, e.g. A consists those rated as A1, A2 and A3.

The selected predictors consists of firm characteristics (1)-(4) and bond characteristics (5)-(6) as follows: (1) ASSET: denotes issuer size, defined as the value of the firm's total asset. (2) LEVERAGE: denotes firm leverage, defined as the ratio of long-term debt to total assets. (3) PROFIT: denotes operating performance, defined as operating income before depreciation divided by sales. (4) CVTA: denotes asset stability, defined as the variance of the firm's total asset in the year prior.[25] (5) OFFAMT: denotes the offering amount, defined as the par value of the bond issued. (6) SENIOR: denotes subordination status, which a dummy variable equals to one if the bond is a senior bond and 0 otherwise. We take the log of both sizing variables (OFFAMT, ASSET) to make all covariates roughly have the same scale as their differences in denominations can be potentially large. Summary statistics of the ratings and explanatory variables can be found in the upper panel of Table 3.2. As motivated in the the behavioral framework, LEVERAGE, OFFAMT and SENIOR are

[25]The definition here follows that from (Kedia *et al.*, 2017).

Figure 3.2: Rating Outcome Distributions Before and After the Dodd-Frank Act



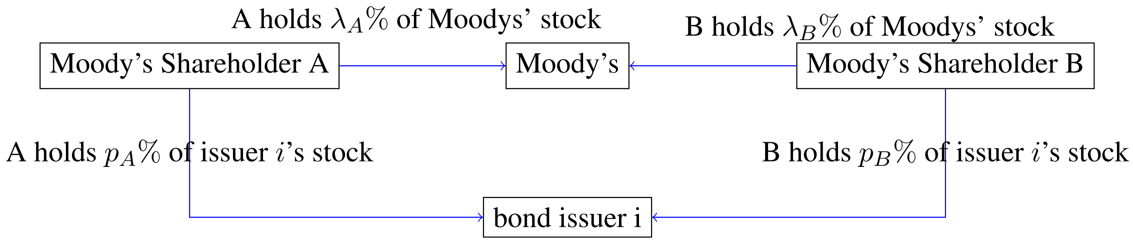
Note: 1. DF is the Dodd-Frank indicator that equals 1 for those after 2010 and 0 if before 2010.

likely to be endogenous as firms might issue more debt when they "foresee" a chance of higher ratings.

The Shareholding Relations

Recall from Section 2 that we need a vector R_i to characterize the shareholding relationship between Moody's and each bond issuer i, so we can address the endogeneity problem by conditioning on this control vector R_i. Specifically, we use three variables to jointly capture the shareholding relationship. That is, $R_i \equiv \{\text{Mshare}_i, \text{Fshare}_i, \text{LargeSH}_i\}$. To convey some intuition on the definition of these variables and why they are selected, consider a bond issuer i that is jointly invested by two shareholders of Moodys, A and B, as described below:[26]

[26]However, Moodys could have shareholders who do not invest in the bond issuer i at all.



The shareholding relationship between Moody's and bond issuer i, in a sense, can be characterized by both *the importance of shareholders to Moody's* (captured by the $\lambda's$) and *the importance of bond issuer i to the shareholders* (captured by the $p's$)[27]. To be more precise, we aggregate the two shareholding percentage measure λ and p across all common shareholders to approximate bond issuer i's overall ownership interaction with Moody's (in this illustrative example, namely $\lambda_A\% + \lambda_B\%$ and $p_A\% + p_B\%$, respectively). Extending to the case with M_i common shareholders, we define:

$$Mshare_i \equiv \sum_{j=1}^{M_i} \lambda_j\%, \quad Fshare_i \equiv \sum_{j=1}^{M_i} p_j\%,$$

to capture the importance of shareholders to Moody's and the importance of bond issuer i to the shareholders. In addition, to highlight the individual shareholder's influence, we define:

$$\text{largeSH}_i \equiv 1\{\text{issuer } i \text{ is invested by at least one large shareholder of Moody's}\}$$

where $1\{E\}$ takes value one if E is true and zero otherwise. In particular, "large" shareholder are those who own at least 5% of Moodys' stock. The significant influence of large shareholders is also documented in Kedia *et al.* (2017).

The descriptive statistics for these three measures are presented in the lower panel of Table 3.2. In general, a larger Mshare, Fshare or largeSH indicates a stronger connection with the CRA. Mshare and largeSH provide the only channel for the

[27] This characterization is enlightened by Kedia *et al.* (2017), in which the authors find Moody's has an upward bias towards issuers that are large investees or subsidiary firms of its large shareholders.

conflict-of-interest to impact ratings, with the magnitude being mediated by the level of Fshare. In contrast, Fshare is supposed to pick up the unobserved bond or firm quality that is also contained in the CRA's soft adjustment. By assumptions, after conditioning Mshare, Fshare and largeSH, issuers should bear no more soft information related to issuing decisions.

As a robustness check, we have also experimented with other measures such as number of common shareholders, number of influential shareholders, number of bonds rated before as well as the weighted shares in common shareholders' portfolio. Moreover, we checked the quadratic and cubic functional forms. However, inclusion of additional variables or higher order terms increases the collinearity of the control index and lead to nonsensical coefficient estimates. Therefore, we stick to the above parsimonious specification.

Table 3.2: Descriptive Statistics

	Before DF: 2000-2010				After DF: 2011-2016			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
	Issuer and Bond Financial Characteristics							
ASSET	161.301	325.751	0.078	3065.556	166.851	446.162	0.175	2601.652
CVTA	0.181	0.146	0.000	1.342	0.169	0.175	0.008	1.504
LEVERAGE	0.259	0.172	0.000	1.283	0.260	0.156	0.000	1.021
PROFIT	0.028	0.058	-0.682	0.503	0.046	0.057	-0.366	0.264
OFFAMT	595.5	3860.1	0.0	250000.0	774.3	2017.2	0.0	100000.0
SENIOR	0.823	0.381	0	1	0.908	0.289	0	1
CRISIS	0.384	0.486	0	1	0	0	0	0
	Common Shareholder Information							
Mshare	44.896	15.468	0	98.794	45.808	15.073	0	89.296
Fshare	46.508	17.066	0	93.204	45.909	17.724	0	99.528
largeSH	0.620	0.485	0	1	0.898	0.302	0	1
Obs.N	6,618				4,516			

Note: 1. ASSET and OFFERAMT are measured in thousand dollars (1000 $) whereas logged asset and offering amount are used in estimation. 2. Mshare and Fshare are measured in percentage. 3. The crisis dummy is equal to 1 if year is in between 2007 and 2010.

3.4.3 Correlation Analysis

To motivate our selected control covariates, we start by presenting some simple correlation analysis between the cumulative rating outcomes and control variables in Table 3.3. If the control varibles could indeed capture the effect of CRA-issuer liasion on ratings, we ought to see some co-movement between them: issuers that are close to Moody's ownership-wise should be assigned higher ratings. In Table 3.3, we divide the whole sample into the before and after Dodd-Frank periods, recognizing the structural change of regulatory environment.

For Mshare, the correlations are positive and consistent, suggesting that a strong firm-Moody tie always corresponds to a upward pressure for ratings, though the magnitudes overall decease after the Dodd-Frank. But for Fshare, a negative correlation is found for investment-level grades, especially after the regulatory change. The magnitudes of having at least one influential common investor are small relative to the other measures and its effects are even lower after the reform. To sum up, our control covariates do convey some predictive power on ratings and should not be simply left out of the model. From the correlation table, we can also conjecture that the effect of CRA-issuer relation on ratings might be highly heterogeneous. For bonds with extremely high or low ratings, the correlation between the rating and the control variables is quite small. Our empirical model is able to capture such heterogeneity as the functional form of the soft adjustment is modeled flexibly other than the single-index restriction. Besides, the structural change of correlation pattern suggest us to estimate the model separably for the periods before and after the Dodd-Frank.

3.5 Empirical Results

We report results in four subsections. First, we estimate a series of parametric rating models in the literature, and argue that the parameters of interest are sensitive to model specification. Hence, it is important to consider a more flexible approach to understand

Table 3.3: Correlation between Control Variables and Rating Outcomes

	Before DF: 2000-2010 (N=6,618)								
	Aaa	>=Aa	>=A	>=Baa	>=Ba	>=B	Mshare	Fshare	largeSH
Mshare	0.106	0.430	0.498	0.455	0.392	0.130	1.000		
Fshare	-0.018	0.073	0.101	0.097	0.125	0.082	0.470	1.000	
largeSH	0.047	0.220	0.150	0.112	0.079	0.026	0.463	0.191	1.000
	After DF: 2011-2016 ($N = 4,516$)								
	Aaa	>=Aa	>=A	>=Baa	>=Ba	>=B	Mshare	Fshare	largeSH
Mshare	0.166	0.216	0.375	0.342	0.308	0.165	1.000		
Fshare	-0.018	-0.166	-0.034	0.048	0.107	0.128	0.556	1.000	
largeSH	0.045	-0.062	0.044	0.001	0.016	-0.010	0.587	0.560	1.000

Note: 1. >=X represents cumulative ratings above or equal to notch X.

the rating process. In the second and third subsection, we estimate the structural rating model proposed in Section 3.2 and reports the estimates of index parameters and soft adjustment. Lastly, we discuss patterns of threshold parameters over a changing shareholding relationship.

3.5.1 Parametric Results

The results of parametric regressions are presented in Table 3.4. The first three specifications are estimated using ordered probit models while the last two are from ordered logit models. Both models have been employed extensively in the empirial credit rating literature. Recall that a CRISIS dummy with full interactions with all covariates are deployed to control for the financial crisis effect for the period before Dodd-Frank.

There are two main findings. First, some regression coefficients change significantly after the passage of Dodd-Frank. PROFIT and OFFAMT have a much larger impact on default risk in the post Dodd-Frank period, indicating that Moody's has become more stringent on firms with low profitability ratio and high debt. The disparity reconfirms the existence of a structural break of rating models. Second, we note that the regression coefficients are sensitive to the distributional assumption of the error term. Oprobit-2 and Ologit-1 (or Oprobt-3 and Ologit-2) differs only by the error distribution. Despite most of the regression coefficients have the same signs as predicted, the logit regression coefficients on firm characteristics (ASSET, CVTA, LEVERAGE and PROFIT) are nearly twice as large as the probit coefficients.[28] The somewhat inconsistency results from different parametric specifications calls for a robust approach which does not overly restrict error distributions.

3.5.2 Semiparametric First Stage: Index Parameters

In this section, we estimate the normalized index parameters defined in Section 2. Recall that these index parameters reflect how ratings are driven by observed "hard information", such as ASSET, LEVERAGE, etc. To fix idea, we focus on comparing the estimation results along two dimensions: (i) between the semiparametric framework proposed in this paper and the baseline parametric model, and (ii) before and after the passage of Dodd-Frank.

The estimation results are shown in Table 3.5 and 3.6, respectively for the period before and after Dodd-Frank. To facilitate comparison, the first column of each table "Oprobit-R" gives ratios of estimated coefficients relative to ASSET and Mshare, the variables that we choose to normalize on, respectively.[29] Estimation results of the suggested semiparametric double-index model, "Semi-R", is reported in the last column. As opposed to column 1, we allow arbitrary interactions between shareholding relation with other characteristics and do not need to specify the error term distribution. Another interesting experiment we do, in order to assess the impact of ignoring soft

[28]Most coefficients have the correct predicted signs across the board: the amount of total asset, profitability and being a senior bond all have negative impact on the default risk index and thus leads to higher ratings. On the other hand, high variance of assets, measured by CVTA as well as high leverage ratios are in accordance with a larger default risk index.

[29]The standard errors of parameter ratios are calculated using the delta-method: via a first order Taylor expansion around true values, $\widehat{\beta}_2/\widehat{\beta}_1 - \beta_{20}/\beta_{10} \approx -(\widehat{\beta}_1 - \beta_{10})\beta_{20}/\beta_{10}^2 + (\widehat{\beta}_2 - \beta_{20})/\widehat{\beta}_1$ and then compute the standard error. Admittedly, the semiparametric model can identify only the relative ratios of coefficients. To construct comparable default risk index, one can back out individual coefficients assuming $\widehat{\beta}_1$ is close to that of ordered probit model.

Table 3.4: Parametric Specifications of Preliminary Estimates

	Before DF: 2000-2010					After DF: 2011-2016				
Variables	Oprobit-1	Oprobit-2	Oprobit-3	Ologit-1	Ologit-2	Oprobit-1	Oprobit-2	Oprobit-3	Ologit-1	Ologit-2
ASSET	-0.587	-0.577	-0.485	-1.131	-0.980	-0.506	-0.500	-0.432	-0.911	-0.792
	(0.010)	(0.012)	(0.014)	(0.023)	(0.028)	(0.013)	(0.013)	(0.014)	(0.025)	(0.027)
CVTA	0.394	0.554	0.634	1.335	1.399	0.571	0.552	0.618	0.990	1.092
	(0.099)	(0.123)	(0.124)	(0.223)	(0.225)	(0.106)	(0.105)	(0.107)	(0.187)	(0.191)
LEVERAGE	2.336	2.326	2.289	4.096	4.090	2.054	2.074	2.099	3.845	3.899
	(0.103)	(0.123)	(0.123)	(0.230)	(0.233)	(0.129)	(0.128)	(0.129)	(0.230)	(0.232)
PROFIT	-9.649	-9.192	-8.457	-21.376	-19.406	-16.620	-16.488	-15.265	-29.883	-27.718
	(0.278)	(0.332)	(0.341)	(0.781)	(0.799)	(0.386)	(0.384)	(0.398)	(0.736)	(0.750)
OFFAMT	0.018	0.033	0.057	0.056	0.091	0.179	0.174	0.188	0.325	0.352
	(0.009)	(0.013)	(0.013)	(0.022)	(0.022)	(0.025)	(0.025)	(0.025)	(0.044)	(0.045)
SENIOR	-0.523	-0.572	-0.538	-1.056	-1.021	-0.817	-0.806	-0.784	-1.376	-1.351
	(0.038)	(0.045)	(0.046)	(0.082)	(0.083)	(0.062)	(0.062)	(0.062)	(0.117)	(0.118)
Mshare			-0.026		-0.040			-0.025		-0.046
			(0.002)		(0.004)			(0.002)		(0.003)
Fshare			0.011		0.018			0.007		0.012
			(0.001)		(0.002)			(0.001)		(0.002)
largeSH			0.081		0.099			0.654		1.245
			(0.039)		(0.070)			(0.077)		(0.140)
CRISIS		0.634	1.354	1.408	2.734					
		(0.326)	(0.353)	(0.588)	(0.645)					
CRISIS*VAR		Y	Y	Y	Y					
Year	Y					Y				
Obs. N	6618	6618	6618	6618	6618	4516	4516	4516	4516	4516

Note: 1. CRISIS*VAR indicates the inclusion of full interaction terms with the crisis dummy. 2. Oprobit-1 is the specification controlling for year fixed effects. 3. s.e. are presented in parenthesis.

adjustment, is estimating a single-index model “Semi-X” in which we deliberately drop the shareholding relation index $R_i(\alpha)$.

We find the shareholding relation and the CRISIS dummy have very different estimated impacts when switching from ordered-probit to the semiparametric framework. Taking the number of common large shareholder (largeSH) as an example. In the post Dodd-Frank period (reported in Table 3.6), Oprobit-R predicts a significantly negative impact on default risk, suggesting that Moody’s ratings are more favorable to issuers who are invested by its own large shareholders. In contrast, Semi-R predicts a positive impact with similar magnitude, suggesting Moody’s has become more stringent on these related firms. The disparity is not only statistically significant, but economically large. Estimated impact of the CRISIS dummy also differs across the board. From Semi-R, most of the interaction terms have insignificant coefficients, meaning that firm and bond characteristics roughly have the same impact on ratings in and out of economic downturns. In contrast, from Oprobit-R, we find firm’s financial stability (CVTA) and profitability (PROFIT) have significantly less impacts on rating during the crisis period. As can be seen in Semi-X, interaction terms also have differential impacts on ratings after we drop the shareholding relation variables.

Turning to the comparison before and after the Dodd-Frank, the differences in estimated coefficients are equally striking. First, Moody’s attention to firm and bond characteristics has clearly changed over time. Specifically, the impact of firm stability (CVTA) decreases by half after the passage of Dodd-Frank, whereas the impact of profitability (PROFIT) increases by half. The biggest difference comes from the relative importance of issuing amount: the estimated effect increasing by a factor of ten, reflecting Moody’s has increased its scrutiny on the amount of debt that an issuer is taking. Second, the impact of shareholding relation also changes over time. In particular, relationship with Moody’s large shareholders led to higher ratings before the Dodd-Frank. Such effects, however, reverse sign and enlarge in magnitude after the Dodd-Frank.

The aforementioned findings suggest that Moody's rating model has changed significantly after the Dodd-Frank, in terms of its relative focus on specific characteristics as well as its treatment to issuers in terms of shareholding relations. Recall that we use the shareholding relation to "anchor" the amount of soft information that the CRA may receive from common shareholders. Thereby, the differential impact of shareholding relation before and after the Dodd-Frank may reflect a substantial change in terms of how the CRA utilize soft information to determine ratings. To investigate this issue further, in the next section we estimate the soft adjustment for each bond issuer given its shareholding relation with Moody's and report the distributional pattern of soft adjustment.

Table 3.5: Estimation Results of Creditworthiness Index Parameters before Dodd-Frank Act

	Parametric	Semiparametric	
Variables	Oprobit-R	Semi-X	Semi-R
Structural Financial Risk Parameters			
CVTA	-1.306 ***	-1.773 ***	-1.884 ***
	(0.259)	(0.204)	(0.213)
LEVERAGE	-4.715 ***	-3.467 ***	-3.862 ***
	(0.320)	(0.254)	(0.298)
PROFIT	17.422 ***	24.339 ***	23.919 ***
	(0.706)	(0.697)	(0.788)
OFFAMT	-0.118 ***	-0.014	-0.047 **
	(0.027)	(0.016)	(0.019)
SENIOR	1.109 ***	0.928 ***	0.994 ***
	(0.101)	(0.079)	(0.082)
CRISIS	-2.790 ***	-1.794 ***	-1.702 ***
	(0.757)	(0.514)	(0.575)
CRISIS*ASSET	0.222 ***	0.110 ***	0.168 ***
	(0.045)	(0.029)	(0.034)
CRISIS*CVTA	0.821 **	0.310	0.073
	(0.401)	(0.314)	(0.329)
CRISIS*LEVERAGE	0.117	0.898 **	0.440
	(0.444)	(0.366)	(0.443)
CRISIS*PROFIT	3.850 ***	-1.994 *	-1.359
	(1.184)	(1.148)	(1.155)
CRISIS*OFFAMT	0.044	0.031	-0.011
	(0.038)	(0.022)	(0.027)
CRISIS*SENIOR	-0.389 **	-0.575 ***	-0.526 ***
	(0.174)	(0.148)	(0.173)
Control Index Parameters			
Fshare	-0.444 ***		-1.288 ***
	(0.106)		(0.036)
largeSH	-3.185 *		-1.903 **
	(1.735)		(0.812)
CRISIS*Mshare	-0.765 ***		0.046
	(0.247)		(0.048)
CRISIS*Fshare	0.428 ***		-0.068
	(0.120)		(0.049)
CRISIS*largeSH	1.558		-6.370 ***
	(2.760)		(1.423)
ASSET	-0.485 ***		
Mshare	-0.026 ***		
$N = 6618$			

Note: 1. The estimation uses data from 2000 to 2010. 2. Estimates represent normalized coefficient ratios with respect to log of asset and Mshare, respectively for financial and control parameters. 3. Oprobit-R is estimated by MLE. Semi-X and semi-R are estimated by pseudo-MLE. 4. Standard errors are in parentheses. 5. Significant level: *10 percent, **5 percent, ***1 percent.

Table 3.6: Estimation Results of Creditworthiness Index Parameters after Dodd-Frank Act

	Parametric	Semiparametric	
Variables	Oprobit-R	Semi-X	Semi-R
Structural Financial Risk Parameters			
CVTA	-1.397 ***	-0.447 ***	-0.915 ***
	(0.265)	(0.182)	(0.263)
LEVERAGE	-4.893 ***	-3.140 ***	-3.939 ***
	(0.380)	(0.282)	(0.343)
PROFIT	35.361 ***	28.975 ***	32.416 ***
	(1.191)	(0.890)	(1.416)
OFFAMT	-0.445 ***	-0.558 ***	-0.430 ***
	(0.055)	(0.041)	(0.065)
SENIOR	1.834 ***	1.308 ***	1.154 ***
	(0.148)	(0.158)	(0.215)
Control Index Parameters			
Fshare	-0.279 ***		-2.409 ***
	(0.091)		(0.181)
largeSH	-26.150 ***		26.308 ***
	(4.886)		(4.773)
ASSET	-0.430 ***		
Mshare	-0.025 ***		
$N = 4516$			

Note: 1. The estimation uses data from 2011 to 2016. 2. Estimates represent normalized coefficient ratios with respect to log of asset and Mshare, respectively for financial and control parameters. 3. Oprobit-R is estimated by MLE. Semi-X and semi-R are estimated by pseudo-MLE. 4. Standard errors are in parentheses. 5. Significance level: *10 percent, **5 percent, ***1 percent.

3.5.3 Second Stage: Soft Adjustment

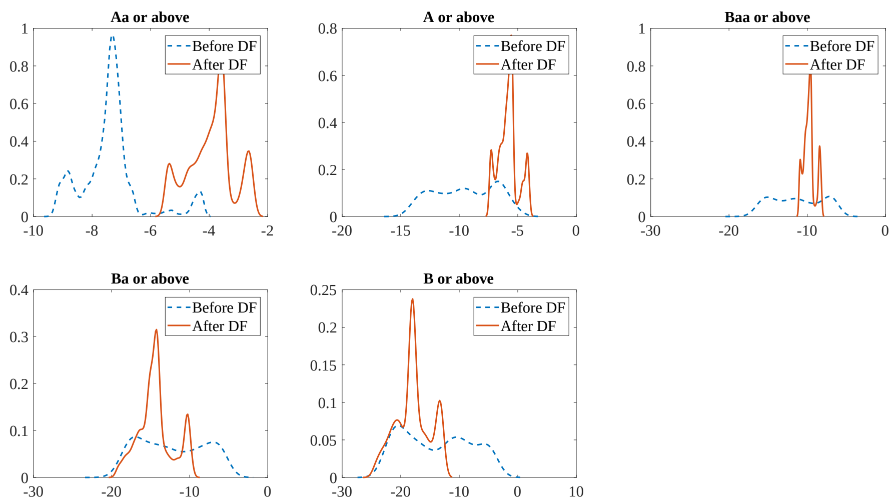
In this section, we estimate Moody's soft adjustment to each bond issuers: that, say, to what extent would Moody's alter the threshold parameters T_{ji} given an issuer's shareholding relation with Moody's. To fix idea, recall that the (unobserved) soft adjustment is represented by the relative conditional mean thresholds $\delta_{j,0}(r) \equiv E[T_{ji} - T_{0i}|R_i(\alpha_0) = r]$. To have a comparable measure of soft adjustment, we choose to normalize the baseline level, Aaa notch that is unlikely to be affected by the shareholding relation. In the appendix, we provide some practical and empirical support for choosing this base category.

Figure 3.3 depicts the empirical distributions of estimated individual-bond soft adjustment $\Delta_{j,0}(r)$ from the semiparametric model, with the dash (dotted) line indicating the period before(after) the Dodd-Frank. We only plot the distribution for the five ratings categories from Aa to B, as the adjustment for Aaa category is already normalized to zero.

There are three main implications of the results. First, the dispersion of soft adjustment decreases substantially after the Dodd-Frank as the empirical distribution becomes more concentrated around the mean, especially for bonds with median level of credit worthiness. Since the soft adjustment reflects the CRA's private information, the decline of the soft adjustment plausibly suggests a more transparent rating methodology after the Dodd-Frank Act. Second, the distribution of the soft adjustment shifts in means after the Dodd-Frank. For investment grade bonds, the soft adjustment, on average, shifts towards the right, implying more stringent rating criteria. Put differently, receiving an investment grade has become more difficult after 2010 for observationally identical bonds. For speculative bonds of Ba or below, the mean thresholds have become smaller, indicating more relaxed criteria of the CRA. The estimated densities also exhibit the apparent tri-modal feature, especially for the investment grade bonds, which contradicts the conventional parametric assumption that the unobserved rating adjustment is normally distributed.[30].

In broad, our estimation results on the soft adjustment suggest that the CRAs took steps to tighten their models and limit the opportunity for conflicts of interest coming from the human element. The passage of Dodd-Frank may largely contribute to the reduction of soft adjustment. On top of the discipline effect of Dodd-Frank, CRAs could have voluntarily become more conservative and used fewer subjective adjustments after the public condemnation they faced in the crisis.

Figure 3.3: Empirical Distributions of Soft Adjustment before and after the Dodd-Frank Act



Note: 1. Sample periods. Before the Dodd-Frank: 2000-2010; After the Dodd-Frank: 2011-2016. 2. The soft adjustment is estimated as mean thresholds relative to the base level Aaa using the two-step semiparametric estimator.

Recall that the proposed semiparametric model allows $\delta_{j,0}$ to be driven by $R_i(\alpha_0)$, a bond issuer's relationship with Moody's, whereas the ordered-probit assumes that $\delta_{j,0}$ is a constant. In Table 3.7, we report estimation results from the two approaches to highlight the heterogeneity in threshold captured by the semiparametric model. The

[30]Estimating the distribution of soft adjustment has been difficult in the parametric context. Strict distributional assumptions, usually a normal random variable with unknown means, have to be imposed to permit estimation. However, as can be seen from Figure 3.3 in which the soft adjustments are estimated in a distribution-free manner, the normality assumption on soft adjustments is highly suspicious.

first two columns report relative thresholds from ordered probit/logit specifications,[31] and the third column from the single-index semiparametric model without assuming the error term distribution.[32] Noticeable differences between Semi-X and the two parametric models suggests that neither ordered probit nor logit correctly describes the underlying data. Turning to the proposed model Semi-R, we present estimated thresholds conditional on various percentiles of the control index. It can be inferred from the heterogeneous pattern of threshold that the extent of soft adjustment varies with the control index $R_i\alpha_0$. Comparing the standard errors of $\delta_{j,0}$ before and after the Dodd-Frank period, the smaller standard errors in the later period confirms the earlier finding that soft adjustment become less dispersed.

3.5.4 Patterns of Soft adjustment over Shareholder Relationship

Following our discussion on the threshold parameters $\Delta_{j,0}$, in this section we provide in-depth analysis on how exactly are soft adjustments driven by the shareholding relation. Recall that the shareholding relation $R_i(\alpha_0) \equiv \text{Mshare}_i + \alpha_1\text{Fshare}_i + \alpha_2\text{LargeSH}_i$. We assess the pattern of $\Delta_{j,0}(r)$ over $R_i(\alpha_0)$, having in mind that $\alpha_1 < 0$ and $\alpha_2 < 0$ before the Dodd-Frank and $\alpha_1 < 0$ and $\alpha_2 > 0$ after the Dodd-Frank[33].

In Figure 3.4, we plot the relationship between $R_i(\alpha_0)$ and the soft adjustment across different rating categories, with the left(right) panel indicating the period before(after) the Dodd-Frank. First, the soft adjustment and $R_i(\alpha_0)$ have a "U-shape" relationship before the Dodd-Frank. As the bond issuer builds a tighter relationship with common shareholders (a higher Fshare and/or LargeSH inducing a lower $R_i(\alpha_0)$)

[31]The base level is Aaa. Relative thresholds are defined as $(T_{ji} - T_{0i})/\widehat{\beta}_1$, where $\widehat{\beta}_1$ is the estimated log asset coefficient.

[32]The bootstrapped standard errors are presented in parentheses for the semiparametric models and we use the delta-method to compute them for Ologit and Oprobit similar to those of relative coefficients.

[33]In the online appendices, we further examine the empirical relationship between the soft adjustment and each of the shareholding measures: Mshare, Fshare and LargeSH, for various counterfactual scenarios. The patterns of $\Delta_{j,0}$ over all three measures are roughly the same, albeit to minor disparity, as if we only use the aggregate the measure $R_i(\alpha_0)$. Detailed discussion can be found in the online appendices.

Table 3.7: Estimation Results of Soft Adjustment ($\widehat{\delta}$) at Control Index Percentiles

	Oprobit	Ologit	Semi-X	Semi-R								
				.1	.2	.3	.4	.5	.6	.7	.8	.9
Case 1: 2000-2010 Before Dodd-Frank												
$\widehat{\delta}_{0,1}(R)$	-4.281	-4.137	-6.863	-4.674	-8.813	-8.504	-7.236	-6.958	-6.982	-7.355	-7.258	-7.691
	(0.323)	(0.322)	(1.333)	(1.135)	(1.232)	(1.198)	(0.999)	(0.977)	(1.032)	(1.209)	(1.423)	(1.620)
$\widehat{\delta}_{0,2}(R)$	-7.453	-7.053	-8.340	-5.629	-12.466	-13.588	-11.632	-10.004	-9.073	-7.545	-6.520	-6.686
	(0.431)	(0.416)	(1.407)	(1.298)	(1.712)	(2.131)	(2.283)	(1.671)	(1.467)	(1.519)	(1.550)	(2.187)
$\widehat{\delta}_{0,3}(R)$	-10.766	-10.145	-10.978	-7.352	-14.842	-15.938	-13.779	-12.064	-10.601	-8.485	-6.754	-9.483
	(0.535)	(0.507)	(1.607)	(1.498)	(1.852)	(2.362)	(2.418)	(1.921)	(1.959)	(2.028)	(2.310)	(2.980)
$\widehat{\delta}_{0,4}(R)$	-12.601	-11.860	-12.411	-8.536	-17.543	-17.422	-15.502	-13.386	-10.845	-7.498	-6.007	-11.544
	(0.592)	(0.557)	(1.793)	(2.003)	(2.176)	(2.593)	(2.560)	(2.422)	(2.972)	(3.293)	(4.044)	(3.256)
$\widehat{\delta}_{0,5}(R)$	-16.676	-16.082	-14.627	-11.387	-20.567	-21.608	-18.938	-15.581	-10.341	-7.626	-4.369	-13.453
	(0.719)	(0.681)	(2.471)	(4.462)	(4.273)	(4.283)	(3.827)	(3.885)	(4.540)	(4.460)	(5.640)	(3.489)
Case 2: 2011-2016 After Dodd-Frank												
$\widehat{\delta}_{0,1}(R)$	-2.667	-2.734	-3.676	-3.874	-3.550	-3.541	-3.909	-4.552	-5.441	-4.544	-3.268	-2.637
	(0.231)	(0.252)	(0.856)	(0.277)	(0.272)	(0.327)	(0.421)	(0.711)	(0.577)	(0.414)	(0.493)	(0.650)
$\widehat{\delta}_{0,2}(R)$	-6.372	-6.334	-5.333	-5.723	-5.460	-5.678	-5.968	-6.529	-7.355	-6.433	-4.929	-4.204
	(0.351)	(0.371)	(0.861)	(0.267)	(0.277)	(0.401)	(0.526)	(0.775)	(0.615)	(0.425)	(0.514)	(0.724)
$\widehat{\delta}_{0,3}(R)$	-11.165	-11.047	-8.882	-9.992	-9.486	-9.645	-9.866	-10.294	-10.987	-10.027	-8.790	-8.402
	(0.496)	(0.514)	(0.879)	(0.317)	(0.319)	(0.453)	(0.550)	(0.802)	(0.674)	(0.478)	(0.557)	(0.762)
$\widehat{\delta}_{0,4}(R)$	-13.037	-12.888	-12.339	-17.919	-16.441	-15.341	-14.015	-14.473	-15.208	-13.837	-11.586	-10.280
	(0.552)	(0.569)	(0.901)	(0.609)	(0.773)	(0.849)	(0.681)	(0.893)	(0.729)	(0.538)	(0.612)	(0.773)
$\widehat{\delta}_{0,5}(R)$	-17.775	-17.817	-15.064	-22.847	-20.972	-19.572	-17.944	-17.917	-18.174	-16.381	-14.128	-13.094
	(0.700)	(0.726)	(0.878)	(0.587)	(0.751)	(0.793)	(0.597)	(0.837)	(0.703)	(0.499)	(0.604)	(0.780)

Note: 1. Each soft adjustment $\widehat{\delta}(\cdot)$ is evaluated at the percentile of R or $\widehat{L}$. For instance, .1 denotes 10 percentile. 2. Scaled thresholds estimates are computed relative to the base level of $Y = 0$, indicating the Aaa notch. 3. Semi-X refers to semiparametric estimation without controlling soft information and Semi-R does the control. 4. In parenthesis are the standard errors. Those of Oprobit and Ologit are computed using first order Delta approximation. Those of semiparametric models are bootstrapped s.e.(s) with 50 times of draws.

Moody's starts to relax its rating criteria: it is easier for firms with a stronger connection to receive higher ratings. Possibly due to the worry of conflicts of interest, Moody's starts to tighten its rating criteria when the relationship gets too strong. This pattern is nearly uniform for all rating categories before the Dodd-Frank, with Aa being the only exception. Interestingly, for the period after the Dodd-Frank, this pattern has changed completely: as $R_i(\alpha_0)$ strengthens, Moody's uniformly tightens the rating criteria.[34] Given that largeSH has a positive impact of $R_i(\alpha_0)$ in this period(e.g. $\alpha_2 > 0$), we conclude that Moody's has become more stringent on issuers related with its large shareholders after the Dodd-Frank. This change of pattern could be a result of rating agency's greater concern of conflicts of interest.

Another sharp contrast between the two panels in Figure 3.4 is that threshold parameters for different rating categories converge to each other before the Dodd-Frank, but not after. Recall that those parameters partition the latent default risk into different rating categories. It can be seen from the left panel that when the relationship index takes value between -10 and 0 (about 70-80 percentile), threshold parameters for different rating categories become indistinguishable. In terms of rating criteria, this indicates Moody's does not have a clear criteria to separate the safe bonds from the bad bonds issued by highly connected firms. Instead, actual rating assignments must involve discretion. This pattern also disappear after the Dodd-Frank, as the threshold parameter maintain an ordered relationship throughout.

3.5.5 Discussion

The time variation in the soft adjustment that we captured plausibly suggests the effectiveness of Dodd-Frank in improving the transparency of credit rating procedures. In particular, we have witnessed that Moody's soft adjustments have less explanatory

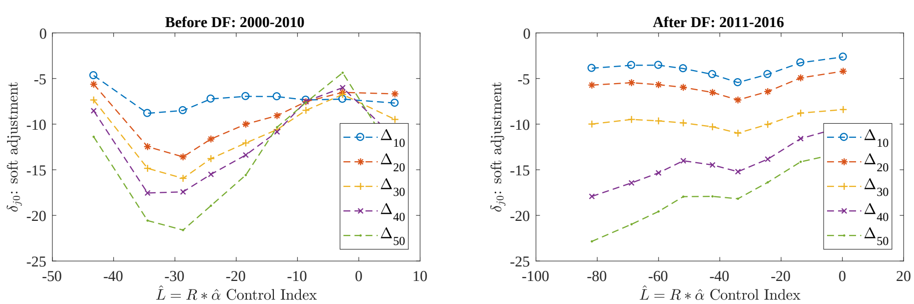
[34] Since in the post Dodd-Frank period, largeSH and Fshare have opposite impact on $R_i(\alpha_0)$, it is unclear whether a increasing $R_i(\alpha_0)$ represents a stronger shareholding relation. Therefore we provide more analysis on the impact of individual variable on soft adjustment later.

power on the distribution of credit ratings. In addition, Moody's has also become more stringent when it comes to rating bonds issued by closely connected firms, as the threshold parameters increase monotonically as the relationship index strengthens.

Both findings significantly contribute to the extant literature on the time series pattern of credit rating criteria. Researchers have studied this characteristic of rating agencies, and many have already found that criteria tighten and loosen over time (examples include Alp (2013); Jorion *et al.* (2009); Griffin and Tang (2012). Our proposed measure of soft adjustment can also be used to quantify such rating criteria variation across issuers and over time, which cannot be simply achieved with conventional parametric models. Moreover, we connect CRA's changing criteria to conflicts of interest with shareholders by estimating the pattern of soft adjustment conditional on various level of CRA-issuer connectedness.

We admit that there are other confounding factors that might lead to the structural change in CRA's rating methodology, such as the CRA's increasing awareness of reputation concerns as well as the emergence of CDS markets. Instead of making a causal statement that the passage of Dodd-Frank is the root cause of such changes, we resort to a weaker conclusion that the tightening regulatory requirement is one but very plausible explanation and leave the causal inference to future work.

Figure 3.4: Estimated Relationship between Soft Adjustment and Shareholding Control Index



Note: 1. Left panel: before the Dodd-Frank; right panel: after the Dodd-Frank. 2. Y-axis plots the estimated soft adjustment as conditional mean thresholds relative to Aaa level. 3. X-axis plots various percentiles of estimated control index for shareholding relations.

3.6 Conclusions

This paper considers the role of soft adjustments when estimating a structural corporate bond rating model used by CRAs and empirically assesses the extent a reduction in the conflict-of-interest after the enactment of the Dodd-Frank Act, specific to the credit rating industry. From an empirical point of view, the presence of hidden soft adjustments could cause endogenous determination of firm and bond characteristics, resulting in inconsistent default risk index estimates and incorrect rating probability functions. To resolve this identification issue, we model the soft adjustment as the bond-specific stochastic thresholds in a fully nonparametric way and approximate it using the shareholding structures between a bond-issuer and a publicly listed CRA. The empirical method in this paper can be applied to other contexts featuring ordered response with unobserved heterogeneous thresholds such as subjective health status, life happiness, etc. For empirics, we focus on initial bond ratings of listed firms after the Moody's went public in 2000 and until 2016, covering a fews year after the passage of the Dodd-Frank Act. Our empirical results suggest that there is a significant reduction of soft adjustment over all rating categories in terms of dispersion after the reform. We also find that it becomes more difficult to be rated as investment grade bonds on average, reflected by the shift of mean thresholds.

Our model does not consider the competition effect in the rating process, especially given the "triopoly" market structure of the U.S. credit rating industry. Under the current issuer-pays model, the CRA charge fees to issuers whose debt will be rated by the same agency. This would produce another aspect of conflict-of-interest where CRAs attempt to make more profit and attract more clients by lowering standards. However, our measure of soft adjustment is not designed to capture such bias. Future studies may consider explicitly incorporating strategic rating behaviors among CRAs.

Chapter 4

A Hausman Test for Partially Linear Models with an Application to Implied Volatility Surface

4.1 Introduction

One key assumption behind the well known Black-Scholes (B-S) formula is a constant volatility function, which has been frequently challenged by both theorists and practitioners after the October 1987 market crash. The extent to which the market deviates from this assumption can be tested by examining the implied volatilities (IV), which by definition is the empirically determined parameter that makes the B-S formula fit market prices of the options. To see this, we denote the price of a call option at time t as C_t, its strike price as K, its time to expiration as T, and the (fair) forward price for delivery at expiration as $F_t \equiv e^{(r-\delta)T}S_t$. As shown in Aıt-Sahalia, Bickle and Stoker (2001, ABS henceforth), inverting the B-S formula with respect to the volatility parameter would give to the following model for IVs (σ_{iv}),

$$\sigma_{iv} = m(K/F_t, T) + \epsilon \quad \text{with} \quad E[\epsilon | K/F_t, T] = 0 \tag{4.1}$$

where K/F_t is the "moneyness" of an option and ϵ summarizes potential sources of noise, e.g., bid-ask spread. The unknown transformation $m(\cdot,\cdot)$ captures the dependency of IV on K/F_t and T. If options are indeed quoted based on the B-S formula, one should expect a constant IV which does not vary across moneyness or time-to-expiration. However, out-of-the-money (OTM) put options, i.e., put options with $K/F_t < 1$, are traded at higher implied volatility than at-the-money (ATM) options and OTM calls, also known as “volatility smiles”.

ABS further show that a semiparametric model permitting a flexible “volatility smile” as well as an additive quadratic time effect, i.e., $m(K/F_t, T) = g(K/F_t) + \theta_1 T + \theta_2 T^2$, is a statistically adequate depiction of the IV data. The above partially linear specification, however, rules out potential interaction effect between moneyness and time-to-expiration. If we plot IV against moneyness and time-to-expiration in 3-D, which gives the so-called *Implied Volatility surface*, a partially linear structure implies that the term structures of IV across different moneyness values should roughly have the same shape and only differ by a level shift. However, as shown in Fig 4.1, which is taken from Fengler (2006) and confirmed by many other studies, this is not the case: there is a slightly increasing slope for ATM and OTM call IV term structure, while OTM put IV displays a decreasing term structure. This paper is primarily motivated by these conflicting findings.

Researchers have tried at theoretically explaining these stylized facts of IV term structure using stochastic volatility (Renault and Touzi, 1996; Hull and White, 1987) and jump diffusion models (Jorion, 1988; Bates, 1996). However, there is an absence of prior work that formally tests whether options across different moneyness value have distinct term-structure. To be specific, we will develop a Hausman type specification test for the partially linear structure in ABS against a semiparametric model that permits interaction effects, based on the observation that the two estimates should drift apart if the partially linear structure does not hold. In Section 2, we describe the testing strategy and construct the Hausman statistic in a general econometric setting. We carry out some Monte Carlo experiments to study the finite sample properties of the proposed test statistic in Section 3, and report the empirical results in Section 4 using traded option data of S&P 500 index after the recent crisis. While the focus of this paper is on option implied volatility, the method can be applied to test additivity of other econometric models.

Figure 4.1: Fig 2.8 in Fengler (2006)

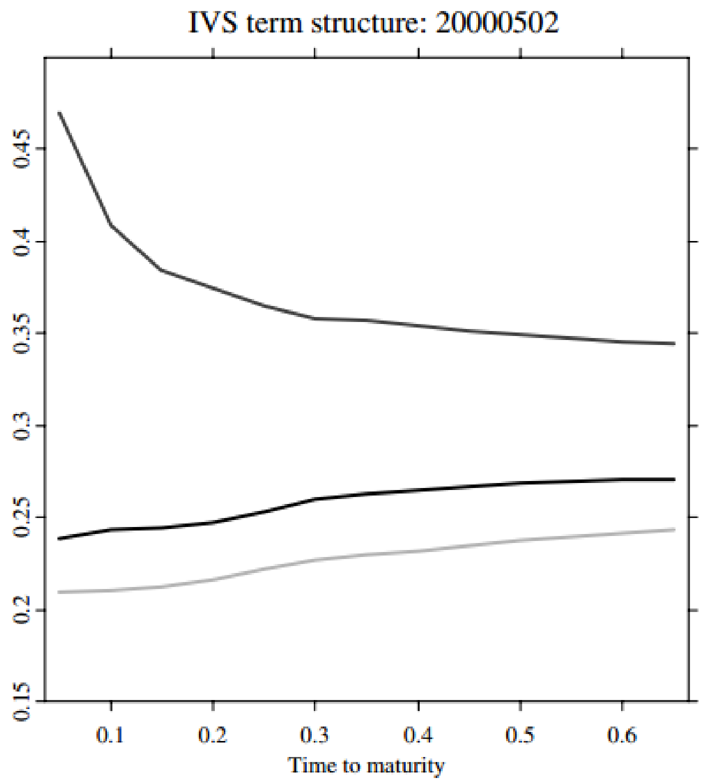


Fig. 2.8. *Top panel*: DAX option IVS on 20000502. IV observations are displayed as black dots; the surface estimate is obtained from a local quadratic estimator with localized bandwidths. *Bottom panel*: term structure of the IVS. $\kappa_f = 0.75$ top line, $\kappa_f = 1$, i.e. ATM, middle line, $\kappa_f = 1.1$, bottom line SCMivsts.xpl

4.2 A Hausman-type Specification Test

Using the notation defined in (4.1), we test

$$\mathbb{H}_0 : m(K/F_t, T) = g(K/F_t) + \beta_1 T + \beta_2 T^2 \tag{4.2}$$

against a two-index model which permits the interaction between moneyness and time-to-maturity:

$$\mathbb{H}_a : m(K/F_t, T) = H(\beta_1 T + \beta_2 T^2, K/F_t) \tag{4.3}$$

The motivation of this test comes from Hausman (1978). I find estimator $\hat{\beta}_A$ for $\beta_0 \equiv [\beta_1, \beta_2]$ that is consistent and efficient only under the null, and another estimator $\hat{\beta}_B$ that is robust to the additive structure that we want to test. If the data is generated from the null DGP, then the two estimators should be close to each other. We use the following notation $Y_i \equiv \sigma_{iv}$, $X_i \equiv [T, T^2]$, $Z_i \equiv K/F_t$ to illustrate our testing strategy, and this test can certainly be applied to other contexts.

Remark 1. For this particular application in which X's contains only T and T^2, we need the assumption that T^2 belongs to the model, that is, $Pr(\beta_2 \neq 0) \rightarrow 1$, to identify parameters in the two-index model. This assumption is plausible since the term structure of volatility is rarely found to be flat. In other contexts that the information of X's cannot be summarized in a single variable, this assumption is not needed.

One candidate for $\hat{\beta}_A$ comes from the "Robinson differencing" procedure: (1) take conditional expectation of Z_i on both sizes e.g, $E[Y_i|Z_i] = E[X_i|Z_i]\beta_0 + G(Z_i)$. (2) subtract the conditional expectations from Eq (4.2), e.g, $Y_i - E[Y_i|Z_i] = (X_i - E[X_i|Z_i])\beta_0 + U_i$, and (3) run OLS. Robinson (1988) shows that the final OLS estimator is root-N-consistent and asymptotically efficient. One candidate for $\hat{\beta}_B$ can be obtained by implementing the semiparametric least square (SLS) estimation on (4.3). Ichimura (1993) and Ichimura and Lee (1991) develop a consistent estimator $\hat{\theta}_B$ for the "normalized parameter" $\theta_0 \equiv \frac{\beta_2}{\beta_1}$. Klein and Shen (2010) employ a two-stage estimator, so that the conditional expectation $E[Y|X\beta_0, Z]$ can be estimated with optimal kernel bandwidth.

Since the results from Robinson (1988) and Klein and Shen (2010) are crucial to the construction of our test statistic, we give these results in the following two propositions[1].

Proposition 4.2.1. *Assume (Y_i, X_i, Z_i) are i.i.d, $g(\cdot)$ satisfy certain differentiability*

---

[1]Some trimming parameters have been suppressed in order to facilitate the exposition, and readers should refer the original papers for full details. Since both methods are widely used in estimating non/semi-parametric models, the proofs are omitted.

and moment condition, as in Robinson (1988), we have

$$\sqrt{N}(\hat{\beta}_A - \beta_0) \sim N(0, \Sigma_A)$$

where $\Sigma_A \equiv \sigma^2 E[(X_i - E[X_i|Z_i])'(X_i - E[X_i|Z_i])]^{-1}$ *and* σ^2 *is the variance of U.*

Proposition 4.2.2. *Under the alternative and additional assumptions as in (A1-A6) in Klein and Shen (2010), let* $\hat{\theta}_B$ *be the maximizer of the following quasi-likelihood function,*

$$\hat{Q}_2 \equiv -\frac{1}{2N}\sum_{i=1}^{N}\hat{t}_{v_i}[Y_i - \hat{E}_a(Y_i|X_i\theta, Z)]^2$$

in which $\hat{t}_{v_i}$ *is a trimming function on the basis of the (estimated) index and* $\hat{E}_a$ *is an adjusted expectation that protects the estimated denominator away from zero in a small neighborhood of the true* θ*. Letting* $G_0 \equiv \nabla_\theta Q_2(\theta_0)$ *and* $H_0 \equiv \nabla_{\theta'\theta} Q_2(\theta_0)$*,*

$$\sqrt{N}(\hat{\theta}_B - \theta_0) \sim N(0, \Sigma)$$

where $\Sigma \equiv H_0^{-1} E[\sqrt{N} G_0' G_0 \sqrt{N}] H_0^{-1}$.

Note that $\hat{\beta}_A$ and $\hat{\theta}_B$ are not directly comparable since the SLS procedure does not identify the regression coefficients β_0, making $dim(\hat{\beta}_A) = dim(\hat{\theta}_B) + 1$. To conduct a feasible Hausman-type test, we compute the "normalized parameter" in the partial linear model as $\hat{\theta}_A \equiv \frac{\hat{\beta}_{A2}}{\hat{\beta}_{A1}}$, and apply the "delta method" to calculate its asymptotic variance:

$$VAR(\sqrt{N}\hat{\theta}_A) = \nabla G(\hat{\beta}_A)^T \cdot \Sigma_A \cdot \nabla G(\hat{\beta}_A) \quad (4.4)$$

$$\nabla G(\hat{\beta}_A) = \begin{bmatrix} -\frac{\hat{\beta}_{A2}}{\hat{\beta}_{A1}^2} \\ \frac{1}{\hat{\beta}_{A1}} \end{bmatrix} \quad (4.5)$$

where Σ_A is the variance-covariance matrix of $\sqrt{N}\hat{\beta}_A$ in Proposition 1. The gradient matrix $\nabla G(\hat{\beta}_A)$ comes from first-order Taylor approximation.

With the two estimators formulated above, a Hausman-type test statistic comes naturally:

$$T_n \equiv \sqrt{N}(\hat{\theta}_A - \hat{\theta}_B)' M^{-1}(\hat{\theta}_A - \hat{\theta}_B)\sqrt{N} \tag{4.6}$$

where the appropriate scale matrix M is

$$M \equiv VAR(\sqrt{N}\hat{\theta}_B) - VAR(\sqrt{N}\hat{\theta}_A) \tag{4.7}$$

The first component in M, $VAR(\sqrt{N}\hat{\theta}_B)$, as shown in Klein and Shen (2010), converges to $\Sigma \equiv H_0^{-1}E[\sqrt{N}G_0' G_0\sqrt{N}]H_0^{-1}$. The second component $VAR(\sqrt{N}\hat{\theta}_A)$ can be estimated consistently using the aforementioned Delta-method. Under the null, T_n follows a χ_1^2 distribution.

4.3 Monte Carlo Experiments

In this section we carry out some Monte Carlo experiments to study the finite sample properties of the test statistic. The DGP is given as:

$$Y \;=\; \beta_1 X_1 + \beta_2 X_2 + \sqrt{Z} + \delta * (\beta_1 X_1 + \beta_2 X_2) * Z + U \quad \beta_1, \beta_2 = 1 \tag{4.8}$$

The data is constructed by generating $X_1, X_2 \sim \chi^2(1)$, $Z = X_1 + X_3 + 3$, $X_3 \sim \chi^2(1)$ and $U \sim N(0,1)$. Some truncations are applied to ensure X and Z are finite. When $\delta = 0$, the model is partial linear. As δ increases, the model smoothly transforms from a series of local alternatives to a two-index model in which Z and $X\beta$ have full interaction.

In Table 4.1 we report the Monte Carlo results from 500 replications with $\delta = 0, 0.1, 0.2, 0.5$. The parameter of interest here is the ratio of regression coefficients of X_2 and X_1: $\theta = \beta_2/\beta_1 = 1$. In the first design with $\delta = 0$, the Robinson differencing estimator $\hat{\theta}_A$ performs better than the two-index estimator $\hat{\theta}_B$ in terms of accuracy (MEAN of estimates with N=2000: 1.003 vs 1.045) and efficiency (Root-Mean-Square-Error (RMSE) with N = 2000: 0.006 vs 0.008). As δ deviates from zero, the 2-index models still provides robust estimation of θ, while the Robinson differencing estimator became inconsistent. For example, when $\delta = 0.5$ and N = 1000, the mean of $\hat{\theta}_A$ is 0.863 while the mean of $\hat{\theta}_B$ is 0.997.

The testing results are reported in the lower panel of Table 4.1. Under the DGP in H_0 ($\delta = 0$) with N=2000, the rejection rate in 500 replication is 3.2% and 8.2 %, given the theoretical size is 5% and 10% respectively. When the true model deviates significantly from the partial linear model ($\delta = 0.5$), the test has powers close to one in both samples. As the interaction term becomes smaller, that is, when the true model approaches a series of local alternatives, the power of the test decreases. For each alternative DGP with a different δ, the power increases as the sample size increases.

4.4 Empirical Results

As motivated in the introduction, we study whether an option's moneyness and time-to-maturity affect its implied volatility in an interactive fashion, e.g., do options across different moneyness value have different term structure? ABS address this question by testing whether the in-sample-fit of $m(K/F_t, T) = g(K/F_t) + \theta_1 T + \theta_2 T^2$ is statistically the same as an unrestricted $m(K/F_t, T)$. We revisit this problem with a more recent dataset using the testing strategy developed in this paper.

Our data sample consists of N = 4431 observations on daily S&P 500 index call options traded at the Chicago Board Options Exchange (CBOE) from September 2012-August 2013. The options are European, and within the market we have chosen the most actively traded options with maturities from 1 to 9 months. We compute the option premium using the midpoint of bid and ask price, and solve for option implied volatility based on the BS formula. Following the "Robinson differencing" and Semiparametric Least Square procedures, we estimate the two competing models

Table 4.1: Estimation results and rejection rate of H_0 in 500 replications

	$\delta = 0$			δ= 0.1			$\delta = 0.2$			$\delta = 0.5$		
	Mean	STDC	RMSR	Mean	STDC	RMSR	Mean	STDC	RMSR	Mean	STDC	RMSR
N=1000												
P-Linear	1.012	0.110	0.012	0.950	0.079	0.009	0.898	0.057	0.014	0.863	0.040	0.020
2-index	1.070	0.112	0.017	1.038	0.084	0.008	1.011	0.061	0.004	0.997	0.040	0.002
N=2000												
P-Linear	1.003	0.075	0.006	0.947	0.058	0.006	0.905	0.040	0.011	0.866	0.029	0.019
2-index	1.045	0.078	0.008	1.025	0.060	0.004	1.011	0.040	0.002	1.001	0.030	0.001

Rejection rate of $H_0 : Y = X_i\beta_0 + G(Z_i) + U_i$

	Size			Power								
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
N=1000	0.006	0.038	0.068	0.014	0.094	0.192	0.150	0.374	0.532	0.718	0.898	0.938
N=2000	0.002	0.032	0.082	0.068	0.170	0.292	0.380	0.705	0.822	0.978	0.996	0.998

Note 1: In the upper panel, I compare the estimation results from both the partial linear model and the two-index model with varying δ's. For each design, I run 500 replications with N = 1000 and 2000. MEAN, STDC and RMSR refer to the mean, standard deviation and root-mean-square-error of the estimates respectively.

Note 2: In the lower panel, I report the rejection rate in 500 replications in each designs. The critical value are set at 1%, 5% and 10% significant levels. When $\delta = 0$, the rejection rate should be close to the theoretical size. When δ significantly deviates from zero, the test should exhibit a theoretical power of 1.

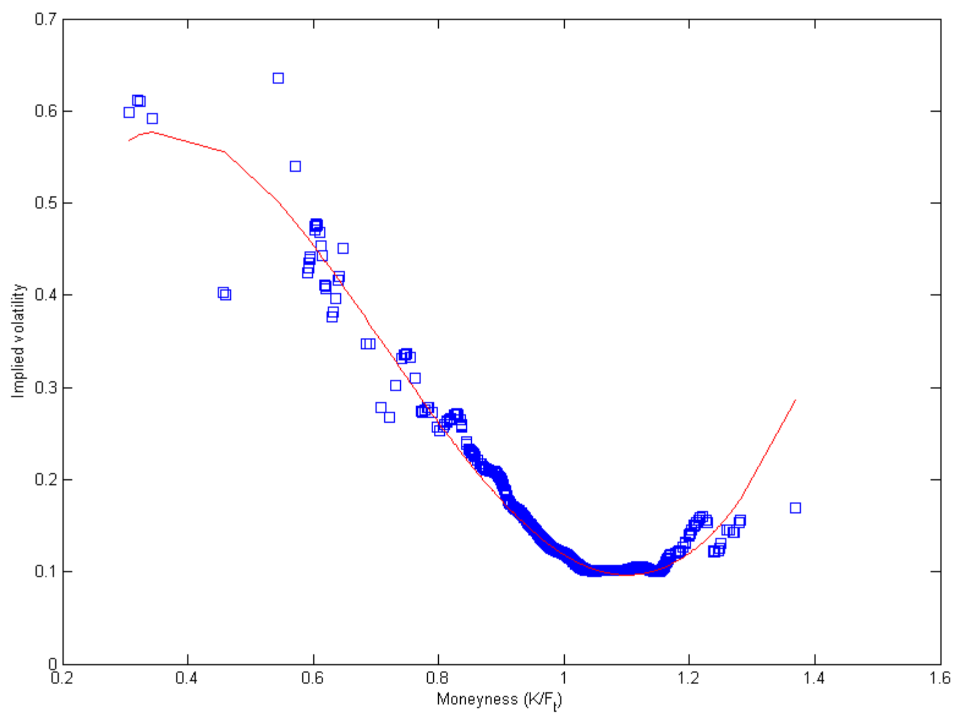
as:

$$\textit{Plinear:} \qquad \hat{m}(K/F_t, T) = \hat{g}(K/F_t) + \underset{(0.0067)}{0.111}\ T - \underset{(0.0098)}{0.089}\ T^2 \tag{4.9}$$

$$\textit{2-index:} \qquad \hat{m}(K/F_t, T) = \hat{H}(T - \underset{(0.138)}{0.663}\, T^2, K/F_t) \tag{4.10}$$

To illustrate the nonparametric components in these models, we plot $\hat{g}(\cdot)$ and $\hat{H}(\cdot,\cdot)$

Figure 4.2: "volatility smile" in partially linear model

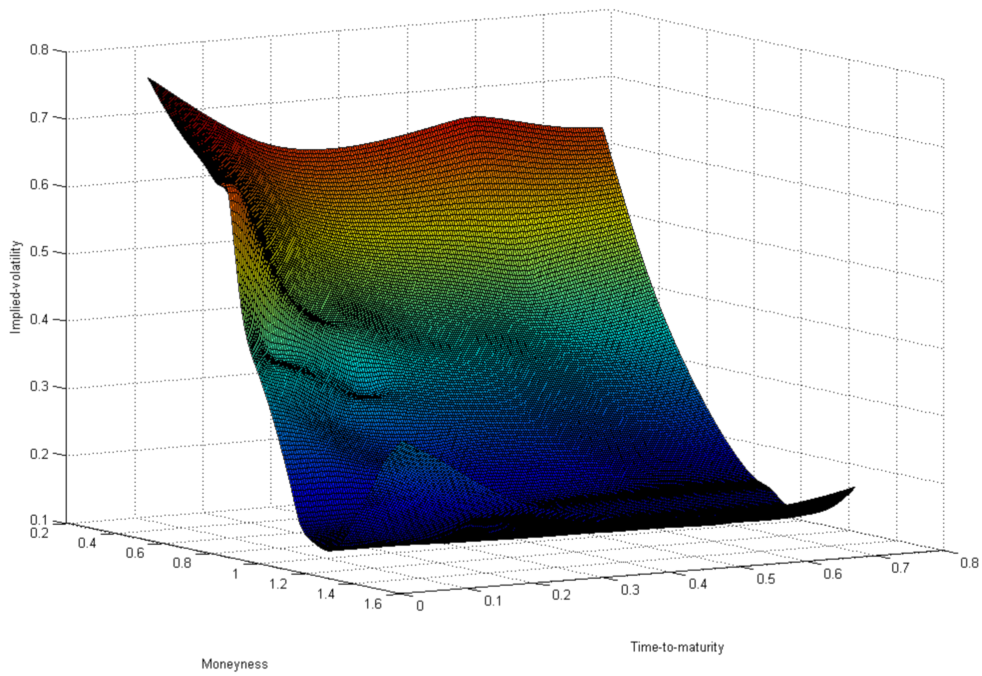


Note: Blue squares are data points in the sample and the red line is a fitted curve using spline interpolation

in Figure 4.2 and 4.3. As can be seen from Figure 4.3, the term structure of volatility is generally downward-sloping for options across different moneyness, which evidents against the hypothesis that time-to-maturity and moneyness affects implied volatility in a interactive fashion. To be sure, we compute the test statistic T_n based on (4.6), which turns out to be 1.18, leading to the acceptance of the partially linear model with a p-value equals 0.27.

This finding is in accord with ABS, but conflicts with Fengler (2006) and other

Figure 4.3: implied volatility surface in two index model



Note: This graph depicts the interpolated implied volatility surface (IVS). The time-to-maturity axis describes the term structure for options with different moneyness values (K/F_t), while the moneyness axis describes the shape of "volatility smile" for options with different time-to-maturity.

conventional thoughts claiming that ATM and OTM options have an upward sloping term structure. Fengler (2006) attributes this pattern to a more 'shallow" smile for longer term maturities: a higher implied volatility for ATM and OTM call options with longer time-to-maturity. Considering the timing in Figure 4.1, such a pattern may be related to the Dot-com bubble during the late 90s. A persistent rapid rise in equity values makes OTM call options more appealing to speculators, as they may have intrinsic values. Such a growing demand for those options bids up the premium, and thus induces a higher IV for long term calls. Considering the stable market conditions[2] in our sampling period, investors may more concern about the short-run risk, making

[2]S&P 500 index steadily grew in the second half of 2012 and the level of VIX was around 12-20

the term structure of implied volatility downward-sloping.

4.5 Conclusion

Through a kernel-based goodness-of-fit test, ABS documented that a partially linear model permitting a flexible ”volatility smile” and an additive quadratic time effect is a statistically adequate depiction of the option implied volatility data. This paper develops an alternative specification test based on the shape of implied volatility surface at different ”moneyness” values. Our test statistic has a conventional Hausman form and can be applied to test additivity of other econometric models.

Chapter 5

Conclusion

Semiparametric models have received tremendous attention in the literature for decades. Much of the empirical papers stick to the single-index model, where the explanatory variables affect the outcome through an unknown linear combination. However, in many contexts, it is necessary to deviate from that specification. For example,

1. Behavioral models with non-separable utilities could suggest explanatory affect the outcome variables in a interactive fashion.

2. In models where the error terms are naturally heteroscadasticiy, it is convenient to have one index drives the conditional expectation and another index drives the conditional variance

3. In decision making problems involves multiple players, it is also natural to use a separate index to capture each player's utility.

Set against this background, the goal of this dissertation to develop econometric framework for estimating and hypothesis testing models that allow for multiple indices. One key methodological challenge in this type of problem is to address the bias in estimating conditional expectation in high dimensions. Higher-order kernels can reduce the bias to any order but do not perform well when outcome variables are discrete. A recursive-differencing estimator, recently proposed by Klein and Shen (2016), perform much more stable; yet the asymptotic normality result is not explicitly proved for the multiple-index case. The authors assert that a component that representing the asymptotic bias can be written in a higher-order degenerate U-statistic and therefore vanishes. I show this is indeed the case.

This dissertation also contributes to the empirical literature of credit rating and implied volatility modeling. Researchers (Mathis *et al.*, 2009; Jiang *et al.*, 2012; He *et al.*, 2015) have concerned with rating qualities on account of the issuers-paid model, whereby CRAs are paid by the issuers seeking ratings. Recently, several studies also document a "premium" on ratings for firms that share a particular form of non-rating relation with the rating agency, such as consulting services (Baghai and Becker, 2016) and rating-based contracts (Kraft, 2015). To the best of my knowledge, the only other paper that focuses on the impact of common ownership on rating is Kedia *et al.* (2017), in which the authors find that Moody's, the leading CRA in the U.S., assigned favorable ratings to firms that are associated with its long term large shareholders.

Much of the empirical evidence presented above, however, have been based on generalized linear models (GLMs); see Ederington (1985) for a survey of prevalent bond rating models[1]. These methods, however, leverage strong assumptions not only on the linearity of covariates but also the additivity of unobservables: that, say, rating agency's private information does not depend on or correlated with observed characteristics of the bond issuer. In this context in which common shareholders may transmit material non-public information to the CRAs, the mechanical interaction between publicly-available and private information makes GLMs prone to systematic bias.

I contribute to the above literature by evaluating rating quality using a semiparametric ordered model. Compared to existing models in the bond rating literature, the semiparametric model proposed here allows a richer set of interactions among covariates. In summary, I find that Moody's is likely to assign favorable ratings to firms that have a strong interaction with Moody's large shareholders. Based

[1]The extant literature use ordered-probit/logit model (Kaplan and Urwitz, 1979; Blume *et al.*, 1998; West, 1970), the linear probability model for the rating process (Jiang *et al.*, 2012; Campbell and Taksler, 2003; Kedia *et al.*, 2017), or discriminant analysis (Pinches and Mingo, 1973, 1975)

on marginal effect analysis, I found being a large investee firm of Moody's large shareholder could increase the probability of receiving favorable treatment by as much as 14 %, meaning that, on average, one out of seven bonds issued by those firms received favorable treatment. However, we found Moody's does not assign favorable ratings to firms related to Moody's small shareholders. This "large shareholder bias" is in accord with the literature on the role of large shareholders in corporate governance. In addition, we found low credit bonds issued by any firms, regardless their ownership interaction with Moody's, will unlikely to be treated with favor, which also seems credible because overrating a low credit bond would incur a greater expected reputation loss than overrating a safe bond. The policy relevance of the findings in this paper is that when credit rating agencies are publicly held by diffuse owners, their ratings are still highly trustworthy.

Appendix A

Supplemental Materials to Semiparametric Estimation of a Credit Rating Model

A.1 A cheap-talk model for credit rating

In the environment that I consider, a credit rating agency (CRA) is asked to rate a bond. The CRA only has partial knowledge about the bond's default risk, but can seek advice from a "shared-owner" - typically a large financial institution who owns both the CRA and issuer firm equities. Due to a frequent and personal contact with the bond issuer, these institutional investors have private information[1] about the bond issuer, which they could reveal in meetings with the CRA by sending an message m. Because the interests of the two parties are not perfectly aligned, the shared-owner may intentionally offer biased advice; the CRA will also contemplate the informational content of m. The model is a stylized version of the cheap-talk model considered by Crawford and Sobel (1982), CS henceforth.

A.1.1 Model

To fix ideas, consider the rating process of a corporate bond in which two risk-neutral players are involved, a credit rating agency (CRA) and a biased shared owner who holds the stock of both the CRA and the bond issuer. The bond's default risk is determined by $\pi = V + U$, in which $V = X\beta_0$ represents the influence of hard

[1] Examples of such private information may include soft factors, such as the manager's abilities. These information are not reducible to numerical scores and therefore hard for the bond issuer to communicate directly with the CRA.

factors (The firm's asset, leverage ratio...etc), and U summarizes other "soft" factors such as the manager's skills and abilities. The CRA can figure out V with their rating methodology, but only knows U is draw from a uniform distribution between 0 and 1.

Due to a better information access, the shared owner observes a noisy signal about U in the form of $z = U + \epsilon$, where ϵ is a small disturbance term, and sends a message m to the CRA as an "advising device". Upon receiving the message, the CRA (the "receiver" of the message) chooses an action y to maximize:

$$U^R(y, \pi) \quad = \quad -(y - \pi)^2 \tag{A.1}$$

Given the action y chosen by the CRA, the shared owner (the "sender" of the message) gains the utility of:

$$U^S(y, \pi, b) \quad = \quad -(y - \pi + b)^2. \tag{A.2}$$

where $b > 0$ is a scalar "bias" parameter that measures how closely aligned the preferences of the two are. b represents the shareholder bias because the utility-maximizing action is π for the CRA but $\pi - b$ for the shared owner. That is, the shared owner intends to inflate the rating by b through exaggerated advice. All aspects of the game except the realization of U are common knowledge.

Remark 2. For the purpose of building intuitions and obtaining solutions in closed-form, the model will be solved assuming the above utility functions and $\epsilon = 0$: that, shared owner observes soft factors perfectly[2]. Predictions of this model, however, will hold so long as for $i = R, S : U^i_{11} < 0, U^i_{12} > 0$, where subscripts denote for partial derivatives.

---

[2]This so-called "uniform-quadratic" specification is employed by many studies in the strategic information transmission literature (Adams and Ferreira, 2007; Kamenica and Gentzkow, 2011) for its tractability. In the case that shareholders observe a noise signal, it can be shown that the equilibria has the same structure as described below, provided that the conditional distribution of soft information $F(\cdot|z')$ dominates $F(\cdot|z)$ in the first stochastic sense for $z' > z$. In our case that $z = U + \epsilon$, this is true.

A.1.2 Equilibrium

Following CS it is possible to show that the CRA's action in equilibrium is:

$$y^* \;=\; V + \sum_{i=1}^{N(b)} \frac{a_i(b) + a_{i+1}(b)}{2} 1\{a_i(b) \leqslant m < a_{i+1}(b)\} \tag{A.3}$$

where m is the message received from the shared owner and the breakpoints a_i is parametrized by

$$a_i = \frac{i}{N(b)} + 2bi(N(b) - i),\; i = 0, 1, \cdots, N(b),\;\; a_0 = 0, a_N = 1 \tag{A.4}$$

and $N(b)$, the number of information partition, is the smallest integer greater or equal to $-\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{2}{b}}$. On the other hand, the shared owner's advising rule $q(m|U)$ is uniform, supported on $[a_i, a_{i+1}]$ for $U \in (a_i, a_{i+1})$[3].

A.1.3 Implication

Looking at the equilibrium action in (A.3), the CRA can at most ascertains an interval (a_i, a_{i+1}) wherein the soft information U lies and conjectures that U to be the midpoint of that interval[4]. Intuitively the finer this information partition is, the more accurate the CRA can learn the soft information. In terms of the model, $N(b)$ represents the efficiency of the information transmission, which would decrease as the shareholder

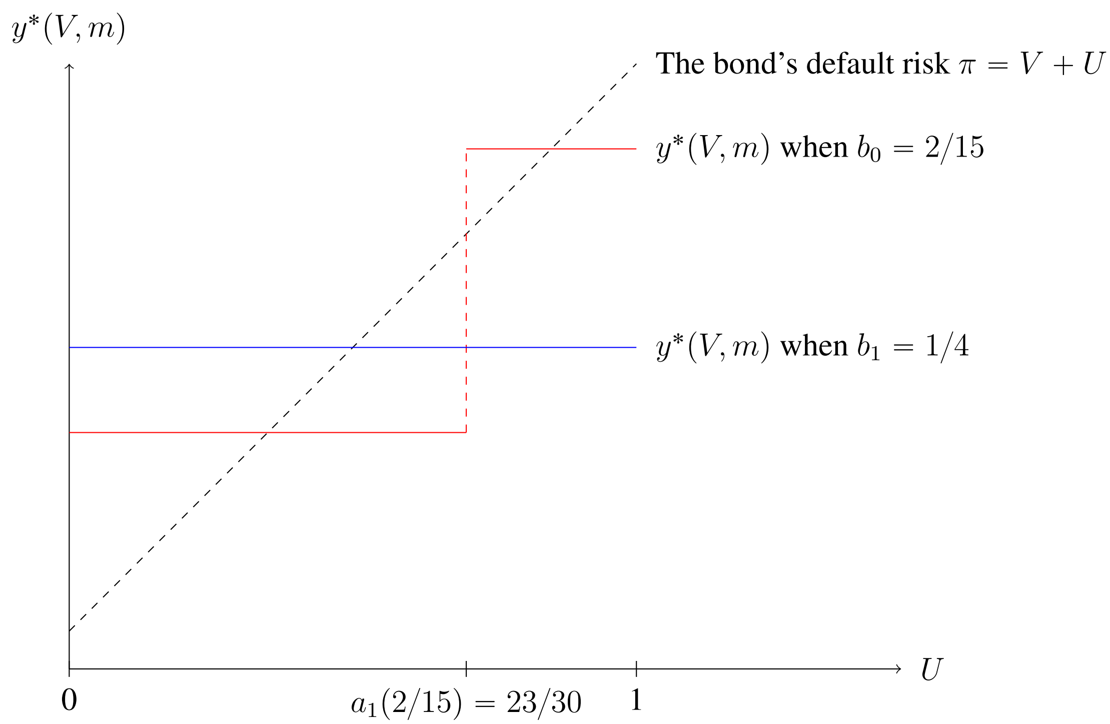
[3]In fact, CS shows that the model has multiple equilibira for every $1 \leqslant N \leqslant N(b)$. Here I focus exclusively on the *Most Informative Equilibrium*, that $N = N(b)$, because (i) for a given b, any other equilibrium with $N < N(b)$ is Pareto-inferior (Theorem 3 of Crawford and Sobel (1982)), and (ii) there are ample empirical evidence suggesting CRAs utilize information outside the issuer's financial reports to adjust their initial ratings. As noted in Kraft (2014), "soft adjustments" are frequently made on ratings to incorporate factors such as management quality, aggressive accounting, weak controls, governance risk, industry structure, and managerial bondholder friendliness. According to the model, one important information source for these soft adjustments is shared owners who have private information about the issuer's soft quality. It is foreseeable that ratings should have incorporate some information obtained from shared owners, as a result of the proposed information transmission mechanism.

[4]The CRA chooses the midpoint as a result of the assumption that U is uniformly distributed. Once this assumption is relaxed, the CRA will choose another action, but still within (a_i, a_{i+1}), to maximize its expected utility

bias increases (i.e., b is larger). In the extreme case when $b > 1/4$, communicating with the shared owner does not convey any meaningful information. To see this, it is easy to verify from (A.4) that with $a_0 = 0$, $a_N(b) = 1$, $N(b) = 1$ when $b > 1/4$. In this case, the CRA's strategy is to set $y^* = V + 1/2$ no matter what the realization of U is. Therefore the only equilibrium left is the "babbling equilibrium" in which no information is transmitted.

Importantly, the game-theoretical model predicts a nonlinear relationship between shareholder bias b and the estimated default risk y^*. From the equations (A.3) and (A.4), b affects y^* through two aspects: the set of cutoff points $a_i(b)$ and the discontinuous mapping $N(b)$. In regions where a increase in b does not change $N(b)$, a larger bias always induces a lower estimated default risk *ceteris paribus* (e.g., for fixed hard and soft information). This implication is consistent with the empirical observation that the CRA assigns more favorable ratings to firms that are associated with its own large shareholders (Kedia *et al.*, 2017). However, I show that such relationship is not monotonic *everywhere*: when a marginal increase in shareholder bias reduces the number of intervals $N(b)$, the net effect on ratings depends on the soft information U. Moreover, when the shareholder bias exceeds some threshold (in this case, 1/4), an increasing bias no longer affects the credit rating decision because the only equilibrium left is the "babbling equilibrium". That is, due to a high conflicts of interest, the CRA does not believe anything that the common shareholder say, so the common shareholder's (biased) advice has no impact on the rating outcome.

To convey more intuition, I present one simple example in which a larger bias indeed makes ratings more conservative (induces a lower credit rating). Consider the uniform quadratic case with b increases from $2/15$ to $1/4$. One can verify that when $b_0 = 2/15$, the only break point $a = 1/2 + 2 * 2/15 = 23/30$, resulting in a information partition $\{(0, 23/30), [23/30, 1]\}$. The optimal rating decision $y^*(V, m)$, depicted by the red solid line, is given by $V + 23/60$ when $U < 23/30$ and $V + 53/60$ when $U > 23/30$. When the bias increases to $b_1 = 1/4$, the game is in the babbling equilibrium with N=1, so the optimal rating function, depicted by the blue solid line, is constant $V + 1/2$. It is clear that a larger bias (from 2/15 to 1/4) induces a lower predicted default, and thus a higher credit rating, only when $U > 23/30$.



A.1.4 The equilibrium strategy beyond the uniform-quadratic model

In a more general model wherein the utilities are not quadratic and U is not uniformly distributed, it can be shown that the CRA's perceived default risk takes the form of $y^* = V + H(b, U)$, where b is a parameter that represents the common shareholder's bias. Based on the discussion above, the following theoretical claims about this function $H(\cdot, \cdot)$ can be made:

Prediction 1-$\mathcal{H}$. *$H(U, b)$ is a nonseparable.*

Prediction 2-$\mathcal{H}$. *The (marginal) impact of b on y^* is not globally monotone and* *affected by the based level b_0, so that a larger bias may lead to rating deflation.*

By analyzing this game-theoretical model, I demonstrate that the impact of conflicts of interest on ratings is not monotonic and depends on the private information U in an unobservable way. Therefore, in terms of econometric modeling, it is essential to take the potentially non-separability of b and U into account.

A.2 Econometric notations and preliminaries

To establish the large sample results in the next section, I require some standard assumptions and a more formal discussion about the conditional expectation estimator $P^k(v)$ and its econometric properties. For presentation simplicity, I use Z to denote the shared-ownership relation proxy $MFOI$.

A.2.1 Definitions and Notations

D.1 Trimming Functions: With W_{ik} as the i th observation on a continuous variable, W_k, k = 1,...K, Let $\hat{\tau}_{ik} \equiv \mathbb{1}\{\hat{a}_k < W_{ik} < \hat{b}_k\}$ and $\hat{\tau}_i = \prod_k \hat{\tau}_{ik}$, where $\hat{a}_k$ $\hat{b}_k$ are, respectively, lower and upper sample quantiles for W_k. when $W_{ik} = X_{ik}$, we refer to τ_{ix} as *X-trimming*; With $\hat{V}_i$ as the estimated index, when $W_{ik} = \hat{V}_i$, we refer to τ_{iv} as *index trimming*[5].

D.2 Kernels: Let v denotes a fixed point of the index and V_j the index value for observation j, define $K_j(v) = \Pi_{s=1}^{d}\phi(\frac{V_j - v}{h})/h$ where $\phi(\cdot)$ is the density of standard Gaussian and h is the window size that vanishes to zero at rate N^{-r}.[6]

[5] In case when a smooth trimming function is needed, define

$$\tau(z,\delta) \equiv (1 + exp(-(ln(N) * ln(N) * (z - \delta))))^{-1} \tag{A.5}$$

as a smoothed approximation to an indicator on $z \geqslant \delta$. A smoothed indicator on $z \in [a, b]$ is then defined as $\tau(z, a) * \tau(b, z)$.

[6] The choice of r will be different depending on the nature of the estimator, and will be discussed in the corresponding asymptotic theorem.

D.3 Density Estimator: Let $g(v)$ denotes the joint density of the indices at a point v, a (leave-one-out) kernel weighted density estimator $\hat{g}(v)$ is defined as

$$\hat{g}(v) = \frac{1}{N-1}\sum_{j=1}^{N} K_j(v) \tag{A.6}$$

D.4 Estimated Probability Referring to D.1, a *initial estimator* for $P^K(v)$ is defined as

$$\widehat{I}^K(v) \equiv \frac{\frac{1}{N-1}\sum_j Y_j^K K_j(v)}{\hat{g}(v)} \tag{A.7}$$

Based on this initial estimator at v and another sample point V_j of the indices, the *final estimator* for $P_i^K(v)$ is defined as

$$\widehat{P}^K(v) \equiv \frac{\sum_j [Y_j^K - \hat{\Delta}_j(v)]K_j(v)}{\hat{g}(v)} = \hat{f}(v)/\hat{g}(v) \tag{A.8}$$

where $\hat{\Delta}_j(v) = \widehat{I}^K(V_j) - \widehat{I}^K(v)$ is an estimate of the localization error. See section A.2.3 below for a detailed description for this estimator.

D.5 Quantile Trimming Define t_{qj} to be the true quantile trimming function which takes value 1 only if Z is in a given quantile: $t_{qj} \equiv 1\{q(\lambda_1) < Z < q(\lambda_2)\}$, where λs are the upper and lower bound for that quantile and $q(\lambda_1), q(\lambda_2)$ the corresponding population quantile for Z. Similarly, we define an estimated quantile trimming function as $\hat{t}_{qj} \equiv 1\{\hat{q}(\lambda_1) < Z < \hat{q}(\lambda_2)\}$ by replacing the population quantiles $q(\lambda) = (q(\lambda_1), q(\lambda_2))$ with the sample quantiles $(\hat{q}(\lambda_1), \hat{q}(\lambda_2))$.

D.6 Quantile Marginal Effects Estimators: Let $V(\hat{\theta}) \equiv [X_{F1} + X_F'\hat{\theta}_0^F, X_{B1} + X_B'\hat{\theta}_0^B, Z]$ be the estimated index. For an observation with $Y_i = K$, an estimator for the marginal effect $ME_i(\theta_0, Z_i, K)$ defined in the main text is:

$$\hat{m}_j(\hat{\theta}, K) \equiv \sum_{k=1}^{K-1} [\widehat{P}_i^k(V_F(\hat{\theta}), V_B(\hat{\theta}), Z + \delta) - \widehat{P}_i^k(V_F(\hat{\theta}), V_B(\hat{\theta}), Z)]$$

Referring to the above definition for quantile trimming t_q, we define the quantile marginal effect QME_q^K and its estimator by:

$$QME_q^K \equiv \frac{E[t_{qj} m_j(\theta_0, K)]}{E[t_{qj}]} \tag{A.9}$$

$$\widehat{QME}_q^K \equiv \frac{\sum_{j=1}^N \hat{t}_{qj} \hat{m}_j(\hat{\theta}, K)}{\sum_{j=1}^N \hat{t}_{qj}} \tag{A.10}$$

D.7 Bahadur Representation: Referring to the above definitions, with $g_Z(\cdot)$ be the marginal density for Z, let:

$$\mathcal{B}_j \equiv \begin{bmatrix} (1\{Z \leqslant q(\lambda_2)\} - \lambda_2)/g_Z(\lambda_2) \\ (1\{Z \geqslant q(\lambda_1)\} - \lambda_1)/g_Z(\lambda_1) \end{bmatrix}$$

The Bahadur representation (Bahadur, 1966; David, 1981) can now be defined as:

$$\sqrt{N}[\hat{q}(\lambda) - q(\lambda)] = \sqrt{N}\overline{\mathcal{B}} + o_p(1), \quad \overline{\mathcal{B}} \equiv \frac{1}{N} \sum_{j=1}^N \mathcal{B}_j \tag{A.11}$$

D.8 First- and Second-stage estimator Based on the above definitions, we define:

$$\hat{\theta}_1 = \underset{\theta}{\text{argmax}}\, Q_1(\theta), \quad Q_1(\theta) = N^{-1} \sum_{i=1}^N \widehat{\tau_{ix}} \{\sum_{k=1}^n Y_i^k Ln(\hat{P}_i^k)\}, \tag{A.12}$$

$$\hat{\theta}_2 = \underset{\theta}{\text{argmax}}\, Q_2(\theta), \quad Q_2(\theta) = N^{-1} \sum_{i=1}^N \widehat{\tau_{iv}} \{\sum_{k=1}^n Y_i^k Ln(\hat{P}_i^k)\}, \tag{A.13}$$

D.9 Quantile trimming Let $W \in [X, Z]$ be the variable whose quantile marginal effects we are interested in. Define t_{qj} to be the true quantile trimming function which takes value 1 only if W is in a given quantile: $t_{qj} \equiv 1\{q(\lambda_1) < W <$

$q(\lambda_2)\}$, where λs are the upper and lower bound for that quantile and $q(\lambda_1), q(\lambda_2)$ the corresponding population quantile for W.

D.10 Quantile Marginal effects estimators Let $V(\hat{\theta}) \equiv [X_{F1} + X_F'\hat{\theta}_0^F, X_{B1} + X_B'\hat{\theta}_0^B, Z]$ be the estimated index. For an observation with $Y_i = K$, the true cumulative marginal effect $CME_i(\theta_0; K)$ defined in (2.6) can be estimated by:

$$\hat{m}_j(\hat{\theta}) \equiv \sum_{k=1}^{L}\{Y_i \geqslant k\}[\hat{P}_i^{ka}(V_F(\hat{\theta}), V_B(\hat{\theta}), Z+\delta) - \hat{P}_i^{ka}(V_F(\hat{\theta}), V_B(\hat{\theta}), Z)]$$

Refer to D.7 for quantile trimming t_q, we define the quantile marginal effect CME_q^K and its estimator by:

$$QME_q^K \equiv \frac{E[t_{qj}m_j(\theta_0)]}{E[t_{qj}]} \tag{A.14}$$

$$\widehat{QME}_q^K \equiv \frac{\sum_{j=1}^N \hat{t}_{qj}\hat{m}_j(\hat{\theta})}{\sum_{j=1}^N \hat{t}_{qj}} \tag{A.15}$$

D.11 Bahadur Representation Refer to D.3 and D.6, with $g_Z(\cdot)$ be the marginal density for Z, let:

$$\mathcal{B}_j \equiv \begin{bmatrix} (1\{Z \leqslant q(\lambda_2)\} - \lambda_2)/g_Z(\lambda_2) \\ (1\{Z \geqslant q(\lambda_1)\} - \lambda_1)/g_Z(\lambda_1) \end{bmatrix}$$

The Bahadur representation (Bahadur, 1966; David, 1981) can now be defined as:

$$\sqrt{N}[\hat{q}(\lambda) - q(\lambda)] = \sqrt{N}\overline{\mathcal{B}} + o_p(1), \quad \overline{\mathcal{B}} \equiv \frac{1}{N}\sum_{j=1}^N \mathcal{B}_j \tag{A.16}$$

To obtain convergence properties for the proposed probability estimator in D.4 and asymptotic normality for the index and quantile marginal effect estimators in the main text, we make the following assumptions.

A.2.2 Assumptions

A.1 DGP: The vector (Y_i^k,X_i) is i.i.d. over $i = 1, \cdots, N$ for each k in $\{1, \cdots, L\}$, and takes on values in a compact and finite support[7]. The columns of $X_i = [F_i, B_i, Z_i]$ are linearly independent with probability 1. In addition, we assume observations are grouped into $g = 1, \cdots, G$ clusters.

A.2 The error term: The error term U_i is conditionally independent of X_i: $E[u_i|X_i] = 0$, and error independence across clusters is assumed so that for $i \neq j$:

$$E[u_{ig}u_{ig'}|X_{ig}, X_{ig'}] = 0, \text{ unless } g = g' \tag{A.17}$$

Errors for bonds belonging to the same group may be correlated, with quite general heteroskedasticity and correlation.

A.3 Index Assumption Write the vector of indices $V(\theta_0) \equiv [F1 + F'\theta_0^F, B1 + B'\theta_0^B, Z]' = [V_F, V_B, Z]$, which depends on two vector, F and B, and a continuous variable Z. We further assume that $F1$ and $B1$ are continuous and functionally independent, and the following index assumption holds:

$$E[Y_i^k = 1|F_i, B_i, Z_i] = E[Y_i^k = 1|V(\theta_0)] \tag{A.18}$$

A.4 Parameter space The vector of true parameters values $\theta_0 \equiv [\theta_0^F, \theta_0^B]$ for the model in lies in the interior of a compact parameter space, Θ.

A.5 Conditional densities With V defined in **A.4**, denote $g(t|y, x, z)$ as its density conditioned on Y = y and X = x, Z = z. Denote $\nabla^d g(t|y, x, z)$ as the partials or cross partials up to order d, with $\nabla^0 g(t|y, x, z) \equiv g(t|y, x, z)$. With g defined on

[7] Since X denotes variables like firm's asset, leverage ratio...which seems to be naturally bounded from above

a compact support, we assume $g > 0$ and $\nabla^d g(t|y, x, z)$ uniformly bounded for d = 0,1,2,3 on the interior of its support.

A.1-A.5 are standard in the literature. Namely we require each index has at least one continuous variable **A.4** and densities for continuous variables and the indices must be sufficiently smooth, as implied by **A.5**. Additional window conditions will be required and stated directly in the theorems for which they are needed.

A.2.3 The estimator for $P^k(v)$ and its convergence property

Let $V_j = [F_{1j} + F_j'\theta_0^F, B_{1j} + B_j'\theta_0^B, Z_j]$ denotes the vector of indices at θ_0, and v is a fixed point. Consider the regression model in the main text: $E[Y_j^k|V_j] = P^K(V_j)$ in a "localized form" for the j^{th} observation:

$$\begin{aligned} Y_j^k &= P^k(V_j) + \epsilon_j \;\text{ with }\; \epsilon_j = Y_j^k - E[Y_j^k = 1|V_j] \\ &= P^k(v) + \underbrace{[P^k(V_j) - P^k(v)]}_{\Delta_j(v)} + \epsilon_j \end{aligned} \tag{A.19}$$

where Y_j^k is a binary variable that takes value one if bond j is rated as category K. This object $\Delta_j(v) \equiv P^k(V_j) - P^k(v)$ is termed as the "localization error".

As described in the main text, one kernel estimator for $P^k(v)$, which becomes a parameter after localization, is usually obtained by minimize the weighted squared sum of $Y_j^k - P^K(v)$ in the following way:

$$\hat{I}^k(v) = argmin_\alpha \sum_j (Y_j^k - \alpha)^2 K_h(V_j - v) \tag{A.20}$$

$$\Longrightarrow \hat{I}^k(v) = \frac{N^{-1}\sum_j Y_j^k K_h(V_j - v)}{N^{-1}\sum_j K_h(V_j - v)} \tag{A.21}$$

The kernel $K_h(V_j - v)$ is employed to downweight observations with index values far away from v. This estimator $\hat{I}^k(v)$, after scaled by $\hat{g}(v) \equiv N^{-1}\sum_j K_j(V_j - v)$, has a bias of order h^2, where h is the window size parameter. In a recent paper, Shen and

Klein (2017) show that by removing an estimate of the localization error, the following estimator:

$$\widehat{P}^k(v) \equiv \frac{N^{-1}\sum_j [Y_j^k - \hat{\Delta}_j(v)]K_j(v)}{N^{-1}\sum_j K_j(V_j - v)} = \hat{f}_1(v)/\hat{g}(v) \tag{A.22}$$

has a "better" convergence property than $\hat{I}^k(v)$ from Lemma A.4.1.

Lemma A.2.1 (Convergence Properties of Estimated Probability after Recursive Differencing)**.** *The following convergence properties hold for the conditional probability estimator defined above:*

(1) $sup_v E\{(\hat{g}(v)[\hat{P}^k(v) - E[\hat{P}^k(v)]])^2\}|_{\theta=\theta_0} = O_p(\frac{1}{Nh^3})$

(2) $sup_v |E[\hat{g}(v)(\hat{P}^k(v) - P^k(v))]|_{\theta=\theta_0} = O(h^4)$

(3) $sup_{v,\theta} \nabla_\theta^t |\hat{P}^k(v) - P^k(v)|$ = $O_p(h^4) + O_p(\frac{1}{N^{1/2}h^{3+t}})$, *with* $t = 0, 1, 2$

Proof. See Theorem 1 and Lemma 11 in Shen and Klein (2017). $\square$

In particular, they demonstrated that a lower order of bias can be achieved after estimating the localization error and subtracted from Y_j^k, without causing the order of variance to shoot up. As illustrated in the first two results, the order of the variance here is the same compared to that with a regular kernel, while a lower order bias is obtained (h^4 vs h^2). In addition, they also derive the uniform rate that this estimated probability and its derivatives goes to the truth. More importantly, they show that by repeating this process, the bias of estimating $P^k(v)$ can be reduced to any order.

A.2.4 "Residual Property" of $\nabla_\theta E[Y_i^k = 1|V_i(\theta)]|_{\theta=\theta_0}$

Lemma A.2.2. *Under the index assumption:* $E[Y_i^k = 1|X_i] = E[Y_i^k = 1|V_i(\theta_0)]$, *we have* $E[\nabla_\theta E[Y_i^k = 1|V(\theta)|_{\theta=\theta_0}]] = 0$.

Proof. This property is formally stated and proved in Klein and Shen (2010), and the authors thank Whitney Newey for mentioning a key idea in a private communication. □

This property plays a key role in reducing the bias of $\hat{\theta}$. To exploit this property as a bias control, however, one needs to estimate the model twice: first obtain a consistent estimate of θ_0, denote it as $\hat{\theta}_1$ and calculate the estimated index as $V(\hat{\theta}_1)$. Then, estimate θ_0 again but based the trimming on $V(\hat{\theta}_1)$.

A.3 Proofs of Asymptotic Results

A.3.1 Proof of Theorem 2.4.1

Proof. Let $\widehat{I_i^k(\theta_0)}$ be a standard Nadaraya-Watson estimator for the conditional expectation $E_i^k(\theta_0)$:

$$\widehat{I_i^k(\theta_0)} = \frac{N^{-1}\sum_j Y_i^k K_h(V_j - v_i)}{N^{-1}\sum_j K_h(V_j - v_i)} \equiv f_i^0/g_i \tag{A.23}$$

The strategy is to show that $\hat{g}(v,\theta)B$ is asymptotically equivalent to another object:

$$\hat{g}(v_i,\theta)B^* \equiv N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{L}\hat{g}(v_i,\theta)\tau_i[\widehat{I_i^k(\theta_0)} - E_i^k(\theta_0)]w_i$$

where the "weight function" $w_i \equiv \nabla_\theta E_i^k|_{\theta=\theta_0}\alpha_i$. This object, as shown in Klein and Shen (2010), is a second-order degenerate U-statistics. Recall from above that

$$\hat{g}(v_i,\theta)B \equiv N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{L}\hat{g}(v_i,\theta)\tau_i[\widehat{E_i^k(\theta_0)} - E_i^k(\theta_0)]w_i$$

Put it differently, I'm establishing an equivalence result between the recursive differencing estimator $\widehat{E_i^k(\theta_0)}$ and the regular kernel estimator $\widehat{I_i^k(\theta_0)}$, for the purpose of estimating index coefficient.

Note also that for each category k, the B component in the gradient has the same structure. Therefore we focus only on a single representative category without worrying about the summation over k. We proceed by first defining two intermediate objects that will simplify the analysis:

$$\begin{aligned}
\widehat{f}^0(v,\theta_0) &\equiv \hat{g}(v_i,\theta)\widehat{I_i^k(\theta_0)} = \frac{1}{N}\sum_j Y_j^k K_h(V_j - v) \equiv \frac{1}{N}\sum f_{0j}(v,\theta_0) \\
\widehat{f}^1(v,\theta_0) &\equiv \hat{g}(v_i,\theta)\widehat{E_i^k(\theta_0)} = \frac{1}{N}\sum_j [Y_j^k - \delta_j(v)]K_h(V_j - v) \equiv \frac{1}{N}\sum_j f_j(v,\theta_0)
\end{aligned}$$

To establish the equivalence result, it is sufficient to show that for each k:

$$\begin{aligned}
\hat{g}(v_i,\theta)[B^* - B] &= \qquad\qquad (A.24) \\
N^{-1/2}\sum_i [\widehat{f^0(v_i,\theta_0)} - \widehat{f^1(v_i,\theta_0)}]\tau_i w_i &\leqslant \sqrt{N}\sup_v |[\hat{f}^0(v,\theta_0) - \hat{f}^1(v,\theta_0)]\tau_i w_i| = o_p(1)
\end{aligned}$$

Using a "residual property" of $\nabla_\theta E_i^k|_{\theta=\theta_0}$ provided in Appendix A, it can be shown that $E[\tau_i f_{1j}(v,\theta_0)w_i] = E[\tau_i f_{2j}(v,\theta_0)w_i] = 0$. Therefore, with $G_n(v)$ as the empirical CDF and $G(v)$ the true density of V_j at θ_0, we have

$$\begin{aligned}
[\hat{f}^0(v) - \hat{f}^1(v)]\tau_i w_i &= \hat{f}^0(v)\tau_i w_i - E[\hat{f}^0(v)\tau_i w_i] - \hat{f}^1(v)\tau_i w_i + E[\hat{f}^1(v)\tau_i w_i] \\
&= \int_{V_j} f_{0j}(v,\theta_0)\tau_i w_i d[G_n(v) - G(v)] - \int_{V_j} f_{1j}(v,\theta_0)\tau_i w_i d[G_n(v) - G(v)] \\
&= \int_{V_j} [\tau_i(f_{0j}(v,\theta_0) - f_{1j}(v,\theta_0))w_i][dG_n(v) - dG(v)] \qquad (A.25) \\
&= \int_{V_j} [\tau_i \delta_j(v) K_h(V_j - v)w_i][dG_n(v) - dG(v)] \qquad (A.26)
\end{aligned}$$

Integrating-by-parts, the above integral equals

$$\tau_i\delta_j(v)K_h(V_j - v)w_i[G_n(v) - G(v)]|_{V_j\in\partial\Omega} - \int_{V_j}[G_n(v) - G(v)]d[\delta_j(v)K_h(V_j - v)w(v)]$$

The first boundary term vanishes because the kernel function K_h decays very fast when V_j is evaluated at boundary and v is a fixed point. For the second term, one can factor $\sup_v |G_n(v) - G(v)|$[8] outside of the integral. Then, since $\int_{V_j \in \Omega} d[\delta_j(v)K(V_j - v)w(v)]$ is $o_p(1)$, the result claimed in (A.24) follows. That is, $sup_v|(\hat{f}^0(v,\theta_0) - \hat{f}^1(v,\theta_0))\tau_i w_i|| = o_p(N^{-1/2})$.

□

A.3.2 Proof of Theorem 2.4.2

Proof. We first "stack" observations by each firm $g = 1, 2, \cdots, N_F$, and rewrite the quasi-likelihood function in (**??**) as:

$$\widetilde{Q}(\theta) = \sum_{g=1}^{N_F} \widehat{\tau_{iv}} \sum_{i \in g} g_i(Y_i|\theta) \tag{A.27}$$

with $g_i(Y_i|\theta) \equiv \sum_{k=1}^{L} Y_i^K Ln(\hat{P^k}(V_i))$. It is important to note that even though each g_i is correlated within firm g, the summation over each firm g, $\sum_{i \in g} g_i(Y_i|\theta)$ is independent across firms. From a taylor expansion of the estimated gradient on $\hat{\theta}$ and the fact that the estimated gradient is zero evaluated at θ_0, we have

$$\begin{aligned} \sqrt{N_F}(\hat{\theta} - \theta_0) &= -\hat{H}(\theta^+)^{-1}\sqrt{N_F}\hat{G}(\theta_0) \quad \theta^+ \in (\theta_0, \hat{\theta}) & \text{(A.28)} \\ &= -\hat{H}(\theta^+)^{-1}\sqrt{N_F}\sum_{g=1}^{N_F} \hat{G}_g(\theta_0)/N & \text{(A.29)} \end{aligned}$$

where $\hat{G}(\theta) = \nabla_{\theta'}\widehat{Q}_2(\theta)$, $\hat{H}(\theta) = \nabla_{\theta'\theta}\widehat{Q_2}(\theta)$, and $\hat{G}_g(\theta) = \nabla_{\theta'} \sum_{i \in g} g_i(Y_i^K|\theta)$ for any θ in its support. From Lemma A.4.5, $\hat{H}(\theta^+)$ converges in probability to $H_0 \equiv$

[8] This term is $O_p(N^{-1/2})$ according to Nadaraya (1965); Eddy and Hartigan (1977)

$E[H(\theta_0)]$. We let $P_i^k \equiv P^k(V_i)$ to simplify the notation,

$$\begin{aligned}
\sqrt{N_F}\hat{G}(\theta_0) &\equiv \sqrt{N_F}\sum_{g=1}^{N_F}\sum_{i\in g}\widehat{\tau_{iv}}\sum_k \frac{Y_i^k}{\hat{P}_i^k}\frac{\partial \hat{P}_i^k}{\partial\theta}|_{\theta=\theta_0} && \text{(A.30)}\\
&= \sqrt{N_F}\sum_{g=1}^{N_F}\sum_{i\in g}\widehat{\tau_{iv}}\sum_k \frac{Y_i^k-\hat{P}_i^k}{\hat{P}_i^k}\frac{\partial \hat{P}_i^k}{\partial\theta}|_{\theta=\theta_0} && \text{(A.31)}\\
&= \sqrt{N_F}\sum_{g=1}^{N_F}\sum_{i\in g}\widehat{\tau_{iv}}[\frac{Y_i^1-\hat{P}_i^1}{\hat{P}_i^1}\frac{\partial \hat{P}_i^1}{\partial\theta}|_{\theta=\theta_0} && \text{(A.32)}\\
&+ \frac{Y_i^2-\hat{P}_i^2}{\hat{P}_i^2}\frac{\partial \hat{P}_i^2}{\partial\theta}|_{\theta=\theta_0}+\dots] && \text{(A.33)}
\end{aligned}$$

The first equality above follows from the fact that$\sum_k \frac{\partial \hat{P}_i^k}{\partial\theta} = \frac{\partial \sum_k \hat{P}_i^k}{\partial\theta} = \frac{\partial 1}{\partial\theta} = 0$. As the first term in (A.33) and all remaining terms have the same structure, it suffices to analyze the first term. From standard argument in Pakes and Pollard (1989) and Klein and Spady (1993), the estimated trimming function $\widehat{\tau_{iv}}$ can be replaced by the truth asymptotically. With $\widehat{w_i^k} \equiv \frac{\partial \hat{P}_i^k}{\hat{P}_i^k\partial\theta}|_{\theta=\theta_0}$, we may write this term as:

$$\begin{aligned}
\sqrt{N_F}\sum_{g=1}^{N_F}\sum_{i\in g}\tau_{iv}[Y_i^k-\hat{P}_i^k]\widehat{w_i^k} &= \sqrt{N_F}[\underbrace{\sum_{g=1}^{N_F}\sum_{i\in g}\tau_{iv}[Y_i^k-P_i^k]\widehat{w_i^k}/N}_{D_1} && \text{(A.34)}\\
&- \underbrace{\sum_{g=1}^{N_F}\sum_{i\in g}\tau_{iv}[\hat{P}_i^k-P_i^k]\widehat{w_i^k}/N]}_{D_2} && \text{(A.35)}
\end{aligned}$$

By making use of the fact that the residual $Y_i^k - P_i^k$ has zero conditional expectation, D_1 could be replaced by $N^{-1}\sum_i \tau_{iv}(Y_i^k - P_i^k)w_i^k$ through a mean-square convergence results (Klein and Spady, 1993). For D_2, with $1/12 < r < 1/8$, from Lemma A.4.7, we have

$$\sqrt{N_F}(D_2 - D_2^*) = o_p(1), \quad D_2^* \equiv \sum_{g=1}^{N_F}\sum_{i\in g}\tau_{iv}[\hat{P}_i^k - P_i^k]w_i^k/N \qquad \text{(A.36)}$$

so that the estimated weight w can be replaced by the truth asymptotically. Recall that the probability P_i^k is estimated by a ratio of kernels according to D.4:

$$\widehat{P}^k(v) \equiv \frac{\sum_{j=1}^N [Y_j^K - \hat{\Delta}_j(i)]K_j(v)}{\sum_{j=1}^N K_j(v)} = \hat{f}/\hat{g} \tag{A.37}$$

where $\hat{\Delta}_j(i) = \widehat{I}^K(V_j) - \widehat{I}^K(v)$ is an estimate of the localization error. Applying Lemma A.4.7 again, we have $\sqrt{N_F}(D_2^* - U_N) = o_p(1)$ where

$$U_N \equiv \frac{1}{N}\sum_{g=1}^{N_F}\sum_{i\in g}(\hat{f}_i/\hat{g}_i - P_i)\tau_{vi}w_i^k\frac{\hat{g}_i}{g_i} \tag{A.38}$$

$$= \frac{1}{N}\sum_{g=1}^{N_F}\sum_{i\in g}(\hat{f}_i - P_i\hat{g}_i)\tau_{vi}\frac{w_i^k}{g_i} + o_p(1) \tag{A.39}$$

Let $\rho_{ij} \equiv [Y_j^k - \hat{\Delta}_j(i) - P_i]K_{ij}\tau_{vi}w_i^k/g_i$, we can rewrite U_K in the form of a second order U-statistic:

$$U_N = \frac{1}{N(N-1)}\sum_{g=1}^{N_F}\sum_{i\in g}\sum_{j\geqslant i}\rho_{ij}^* \text{ with } \rho_{ij}^* = \frac{\rho_{ij}+\rho_{ji}}{2} \tag{A.40}$$

After applying the standard approximation theory of U-statistics (Serfling, 2009; Powell *et al.*, 1989), we obtain $\sqrt{N_F}(U_N - \hat{U}_N) = o_p(1)$ where

$$\sqrt{N_F}\hat{U}_N \equiv \sqrt{N_F}\sum_{g=1}^{N_F}\sum_{i\in g}(E[\rho_{ij}|Obs_i] + E[\rho_{ji}|Obs_i])/N = T_1 + T_2 \tag{A.41}$$

It can be shown that for each term in T_1:

$$E[term_i] \equiv E_i\{E[\rho_{ij}|Obs_i]\} = E[\rho_{ij}] = E_{V_i,V_j}\{E[\rho_{ij}|V_i,V_j]\} = 0$$

$$Var[term_i] = O(1)$$

T_1 is therefore $o_p(1)$ from standard sample mean property. $T_2 = 0$ from the law of iterated expectation and the residual property of w_i^k. Therefore, since U_N converges in probability to zero, we have

$$\sqrt{N_F}\hat{G}(\theta_0) = \sqrt{N_F}\sum_{g=1}^{N_F}\sum_{i\in g}\tau_{iv}\frac{Y_i^k - P_i^k}{P_i^k}\frac{\partial P_i^k}{\partial\theta}/N + o_p(1) \tag{A.42}$$

Then, referring to the expression in (A.28), $\sqrt{N_F}(\hat{\theta}-\theta_0)$ has an asymptotic linear form:

$$\sqrt{N_F}(\hat{\theta}-\theta_0) = H_0^{-1}\sqrt{N_F}\sum_{g=1}^{N_F} G_g(\theta_0)/N + o_p(1) \tag{A.43}$$

$$G_g(\theta_0) \equiv \sum_{i\in g}\sum_{k=1}^{L}\tau_{iv}\frac{Y_i^k - P^k}{P^k}\frac{\partial P^k}{\partial\theta}|_{\theta=\theta_0} \tag{A.44}$$

Since each cluster g consists of different number of observations, the term $G_g(\theta_0)$ is independent but not identically distributed. We apply the Lindberg-Feller CLT and obtain:

$$\sqrt{N_F}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, H_0^{-1}[\sum_{g=1}^{N_F} G_g(\theta_0)G_g'(\theta_0)]H_0^{-1}) \tag{A.45}$$

□

A.3.3 Proof of Theorem 2.4.3

Proof. For consistency, it can be shown that the estimated quantile trimming function $\hat{t}_{qj}$ will converge in probability to the truth. Since GME_q^K is a continuous function of the estimated probability, and $\sup_\theta|\hat{P}_k(\theta) - P_k(\theta)| \xrightarrow{p} 0$ for all k, consistency follows from continuous function theorem and Theorem 3.1. To derive the estimator's limiting distribution, defining a term

$$\widehat{M_q^k} \equiv \frac{\sum_{j=1}^{N}\hat{t}_{qj}[\widehat{P_i^k}(V_F, V_B, Z+\delta;\hat{\theta}) - \widehat{P_i^k}(V_F, V_B, Z;\hat{\theta})]}{\sum_{j=1}^{N}\hat{t}_{qj}}$$

so it follows that

$$\sqrt{N_F}(GME_q^K - \widehat{GME_q^K}) = \sqrt{N_F}\sum_{k=1}^{K-1}(\widehat{M_q^k} - M_q^k) \tag{A.46}$$

In the following proofs, all of the summation $\sum_{j=1}^{N}$ should be taken as $\sum_{g=1}^{N_F}\sum_{j\in g}$. Since all terms within the above summation have the same structure, it suffice to analyze just one term for any k. To simplify notation, let

$$\widehat{m_j^k} \equiv \widehat{P_i^k}(V_F, V_B, Z+\delta;\hat{\theta}) - \widehat{P_i^k}(V_F, V_B, Z;\hat{\theta}) \tag{A.47}$$

$$m_j^k \equiv P_i^k(V_F, V_B, Z+\delta;\theta_0) - P_i^k(V_F, V_B, Z;\theta_0) \tag{A.48}$$

$$\hat{N}_q \equiv \sum \hat{t}_{qj} \tag{A.49}$$

and we proceed with the following decomposition:

$$\begin{aligned}
\sqrt{N_F}(\widehat{M_q^k} - M_q^k) &\equiv \sqrt{N_F}\left(\frac{\sum_{j=1}^N \hat{t}_{qj}\hat{m}_j^k(\hat{\theta})}{\sum_{j=1}^N \hat{t}_{qj}} - \frac{\sum_{j=1}^N t_{qj} m_j^k}{\sum_{j=1}^N t_{qj}}\right) \\
&= \sqrt{N_F}(\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4) \qquad \text{(A.50)} \\
\text{where} \quad \Delta_1 &= (N/\hat{N}_q)\frac{1}{N}\sum_j \hat{t}_{qj}(\hat{m}_j^k(\hat{\theta}) - m_j^k(\theta_0)) \\
\Delta_2 &= (N/\hat{N}_q)\frac{1}{N}\sum_j m_j^k(\theta_0)(\hat{t}_{qj} - t_{qj}) \\
\Delta_3 &= (N/\hat{N}_q)\left(\frac{1}{N}\sum_j t_{qj} m_j^k(\theta_0) - E[t_{qj}]M_q^k\right) \\
\Delta_4 &= (N/\hat{N}_q)\frac{1}{N}\sum_j (E[t_{qj}] - \hat{t}_{qj})M_q^k
\end{aligned}$$

In what follows, we study the limit of these four Δ terms respectively. Loosely speaking, Δ_1 reflects the estimation uncertainly from the parameter θ_0. We show that Δ_1 can be characterized as the sum of one term related with the limiting distribution of $\hat{\theta} - \theta_0$ and another third order U-statistic that vanishes asymptotically. Both Δ_2 and Δ_4 are related with estimation uncertainty from quantiles, which fortunately is a problem that received fair amount of attention in the literature of statistics and econometrics. We apply results from Pakes and Pollard (1989); Shen and Klein (2017) and derive their probability limits. There is no estimated components in Δ_3, so a central limited theorem can be applied directly.

Limiting distribution of Δ_1:

For Δ_1, note that

$$\begin{aligned}
\sqrt{N_F}\Delta_1 &\equiv \Delta_{11} + \Delta_{12} && \text{(A.51)}\\
\Delta_{11} &\equiv (N/\hat{N}_q)\frac{\sqrt{N_F}}{N}\sum_j (\hat{t}_{qj} - t_{qj})(\hat{m}_j(\hat{\theta}) - m_j(\theta_0)) && \text{(A.52)}\\
\Delta_{12} &\equiv (N/\hat{N}_q)\frac{\sqrt{N_F}}{N}\sum_j t_{qj}(\hat{m}_j(\hat{\theta}) - m_j(\theta_0)) && \text{(A.53)}
\end{aligned}$$

Δ_{11} is $o_p(1)$ in a fashion identical to the term Δ_{11} in Theorem 5 of Shen and Klein (2017). Turning to Δ_{12}, we furthur decompose Δ_{12} into four pieces:

$$\begin{aligned}
\Delta_{12} &= (N/\hat{N}_q)\sqrt{N_F}(T_1 + T_2 - T_3 - T_4)\\
T_1 &= \frac{1}{N}\sum_{j=1}^{N} t_{qj}[\widehat{P_j^k}(V_F, V_B, Z+\delta; \hat{\theta}) - \widehat{P_j^k}(V_F, V_B, Z+\delta; \theta_0)]\\
T_2 &= \frac{1}{N}\sum_{j=1}^{N} t_{qj}[\widehat{P_j^k}(V_F, V_B, Z+\delta; \theta_0) - P_j^k(V_F, V_B, Z+\delta; \theta_0)]\\
T_3 &= \frac{1}{N}\sum_{j=1}^{N} t_{qj}[\widehat{P_j^k}(V_F, V_B, Z; \hat{\theta}) - \widehat{P_j^k}(V_F, V_B, Z; \theta_0)]\\
T_4 &= \frac{1}{N}\sum_{j=1}^{N} t_{qj}[\widehat{P_j^k}(V_F, V_B, Z; \theta_0) - P_j^k(V_F, V_B, Z; \theta_0)] && \text{(A.54)}
\end{aligned}$$

For T_1, from a Taylor expansion around θ_0, we have:

$$\begin{aligned}
\sqrt{N_F}T_1 &= \frac{1}{N}\sum_{j=1}^{N} t_{qj}\nabla_\theta \widehat{P_j^k}|_{\theta=\theta*}\sqrt{N_F}(\hat{\theta} - \theta_0) + o_p(1) && \text{(A.55)}\\
&= E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z+\delta; \theta_0)|]\sqrt{N_F}(\hat{\theta} - \theta_0) + o_p(1) && \text{(A.56)}
\end{aligned}$$

The second equality follows from the convergence of $\sup|\nabla_\theta \hat{P^k}_i(\theta) - \nabla_\theta P_i^k(\theta)| \xrightarrow{p} 0$ in Lemma A.4.4. Turning to T_2, in Lemma A.4.9 we show that $\sqrt{N}T_2$ can be written as the sum of two pieces U_1 and U_2. In Lemma A.4.10 we proceed to show U_1 is a degenerate third order U-stat which would vanish asymptotically. In Lemma A.4.11 we show that U_2 can be approximated by a sample mean of iid terms. Therefore we have:

$$\begin{aligned}\sqrt{N}T_2 &= \sqrt{N}(U_1 + U_2) \\ &= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} E[t_{qj}|V_F, V_B, Z+\delta]\epsilon_j^\delta + o_p(1) \qquad \text{(A.57)}\end{aligned}$$

where $\epsilon_j^\delta \equiv Y_j^k - E[Y_j^k|V_F, V_B, Z+\delta]$. The proof strategy for T_3 and T_4 mimic the ones for T_1 and T_2:

$$T_3 = E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z; \theta_0)|]\sqrt{N}(\hat{\theta} - \theta_0) + o_p(1) \qquad \text{(A.58)}$$

$$T_4 = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} E[t_{qj}|V_F, V_B, Z]\epsilon_j + o_p(1) \qquad \text{(A.59)}$$

Combining (A.56)(A.57)(A.58)(A.59), we have

$$\begin{aligned}\sqrt{N_F}\Delta_{12} &= \sqrt{N_F}(T_1 - T_3 + T_2 - T_4) \\ &= \frac{E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z+\delta; \theta_0)|] - E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z; \theta_0)|]}{E[t_{qj}]}\sqrt{N_F}(\hat{\theta} - \theta_0) \\ &+ \frac{1}{E[t_{qj}]\sqrt{N}}\sum_{j=1}^{N} E[t_{qj}|V_F, V_B, Z+\delta]\epsilon_j^\delta - E[t_{qj}|V_F, V_B, Z]\epsilon_j + o_p(1) \qquad \text{(A.60)}\end{aligned}$$

Limiting distribution of Δ_2 and Δ_4:

Both Δ_2 and Δ_4 in (A.50) are related to quantile estimation uncertainty. Turning to Δ_4, we have, up to $o_p(1)$:

$$\begin{aligned}
\Delta_4 &= (N/\hat{N}_q)\frac{1}{N}\sum_j(\hat{t}_{qj} - E[t_{qj}])CME_q^k && \text{(A.61)}\\
&= (N/\hat{N}_q)\frac{1}{N}\sum_j(\hat{t}_{qj} - t_{qj})CME_q^k && \text{(A.62)}\\
&+ (N/\hat{N}_q)\frac{1}{N}\sum_j(t_{qj} - E[t_{qj}])CME_q^k && \text{(A.63)}
\end{aligned}$$

Both Δ_2 and the first piece of Δ_4 take the form $(N/\hat{N}_q)\frac{1}{N}\sum W_j(\hat{t}_{qj} - t_{qj})$ where W_j is either iid term over j or constant. From Lemma 3 of Shen and Klein (2017), who exploits an important result from Pakes and Pollard (1989), the Bahadur representation, and the convergence of $N/\hat{N}_q$ to $1/E[t_{qj}]$, we have:

$$\begin{aligned}
\sqrt{N}\Delta_2 &= \frac{1}{E[t_{qj}]}\nabla_q E[t_{qj}m_j(\theta_0)]\sqrt{N}\overline{\mathcal{B}} + o_p(1) && \text{(A.64)}\\
\sqrt{N}\Delta_4 &= \frac{1}{E[t_{qj}]}\nabla_q E[t_{qj}GME_q]\sqrt{N}\overline{\mathcal{B}} && \text{(A.65)}\\
&+ \frac{GME_q}{\sqrt{N}E[t_{qj}]}\sum_{j=1}^{N}(t_{qj} - E[t_{qj}]) + o_p(1) && \text{(A.66)}
\end{aligned}$$

Note that the normality of $\overline{\mathcal{B}}$, the sample average of $\mathcal{B}_j$ has been derived by Bahadur (1966) and widely used in the econometric literature. The second piece of Δ_4 is an average of iid terms and therefore could be handled by CLT.

Limiting distribution of Δ_3:

Δ_3 in (A.50) is exactly an sample average of iid terms minus its expectation. From Lindeberg-Levy CLT, we shall have

$$\begin{aligned}\sqrt{N}\Delta_3 &= (N/\hat{N}_q)(\frac{1}{N}\sum_j t_{qj}m_j(\theta_0) - E[t_{qj}]GME_q) && \text{(A.67)}\\ &= \frac{1}{\sqrt{N}E[t_{qj}]}\sum_j (t_{qj}m_j(\theta_0) - E[t_{qj}]GME_q) && \text{(A.68)}\end{aligned}$$

After incorporating the results of all four Δs and the asymptotic linear structure of $\sqrt{N_F}(\hat{\theta}-\theta_0)$ in (A.43), the vector of $\widehat{M_q^k} - M_q^k$, where k=1,2,3...L, has an asymptotic linear form:

$$\begin{aligned}\sqrt{N_F}(M_q^k - \widehat{M_q^k}) &= \frac{\sqrt{N_F}}{N}\sum_{g=1}^{N_F}[\Psi_{1g}^k + \Psi_{2g}^k + \Psi_{3g}^k + \Psi_{4g}^k] + o_p(1)\\ \Psi_{1g}^k &\equiv \frac{E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z+\delta;\theta_0)] - E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z;\theta_0)]}{E[t_{qj}]}H_0^{-1}G_g(\theta_0)\\ \Psi_{2g}^k &\equiv \{\nabla_q E[t_{qj}m_j(\theta_0)] - \nabla_q E[t_{qj}CME_q^k]\}\sum_{j\in g}\frac{\mathcal{B}_j}{E[t_{qj}]}\\ \Psi_{3g}^k &\equiv \frac{E[t_{qj}|V_F, V_B, Z+\delta]\epsilon_j^\delta - E[t_{qj}|V_F, V_B, Z]\sum_{j\in g}\epsilon_j}{E[t_{qj}]}\\ \Psi_{4g}^k &\equiv \sum_{j\in g}\frac{t_{qj}m_j(\theta_0) - E[t_{qj}]CME_q^k}{E[t_{qj}]} - \frac{t_{qj} - E[t_{qj}]}{E[t_{qj}]}CME_q^k && \text{(A.69)}\end{aligned}$$

and after applying the Central Limited Theorem to that vector, we have:

$$\sqrt{N_F}\begin{bmatrix}\widehat{M_q^1} - M_q^1\\ \widehat{M_q^2} - M_q^2\\ \vdots\\ \widehat{M_q^L} - M_q^L\end{bmatrix} \xrightarrow{d} Z \sim N(0,\Omega) \qquad \text{(A.70)}$$

where $\Omega \equiv E[(\Psi_{1g}^k + \Psi_{2g}^k + \Psi_{3g}^k + \Psi_{4g}^k)'(\Psi_{1g}^k + \Psi_{2g}^k + \Psi_{3g}^k + \Psi_{4g}^k)]$ and each Ψ_g^k is a L X 1 column vector calculated from the formula above. By construction, the object of interest $\sqrt{N}(Q\hat{M}E_q^K - QME_q^K)$ can be obtain from the following linear transformation on the above vector. With $A = (1, 1, ...1, 0, 0, 0)$, an 1 X L dimension row vector with the first K-1 component equals one and the rest equals zero, we have

$$\sqrt{N}(Q\hat{M}E_q^K - QME_q^K) = \sqrt{N}A \begin{bmatrix} \widehat{M_q^1} - M_q^1 \\ \widehat{M_q^2} - M_q^2 \\ \vdots \\ \widehat{M_q^L} - M_q^L \end{bmatrix} \xrightarrow{d} AZ \sim N(0, A'\Omega A) \quad \text{(A.71)}$$

Then the normality result follows. $\square$

A.4 Intermediate Lemmas and Proofs

Consider the mean regression model in the main text: $E[Y_j^K = 1|V_j] = P^K(V_j)$ in a "localized form" for the j^{th} observation:

$$\begin{aligned} Y_j^K &= P^K(V_j) + \epsilon_j \text{ with } \epsilon_j = Y_j^K - E[Y_j^K = 1|V_j] && \text{(A.72)} \\ &= P^K(v) + [P^K(V_j) - P^K(v)] + \epsilon_j && \text{(A.73)} \end{aligned}$$

where Y_j^K is a binary variable that takes value one if bond j is rated as category K, $V_j = [F_{1j} + F_j'\theta_0^F, B_{1j} + B_j'\theta_0^B, Z_j]$ denotes the vector of indices at θ_0, and v is a fixed point that we choose to localize on. This object $P^K(V_j) - P^K(v)$ is termed as the "localization error".

Normally a kernel estimator for $P^K(v)$, which becomes a parameter after localization, is obtained by minimize the weighted squared sum of $Y_j^k - P^K(v)$ in the following way:

$$\hat{I}^K(v) = argmin_\alpha \sum_j (Y_j^K - \alpha)^2 K_j(v) \tag{A.74}$$

$$\Longrightarrow \hat{I}^K(v) = \frac{\sum_j Y_j^K K_j(v)}{\sum_j K_j(v)} \tag{A.75}$$

The kernel $K_j(v)$ is employed to downweight observations with index values far away from v. This estimator $\hat{I}^K(v)$, after scaled by the density $\hat{g}(v)$, has a bias of order h^2, where h is the window size defined in **D.2**. In a recent paper, Shen and Klein (2017) suggest that the following estimator:

$$\widehat{P}^K(v) \equiv \frac{\sum_j [Y_j^K - \hat{\Delta}_j(v)] K_j(v)}{\hat{g}(v)} = \hat{f}(v)/\hat{g}(v) \tag{A.76}$$

has a "better" convergence property than $\hat{I}^K(v)$, as Lemma A.4.1 suggested. More importantly, they show that by repeating this process, the bias of estimating $P^K(v)$ can be reduced to any order.

Lemma A.4.1 (Convergence Properties of Estimated Probability after Recursive Differencing)**.** *The following convergence properties hold for the conditional probability estimator defined above:*

(1) $sup_v E\{(\hat{g}(v)[\hat{P}^K(v) - E[\hat{P}^K(v)]])^2\}|_{\theta=\theta_0} = O_p(\frac{1}{Nh^3})$

(2) $sup_v |E[\hat{g}(v)(\hat{P}^K(v) - P^K(v))]|_{\theta=\theta_0} = O(h^4)$

(3) $sup_{v,\theta} \nabla_\theta^t |\hat{P}^K(v) - P^K(v)| = O_p(h^4) + O_p(\frac{1}{N^{1/2}h^{3+t}})$*, with* $t = 0, 1, 2$

Proof. See Theorem 1 and Lemma 11 in Shen and Klein (2017). In particular, they demonstrated that a lower order of bias can be achieved after estimating the localization error and subtracted from Y_j^K, without causing the order of variance to shoot up. As illustrated in the first two results, the order of the variance here is the same compared to that with a regular kernel in Lemma A.4.3, while a lower order bias is obtained (h^4

vs h^2). In addition, they also derive the uniform rate that this estimated probability and its derivatives goes to the truth. □

Lemma A.4.2 (Uniform Convergence of Estimates to Expectations)**.** *Let w be a K dimensional vector and assume that m(w) is a sample average of terms m(w; z_i), where z_i are i.i.d. Assume that with h →0, we have uniformly over N:*

$$h^{r+1}|m(w;z_i)| < c, \quad r+1>0 \quad \textit{and} \quad h^s|\partial m(w;z_i)/\partial w| < c, \quad s>0$$

Let E[m(w)] be the expectation of m(w) taken over the distribution of z_i. Then, with w in a compact set and for any $\alpha > 0$:

$$N^{(1-\alpha)/2}h^{r+1}sup|m(w)-E[m(w)]| \xrightarrow{p} 0 \qquad a.s.$$

Proof. See the proof of Lemma A.4.2 on pp 411 in Klein and Spady (1993). The proof utilizes important implication in Hoeffding (1963) and Bhattacharya (1967). □

Lemma A.4.3 (Uniform Convergence for Density Estimator and Its Derivatives)**.** *Let $\hat{g}$ be a estimated density with 3 indices defined as in* ***D.3*** *and $\nabla_\theta^r \hat{g}$ be the r-th density derivative with respect to θ, r = 0,1,2. If all x's are bounded, then:*

(A.4.3.1) $sup_{v,\theta}E\{(\nabla_\theta^r\hat{g}(v)-E[\nabla_\theta^r\hat{g}(v)])^2]\} = O_p(\frac{1}{Nh^{2r+3}})$

(A.4.3.2) $sup_{v,\theta}|E[\nabla_\theta^r\hat{g}(v)]-\nabla_\theta^r g(v)| = O_p(h^2)$

(A.4.3.3) $sup_{v,\theta}|\nabla_\theta^r\hat{g}(v)-\nabla_\theta^r g(v)| = O_p(N^{-\frac{1}{2}}h^{-r-3})+O_p(h^2)$

Proof. The bias and variance calculation is fairly standard in the literature. One can find them in Hansen (2009). We just outline the proof for r = 0. For the bias calculation,

$$
\begin{aligned}
E[\hat{g}(\theta)] &= \int_{-\infty}^{\infty} \frac{1}{h} K(\frac{V_i(\theta) - v(\theta)}{h}) g(v(\theta)) dv && \text{(A.77)}\\
&= \int_{-\infty}^{\infty} K(u) g(v + hu) du \qquad \text{let } u = \frac{V_i(\theta) - v(\theta)}{h} && \text{(A.78)}\\
& \quad \text{After a second order taylor expansion on h=0} && \text{(A.79)}\\
&= \int_{-\infty}^{\infty} K(u)[g(v) + g'(v^+)hu + \frac{1}{2} g''(v^+) h^2 u^2] du && \text{(A.80)}\\
&= A_1 + A_2 + A_3 && \text{(A.81)}
\end{aligned}
$$

Given that K is the regular Gaussian kernel defined in **D.2**, $A_1 = g(v(\theta))$ because $\int K(u)du = 1$. $A_2 = 0$ since K is symmetric by D.4, so $\int K(u)udu = 0$. $A_3 = O_p(h^2)$ because the Gaussian kernel has variance of 1 and g'' is bounded by A.5. Therefore, we have

$$E[\hat{g}(\theta)] - g(\theta) = O_p(h^2) \tag{A.82}$$

The proof for density derivatives are similar. For the uniform convergence of $\nabla_\theta^r \hat{g}(v)$, note that

$$
\begin{aligned}
sup_{v,\theta} |\nabla_\theta^r \hat{g}(v) - \nabla_\theta^r g(v)| &\leqslant \underbrace{sup_{v,\theta} |\nabla_\theta^r \hat{g}(v) - E[\nabla_\theta^r \hat{g}(v)]|}_{\text{This piece is studied in Lemma A.4.2}} && \text{(A.83)}\\
&+ \underbrace{sup_{v,\theta} |E[\nabla_\theta^r \hat{g}(v)] - \nabla_\theta^r g(v)|}_{\text{The bias calculation}} && \text{(A.84)}
\end{aligned}
$$

To utilize Lemma A.4.2 to the study the converge property of the first term, let $m(w; Z_i) = \hat{g}(v)$ where $w = \{v, \theta\}$. By the way that $\hat{g}(v)$ is defined in **D.3**, we have $h^{3+r} \nabla_\theta^r \hat{g}(v) = O(1)$, so it follows that the first term is $O_p(N^{-1/2} h^{-r-3})$ □

Lemma A.4.4 (Mean-square Convergence of Estimated Probability)**.** *Referring **D.4** for the definition of $\hat{P}^k(v)$, let $\nabla_\theta^t \hat{P}$ be the t-th derivative with respect to θ, with t =*

0,1,2. Then, for h = N^{-r} *and under the conditions of Lemma A.4.2 and Lemma A.4.3,*

$$\frac{1}{N}\sum_i(\hat{P^k}_i(\theta) - P_i^k(\theta))^2 = O_p(N^{-1}h^{-3}) + O_p(h^8) \tag{A.85}$$

Proof. Note that $\frac{1}{N}\sum_i(\hat{P}_i^k(\theta) - P_i^k(\theta)) = \frac{1}{N}\sum_i(\frac{\hat{f}_i}{\hat{g}_i} - \frac{f_i}{g_i}) \leqslant sup(\frac{1}{\hat{g}_i})\frac{1}{N}\sum_i(\hat{f}_i - P_i^k(\theta)\hat{g}_i)$. Since we trim $\hat{g}_i$ away from zero, $1/inf(\hat{g}_i)$ is clearly bounded from above. Due to the recursive differencing structure in $\hat{f}$ that we explained in Lemma A.4.1, we have that for some constant B = $O(1)$,

$$\begin{aligned}
\frac{1}{N}\sum_i(\hat{P}_i^k(\theta) - P_i^k(\theta))^2 &\leqslant B^2\frac{1}{N}\sum_i(\hat{f}_i - P_i^k(\theta)\hat{g}_i)^2 \\
&= B^2\frac{1}{N}\sum_i(\hat{f}_i - E[\hat{f}] + E[\hat{f}] - P_i^k(\theta)\hat{g}_i)^2 \\
&= \underbrace{B^2\frac{1}{N}\sum_i(\hat{f}_i - E[\hat{f}])^2}_{\text{variance of } \hat{f}} + \underbrace{B^2\frac{1}{N}\sum_i(E[\hat{f}] - P_i^k(\theta)\hat{g}_i)^2}_{\text{squared bias of } \hat{f}} \\
&+ \underbrace{B^2\frac{2}{N}\sum_i(\hat{f}_i - E[\hat{f}])(E[\hat{f}] - P_i^k(\theta)\hat{g}_i)}_{o_p(1)} \\
&= O_p^2(1)(O_p(N^{-1}h^{-3}) + O_p(h^8))
\end{aligned}$$

□

Lemma A.4.5 (Convergence of Hessian)**.** *Assume the window size h =* N^{-r} *and* $1/16 < r < 1/10$. *Then, under the conditions of lemma A.4.2 and C.2 and with* $\theta^+ \in [\hat{\theta}, \theta_0]$,

$$\hat{H}(\theta^+)^{-1} \xrightarrow{p} H_0 = E[H(\theta_0)]$$

Proof. Given that the Hessian is a continuous function on θ, the desired argument $\hat{H}(\theta^+) \xrightarrow{p} E[H(\theta_0)]$ would follow if we have the following two conditions holds: (a) $\theta^+ \xrightarrow{p} \theta_0$ (b) sup$|\hat{H}(\theta) - E[H(\theta)]| \xrightarrow{p} 0$. Condition (b) implies that $\hat{H}(\theta_0) \xrightarrow{p}$ $E[H(\theta_0)]$. If (a) holds, then by the continuous mapping theorem we have $\hat{H}(\theta^+) \xrightarrow{p}$

$\hat{H}(\theta_0) \xrightarrow{p} E[H(\theta_0)]$. Condition (a) directly follows from consistency because θ^+ is some intermediate point between θ_0 and $\hat{\theta}$. To show (b), note that:

$$\begin{aligned} sup|\hat{H}(\theta) - E[H(\theta)]| &\leqslant sup|\hat{H}(\theta) - H(\theta)| + sup|H(\theta) - E[H(\theta)]| && \text{(A.86)} \\ &\leqslant T_1 + T_2 && \text{(A.87)} \end{aligned}$$

$T_2 \xrightarrow{p} 0$ from Lemma A.4.2. Note that the hessian, by definition, is the second derivative of the quasi-log-likelihood function:

$$H(\theta_0) \equiv \frac{1}{N}\sum_i(\frac{Y_i^k}{P_{k,i}}\nabla_{\theta'\theta}P_{k,i} - \frac{Y_i^k}{P_{k,i}^2}\nabla_\theta P_{k,i})$$

To make T_1 uniformly converge to 0, we need $\nabla_\theta^t \hat{P_i^k}$ uniformly converge to its associated estimand for t = 0,1,2, with the rate at t=2 being the slowest. From Lemma A.4.3, $r < 1/10$ ensures this is the case for t = 2. Therefore we have $\hat{H}(\theta^+)^{-1} \xrightarrow{p} H_0 = E[H(\theta_0)]$. □

Lemma A.4.6 (Double Convergence)**.** *Let* a_i,b_i *be some iid quantity, and* $\hat{a}_i, \hat{b}_i$ *be their estimator respectively. If* $\frac{1}{N}\sum_i(\hat{a}_i - a_i)^2 = O_p(N^{-\alpha_1})$, $\frac{1}{N}\sum_i(\hat{b}_i - b_i)^2 = O_p(N^{-\alpha_2})$, *then* $\frac{1}{N}\sum_i(\hat{a}_i - a_i)(\hat{b}_i - b_i) = O_p(N^{-\alpha_2-\alpha_1})$

Proof. The proof follow directly from the Cauchy-Schwarz:

$$\frac{1}{N}\sum_i(\hat{a}_i - a_i)(\hat{b}_i - b_i) \leqslant \sqrt{\frac{1}{N}\sum_i(\hat{a}_i - a_i)^2}\sqrt{\frac{1}{N}\sum_i(\hat{b}_i - b_i)^2} \qquad \text{(A.88)}$$

□

Lemma A.4.7 (Convergence rate for double sums)**.** *With* $h = O(h^{-r}), 1/12 < r < 1/8$*, then with* $\hat{P} \equiv \hat{f}_{ki}/\hat{g}_{ki}$ *and* $\hat{M}$ *as (i)* $\hat{g}$ *or (ii)* $\hat{w} \equiv \frac{\partial \hat{P_{k,i}}}{\partial\theta}/\hat{P}$*:*

$$\Delta \equiv \sqrt{N}\frac{1}{N}\sum_i(P - \hat{P})(M - \hat{M}) = o_p(1)$$

Proof. To show (i), note that $\frac{1}{\sqrt{N}}\sum_i(\hat{f}-P\hat{g})(g-\hat{g})/g\hat{g} \leqslant \frac{1}{g\inf(g)\sqrt{N}}\sum_i(\hat{f}-P\hat{g})(g-\hat{g})$. Therefore it suffices to show that $\frac{1}{\sqrt{N}}\sum_i(\hat{f}-P\hat{g})(g-\hat{g}) = o_p(1)$ since $\inf(g)$ is bounded below from zero due to trimming. From the Cauchy-Schwarz inequality:

$$
\begin{aligned}
\frac{1}{N}\sum_i(\hat{f}-P\hat{g})(g-\hat{g}) &\leqslant \sqrt{\frac{1}{N}\sum_i(\hat{f}-P\hat{g})^2\frac{1}{N}\sum_i(g-\hat{g})^2} \\
&= \sqrt{\underbrace{\frac{1}{N}\sum_i(\hat{f}-E[\hat{f}]+E[\hat{f}]-P\hat{g})^2}_{\text{This piece is studied in Lemma A.4.4}}\underbrace{\frac{1}{N}\sum_i(g-E[\hat{g}]+E[\hat{g}]-\hat{g})^2}_{\text{This piece is studied in Lemma A.4.3}}} \\
&= \sqrt{(O_p(h^8)+O_p(1/Nh^3))(O_p(h^4)+O_p(1/Nh^3)))} \\
&= o_p(N^{-1/2}) \quad \text{given that } 1/12 < r < 1/6
\end{aligned}
$$

To show (ii), apply (i) and note it suffice to show that $\frac{1}{\sqrt{N}}\sum(\hat{f}-P\hat{g})(w-\hat{w}) = o_p(1)$. Again from Cauchy-Schwarz, we have

$$
\begin{aligned}
\frac{1}{N}\sum_i(\hat{f}-P\hat{g})(w-\hat{w}) &\leqslant \sqrt{\frac{1}{N}\sum_i(\hat{f}-P\hat{g})^2\frac{1}{N}\sum_i(w-\hat{w})^2} \\
&= \sqrt{\underbrace{\frac{1}{N}\sum_i(\hat{f}-E[\hat{f}]+E[\hat{f}]-P\hat{g})^2}_{\text{This piece is studied in Lemma A.4.4}} \times \underbrace{O(h^8)+O(1/Nh^5)}_{\text{By Lemma 10 of Shen and Klein (2017)}}}
\end{aligned}
$$

Therefore the results will be $o_p(1)$, if $\left\{\begin{array}{l}[O_p(h^4)+O_p(1/N^{1/2}h^{3/2})]\times \\ [O_p(h^4)+O_p(1/N^{1/2}h^{5/2})]\end{array}\right\} = o_p(N^{-1/2})$. This condition is satisfied with $1/16 < r < 1/8$. □

Lemma A.4.8 (Residual Property)**.** *Under the index assumption:* $E[Y_i^k = 1|X_i] = E[Y_i^k = 1|V_i(\theta_0)]$, *we have* $E[\nabla_\theta E[Y_i^k = 1|V(\theta)|_{\theta=\theta_0}]] = 0$.

Proof. This property is stated and proved in Theorem 1 of Klein and Shen (2010), and the authors thank Whitney Newey for mentioning a key idea in a private communication. This property plays a key role in reducing the bias of $\hat{\theta}$. To be specific, recall that in analyzing the gradient term in (A.34), we show that the D_2 piece can be approximated by a U-statistic which looks like the sum of two terms T_1 and T_2. Using the law of iterated expectation after conditioning on the index V_j, T_2 has the form of:

$$T_2 = \sqrt{N_F} \sum_{g=1}^{N_F} \sum_{i \in g} E_{V_j}[E[\rho_{ji}|Obs_i, V_j]]/N \tag{A.89}$$

where $\rho_{ji} \equiv [Y_i^k - \hat{\Delta}_i(j) - P_j]K_{ji}\tau_{vj}w_j^k/g_j$ with $w_j^k \equiv \nabla_\theta E[Y_i^k = 1|V(\theta)|_{\theta=\theta_0}]/P_j$. Once conditional on the i'th observation and V_j, the inner expectation becomes

$$\{[Y_i^k - \hat{\Delta}_i(j) - P_j]K_{ji}\tau_{vj}/P_j g_j\} \underbrace{E[\nabla_\theta E[Y_i^k = 1|V(\theta)|_{\theta=\theta_0}]|obs_i, V_j]}_{=0 \text{ by the residual property}} \tag{A.90}$$

so it follows that T_2 is zero.

However, using this residual property as a bias control has a cost. Note that the expectation operator can slide through p_{ji} and hit w_j^k only if the trimming function τ is based on the index (We discussed the index-trimming and X-trimming in D.1). To perform index trimming, we need to estimate θ_0 in a two-stage process: first obtain a consistent estimate of θ_0, denote it as $\hat{\theta}_1$ and calculate the estimated index as $V(\hat{\theta}_1)$. Then, estimate θ_0 again but based the trimming on $V(\hat{\theta}_1)$. In Theorem 3.1, we show that the estimator in the second stage is asymptotically normal, but the first stage estimator (based on X-trimming) is not due to the bias from the U-statistic term. $\square$

Lemma A.4.9. *Refering to T_2 in (A.54), $\sqrt{N}T_2 = \sqrt{N}(U_1 + U_2) + o_p(1)$, where*

$$U_1 \equiv \frac{1}{\sqrt{N}(N-1)(N-2)} \sum_{j=1}^{N} \sum_{i \neq j} \sum_{s \neq i \neq j} \rho_{jis} \tag{A.91}$$

$$U_2 \equiv \frac{1}{\sqrt{N}} \sum_{j=1}^{N} E[t_{qj}|V_F, V_B, Z + \delta]\epsilon_j^\delta + o_p(1) \tag{A.92}$$

$$\rho_{jis} = \frac{\delta_j}{g_j} K_{ij}[\frac{1}{g_j}(Y_s K_{js} - P_j K_{js}) - \frac{1}{g_i}(Y_s K_{is} - P_j K_{is})] \tag{A.93}$$

Proof. To simplify notation, let $\delta_j = t_{qj}$ and $P_j = P_j^k(V_F, V_B, Z+\delta; \theta_0)$.

$$\begin{aligned}
\sqrt{N}T_2 &= \frac{\sqrt{N}}{N}\sum_{j=1}^{N}[\hat{P}_j - P_j]\delta_j && \text{(A.94)}\\
&= \frac{\sqrt{N}}{N}\sum_{j=1}^{N}[\frac{f_j}{g_j} - \frac{\hat{f}_j}{\hat{g}_j}]\delta_j && \text{(A.95)}\\
&= \frac{\sqrt{N}}{N}\sum_{j=1}^{N}[\hat{f}_j - P_j\hat{g}_j]\frac{\delta_j}{g_j} + o_p(1) && \text{(A.96)}\\
&= \frac{\sqrt{N}}{N(N-1)}\sum_{j=1}^{N}\sum_{i\neq j}[Y_i - \hat{\Delta}_i(j) - P_j]K_{ij}\frac{\delta_j}{g_j} && \text{(A.97)}\\
&= \frac{\sqrt{N}}{N(N-1)}\sum_{j=1}^{N}\sum_{i\neq j}[Y_i - \hat{P}_i + \hat{P}_j - P_j]K_{ij}\frac{\delta_j}{g_j} && \text{(A.98)}\\
&= \frac{\sqrt{N}}{N(N-1)}\sum_{j=1}^{N}\sum_{i\neq j}[P_i - \hat{P}_i + \epsilon_i^\delta + \hat{P}_j - P_j]K_{ij}\frac{\delta_j}{g_j} && \text{(A.99)}\\
&= \frac{\sqrt{N}}{N(N-1)}\sum_{j=1}^{N}\sum_{i\neq j}[\Delta_i(j) - \hat{\Delta}_i(j) + \epsilon_i^\delta]K_{ij}\frac{\delta_j}{g_j} && \text{(A.100)}
\end{aligned}$$

The equality of (A.95) and (A.96) follows from double convergence. For the transformation from (A.96)-(A.99), recall that we have a δ-localized model: (The same localized model but conditioning on Z+δ. Denote $V_i^\delta \equiv (V_{Fi}, V_{Bi}, Z_i + \delta)$)

$$\begin{aligned}
Y_i &= P(V_i^\delta) + \epsilon_i^\delta \quad \text{(Note: } \epsilon_i^\delta = Y_i - E[Y_i|V_i^\delta]) && \text{(A.101)}\\
&= P(V_j^\delta) + P(V_i^\delta) - P(V_j^\delta) + \epsilon_i^\delta && \text{(A.102)}\\
&\equiv P(V_j^\delta) + \Delta_i(j) + \epsilon_i && \text{(A.103)}
\end{aligned}$$

and a localized bias reducing estimator:

$$\begin{aligned}
\hat{P}(V_j^\delta) &\equiv \frac{\hat{f}_j}{\hat{g}_j} = \frac{\frac{1}{(N-1)h}\sum_{i\neq j}(Y_i - \hat{\Delta}_i(j))K_{ij}}{\frac{1}{(N-1)h}\sum_{i\neq j}K_{ij}} && \text{(A.104)}\\
K_{ij} &\equiv K(\frac{V_j^\delta - V_i^\delta}{h}) && \text{(A.105)}\\
\hat{\Delta}_i(j) &\equiv \hat{P}_i - \hat{P}_j \quad \text{(difference of first stage estimator)} && \text{(A.106)}
\end{aligned}$$

Proceed from (A.100), note that both $\hat{P}_i, \hat{P}_j$ could be estimated by the first stage kernel estimator, so the difference of Δs in (A.100) could be replaced by

$$\begin{aligned}
\Delta_i(j) - \hat{\Delta}_i(j) &= (\frac{f_j}{g_j} - \frac{\hat{f}_j}{\hat{g}_j}) - (\frac{f_i}{g_i} - \frac{\hat{f}_i}{\hat{g}_i}) && \text{(A.107)}\\
&= \frac{1}{g_j}(\hat{f}_j - P_j\hat{g}_j) - \frac{1}{g_i}(\hat{f}_i - P_j\hat{g}_i) && \text{(A.108)}\\
&= \frac{1}{(N-2)g_j}\sum_{s\neq i\neq j}(Y_sK_{js} - P_jK_{js}) && \text{(A.109)}\\
&- \frac{1}{(N-2)g_i}\sum_{s\neq i\neq j}(Y_sK_{is} - P_iK_{is}) && \text{(A.110)}\\
&= \frac{1}{(N-2)g_j}\sum_{s\neq i\neq j}C_{sj} - \frac{1}{(N-2)g_i}\sum_{s\neq i\neq j}C_{si} && \text{(A.111)}
\end{aligned}$$

So substitute the expression in (A.111) into (A.100) yields the following:

$$\begin{aligned}
\sqrt{N}U &= \sqrt{N}(U_1 + U_2) && \text{(A.112)}\\
\sqrt{N}U_1 &= \frac{\sqrt{N}}{N(N-1)(N-2)}\sum_{j=1}^{N}\frac{\delta_j}{g_j}\sum_{i\neq j}K_{ij}[\frac{1}{g_j}\sum_{s\neq i\neq j}C_{sj} - \frac{1}{g_i}\sum_{s\neq i\neq j}C_{si}]\\
\sqrt{N}U_2 &= \frac{\sqrt{N}}{N(N-1)}\sum_{j=1}^{N}\frac{\delta_j}{g_j}\sum_{i\neq j}\epsilon_iK_{ij}
\end{aligned}$$

□

Lemma A.4.10. *Refering to the proof in Lemma A.4.9,* $\sqrt{N}U_1 = o_p(1)$

Proof. With the definition of U_1 in (A.113), and

$$\rho_{jis} = \frac{\delta_j}{g_j}K_{ij}[\frac{1}{g_j}(Y_sK_{js} - P_jK_{js}) - \frac{1}{g_i}(Y_sK_{is} - P_jK_{is})] \qquad \text{(A.113)}$$

we have:

$$\sqrt{N}U_1 \equiv \frac{1}{\sqrt{N}(N-1)(N-2)}\sum_{j=1}^{N}\sum_{i\neq j}\sum_{s\neq i\neq j}\rho_{jis} \qquad \text{(A.114)}$$

letting

$$\rho^*_{jis} = [\rho_{jis} + \rho_{ijs} + \rho_{isj} + \rho_{sij} + \rho_{jsi} + \rho_{sji}]/6 \tag{A.115}$$

we can write (A.114) as a centered U-statistic:

$$\sqrt{N}U_1 = \sqrt{N}\binom{N}{3}^{-1}\sum_{j=1}^{N}\sum_{i\neq j}\sum_{s\neq i\neq j}\rho^*_{jis} \tag{A.116}$$

From the approximation theory of Serfling (2009) and Powell *et al.* (1989),

$$\sqrt{N}(U_1 - \hat{U}_1) = o_p(1) \tag{A.117}$$

$$\hat{U}_1 = \frac{1}{N}\sum_{j=1}^{N} E[\rho^*_{jis}|obs_j] \tag{A.118}$$

we show $\sqrt{N}U_1$ is $o_p(1)$ through the following steps:

(a) $E[\rho_{jis}|obs_j] = o(N^{-1/2})$

(b) $E[\rho^*_{jis}] = o(N^{-1/2})$

(c) $\sqrt{N}\hat{U}_1$ is $o_p(1)$

For (a), we may write $E[\rho_{jis}|obs_j]$ as $E_1 - E_2$ where

$$E_1 \equiv E\{\frac{\delta_j}{g_j}K_{ij}\frac{1}{g_j}(Y_sK_{js} - P_jK_{js})|obs_j\} \tag{A.119}$$

$$E_2 \equiv E\{\frac{\delta_j}{g_j}K_{ij}\frac{1}{g_i}E[(Y_sK_{is} - P_iK_{is})|V_i]|obs_j\} \tag{A.120}$$

For the interior expectation in E_2 term, we have:

$$\begin{aligned} E[(Y_sK_{is}-P_iK_{is})|V_i] &= \underset{V_s}{E}\,E[(Y_sK_{is}-P_iK_{is})|V_i,V_s] \\ &= \underset{V_s}{E}(E[Y_s|V_s]K_{is}-P_iK_{is})|V_s] \\ &= \int(P_s-P_i)K(\frac{V_i-V_s}{h})g(V_s)d(V_s) \\ &= \int(P(V_i+zh)-P(V_i))K(z)g(V_i+zh)d(z) \\ &= [\frac{g(V_i)sup_w(\nabla^2P(w))}{2}+sup_w(\nabla P(w)\nabla g(w))]h^2+O(h^4) \end{aligned} \tag{A.121}$$

where ∇ and ∇^2 denote a first(second) derivative taken with respect to V_i. Substituting this expression into (A.120) yields:

$$\begin{aligned} E_2 &= E\{\frac{\delta_j}{g_j}K_{ij}\frac{1}{g_i}[[\frac{g(V_i)sup_w(\nabla^2P(w))}{2}+sup_w(\nabla P(w)\nabla g(w))]h^2+O(h^4)]|obs_j\} \\ &= \frac{\delta_j}{g_j}E\{K_{ij}\frac{1}{g_i}[[\frac{g(V_i)sup_w(\nabla^2P(w))}{2}+sup_w(\nabla P(w)\nabla g(w))]h^2+O(h^4)]|obs_j\} \\ &= h^2\frac{\delta_j}{g_j}[\frac{g(V_j)sup_w(\nabla^2P(w))}{2}+sup_w(\nabla P(w)\nabla g(w))]+O(h^4) \end{aligned}$$

Turning to the E_1 term, through a similar derivation, we shall have:

$$\begin{aligned} E_1 &= E\{\frac{\delta_j}{g_j}K_{ij}\frac{1}{g_j}|obs_j\}E\{Y_sK_{is}-P_jK_{is}|obs_j\} \\ &= \frac{\delta_j}{g_j^2}[g_j+h^2\frac{sup_w(g''(w))}{2}+O(h^4)]* \\ &\quad [[\frac{g(V_j)sup_w(\nabla^2P(w))}{2}+sup_w(\nabla P(w)\nabla g(w))]h^2+O(h^4)] \\ &= h^2\frac{\delta_j}{g_j}[\frac{g(V_j)sup_w(\nabla^2P(w))}{2}+sup_w(\nabla P(w)\nabla g(w))]+O(h^4) \end{aligned}$$

Therefore we have $E_1-E_2=O(h^4)=o(N^{-1/2})$ with $h=O(N^{-1/7.99})$.

For (b), the proof follows immediately from (a) and the fact that all terms in ρ^*_{ijk} have the same unconditional expectations. For (c), since (b) would imply that $E[\sqrt{N}U_1]=o_p(1)$ and it can be shown that $Var(\sqrt{N}U_1)=o_p(1)$, the result

$\sqrt{N}U_1 = o_p(1)$ now follows. □

Lemma A.4.11. *the U_2 piece :* $\sqrt{N}U_2 = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} E[t_{qj}|V_j^\delta]\epsilon_j^\delta + o_p(1)$

Proof. Refer to U_2 in (A.113), defining $\rho_{ij} \equiv \frac{\delta_j}{g_j}\epsilon_i^\delta K_{ij}$, we have:

$$\sqrt{N}U_2 = \frac{1}{\sqrt{N}(N-1)}\sum_{j=1}^{N}\sum_{i\neq j}\frac{\delta_j}{g_j}\epsilon_i K_{ij} \quad \text{(A.122)}$$

$$= \sqrt{N}\binom{N}{2}^{-1}\sum_{j=1}^{N}\sum_{i\geqslant j}[\rho_{ji}+\rho_{ij}]/2 \quad \text{(A.123)}$$

$$= \frac{2}{\sqrt{N}}\sum_{j=1}^{N} E[\rho_{ji}+\rho_{ij}|Y_j, X_j]/2 + o_p(1) \quad \text{(A.124)}$$

$$= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} E[\rho_{ij}|Y_j, X_j] + o_p(1) \quad \text{(A.125)}$$

$$= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} E[\frac{\delta_i}{g_i}K_{ji}|Y_j, X_j]\epsilon_j^\delta + o_p(1) \quad \text{(A.126)}$$

$$= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} E[\frac{E[\delta_i|V_i^\delta]}{g_i}K_{ji}|Y_j, X_j]\epsilon_j^\delta + o_p(1) \quad \text{(A.127)}$$

$$= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} E[\delta_j|V_j^\delta]\epsilon_j^\delta + o_p(1) \quad \text{(A.128)}$$

The third equality follows from standard U-statistics projection theorem. (A.127) follows from an iterated expectation conditioning on V_i. The last equality follows from a Taylor expansion on $E[\delta_i|V_i]$. When defining the localized model, recall that we let $\epsilon_j^\delta = Y_j - E[Y_j|V_j^\delta]$, therefore:

$$\sqrt{N}U_2 = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} E[\delta_j|V_j^\delta][Y_j - E[Y_j|V_j^\delta]] + o_p(1) \quad \text{(A.129)}$$

□

Appendix B

Supplemental Materials to Ordered Response Models with Unobserved Correlated Thresholds

B.1 Evidences of Endogenous Characteristics

In this part, we will first provide the definition and identification results of structural and nonstructural rating probability functions. We also show that the disparity between the two functions could be seen as evidence of endogenous selection of bond characteristics. And this point is further confirmed with our data.

B.1.1 Structural and Nonstructural Probabilistic Rating Functions

We first define the non-structural conditional rating probability function in Eq. (B.1) for any $v \in \mathbb{R}$

$$P_j^n(v) \equiv \Pr(Y_i \leqslant j | V_{0i} = v) = \int \{v \leqslant T_j\} dF_{T_j|V_0=v}(t), \quad j = 0, 1, \cdots, J-1 \quad \text{(B.1)}$$

where $F_{T_j|V_0=v}(\cdot)$ is the conditional cumulative distribution function of T_j given the risk index $V_0 = v$. $P_j^n(\cdot)$ measures the probability of being rating equal or above notch j given some true risk index. However, $P_j^n(\cdot)$ is non-structural as the marginal effects of changes in bond characteristics on the this probability are confounded with the effect from changes of conditional distribution functions. This confoundedness is attributed to the dependency between the risk index and thresholds that consist of the hidden adjustment at bond level. Those probabilities and probability functions of this type are always computed in empirical works. However, its validity of predictions and causal effects would not generally be true.

This paper suggests a more interesting object, which only capture the partial effect on the probabilities due to the change in V_{0i} while holding the thresholds distributions fixed. This effect is summarized by the structural cumulative conditional rating probability function in Eq. (B.2) given $V_{0i} = v$,

$$P_j^s(v) \equiv \Pr(v \leqslant T_j) = \int \{v \leqslant T_j\} dF_{T_j}(t), \quad j = 0, 1, \cdots, J-1 \tag{B.2}$$

where $F_{T_j}(\cdot)$ is the marginal distribution of T_j. $P_j^s(v)$ corresponds to the average structural functions considered in Blundell and Powell (2004); Imbens and Newey (2009). In our example, $P_j(v)$ calculates the probability of being rated less than or equal to notch j, holding the threshold distribution unchanged for some default risk v. For models with only exogenous variables, $P_j^s(v)$ and $P_j^n(v)$ coincide with each other but diverge if the soft adjustment considered here does exist and its effect cannot be ignored.[1] Based on this observation, one can even have a test for the endogeneity.

The identification result of average structural functions for nonseparable models often relies critically on a large support condition in A-I.3.

A-I.3 Large Support. $\mathcal{R} = \mathcal{R}^v, \quad \forall v \in \mathbb{R},$ *a.s. where* $\mathcal{R} =$ *supp*$(R_i), \mathcal{R}^v =$ *supp*$(R_i | V_{0i} = v)$.

A-I.3 requires the conditional support of R_i to be the same as the unconditional support. This large support condition is often invoked in the control function literature to obtain point identification results of average structural functions. We require A-I.3 only for the point identification of $P_j^s(\cdot)$. But for identification of index parameters and soft

[1]The structural probability function of being rated exactly at notch j given $V_0 = v$ can be obtained straightforwardly by $\Pr(T_{j-1} < v \leqslant T_j) = P_j^s(v) - P_{j-1}^s(v), \;\; j = 0, 1 \cdots, J$.

adjustment alone, A-I.3 is not necessary. Proposition A states that $P_j(v)$ in Eq. (B.2) can be identified if the large support condition is invoked. The identification is achieved by marginally integrating out R_i for each index value $v \in \mathbb{R}$. The proof is standard and we leave it in the appendix.

Under Assumption A-I.1 and A-I.3, $P_j^s(v)$ are identified for each $v \in \mathbb{R}$ and $j \in \mathcal{Y}$.

Proof of Proposition A. The proof resembles the line of reasoning in Blundell and Powell (2003) and Imbens and Newey (2009) for the identification of average structural functions.

□

Finally, consider estimators of the structural and non-structural conditional probability functions defined in Eq. (B.1) and Eq. (B.2). Proposition A shows that $P_j^s(v)$ can be identified by integrating the conditional expectation function with respect to the CDF of R_i. Substitution with their consistent estimators gives us the estimators, $\widehat{P}_j^s(v)$, for each j. Consider the partial mean estimator in Eq. (B.3).

$$\widehat{P}_j^s(v) = \frac{1}{N}\sum_{i=1}^{N}\widehat{P}_j(v, R_i(\widehat{\alpha})) \tag{B.3}$$

In contrast, the nonstructural conditional probability functions can be straightforwardly estimated as the conditional expectation function in Eq. (B.4) in which $\widehat{E}$ denotes the local constant estimator

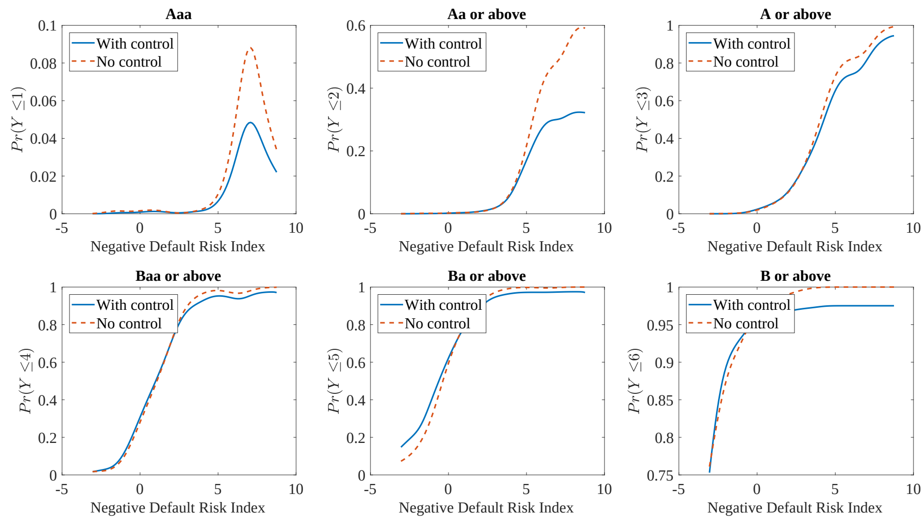
$$\widehat{P}_j^n(v) = \widehat{E}(\{Y_i \leqslant j\}|V_i(\widehat{\beta}) = v) \tag{B.4}$$

In the next section, we plot $\widehat{P}_j^n(v)$ against $\widehat{P}_j^s(v)$ to empirically examine the endogeneity issue of bond characteristics.

B.1.2 Empirical Evidence of Endogenous Characteristics

The key implication of our structural bond rating model is that some firm and bond characteristics are endogenously determined as issuers tend to form expectations of CRAs' soft adjustment conditional on its own private information. If this is true, then traditional estimators without taking into account the omitted soft adjustment could be biased. But the validity of above story remains to be verified by empirical evidences. In Figure B.1, we plot the cumulative structural (in solid) and non-structural (in dash) rating probability functions, defined in Eq. (B.2) and (B.1), for each category. To produce the graph, we have to estimate both the risk parameter and control relationship index first and then evaluate the conditional probability function nonparametrically. Figure B.1 is generated using the entire sample. For example of Aa or above, shown in the middle of the first row, the probabilities of being rated to Aa or above are drawn against the negative default risk index. In the case of no endogenous bond characteristics, one curve should match the other perfectly. However, this observation is obviously untrue for most of our subplots, but Baa or Ba above. For Aaa, the two functions diverge apart when the negative default risk index is large enough. For Aa or above, the nonstructural one explodes rapidly, as opposed to the more stable structural function. In general, the nonstructural tend to overestimate the probability at large index values. A formal test is unnecessary here because rejections are not hard to find as long as at least one of the categories does depart largely at some index value. Finally, one caveat of this plot: the converges of curves is the sufficient but not necessary condition for the presence of endogenous financial characteristics. Therefore, one cannot rule out the possibility of endogenous bond attributes even for subplots Baa or Ba above.

Figure B.1: Structural and Nonstructural Conditional Rating Probabilities



Note: 1. Data range is 2000-2016. 2. Estimates represent normalized coefficient ratios with respect to log of asset and Mshare, respectively for financial and control parameters. 3. Oprobit-R is estimated by MLE. Semi-X and semi-R are estimated by pseudo-MLE. 4. The rule-of-thumb bandwidths, $h = 1.06 s.d.(R) N^{-r}$ are used, with the optimal rate i.e. $r = 1/6$ for double index models. 5. Standard errors are in parentheses. 6. Significant level: *10 percent, **5 percent, ***1 percent.

B.2 Asymptotic properties

We now discuss the asymptotic properties for the two-stage estimators of both $\widehat{\theta}$ and $\widehat{\Delta}(r)$ for each r in the support. In particular, the theorem below presents consistency results on index parameter estimators and conditional relative thresholds estimators. Asymptotic assumptions A-A.1 to A-A.6 are all standard in the non/semi-parametric literature.

A-A.1. DGP. $\{(Y_i, X_i', R_i', \boldsymbol{T}_i')\}_{i=1}^N \in (\mathcal{Y}, \mathcal{X}, \mathcal{R}, \mathcal{J})$ is an i.i.d. vector of random variables defined on a complete probability space $(\Omega, \mathcal{F}, P)$, where (Y_i, X_i', R_i') are observed and $\boldsymbol{T}_i'$ are unobserved.

A-A.2. Smoothness. For each $j \in \mathcal{Y}$ and $(v, r) \in \mathbb{R} \times \mathcal{R}$, $0 < P_j(v, r) < 1$. The CDF F_V and F_R has the uniformly continuous and bounded Radon-Nikodym second order density derivatives with respect to Lebesgue measure. i). f_V is continuous in v and $f_{V|R}$ is continuous in (v, r). ii). There exists $C > 0$ such that $\inf_{\mathbb{R}_0} f_V > C$ and $\inf_{\mathbb{R}_0} f_{V|R} > C$.

A-A.3. Dominance. For each $j \in \mathcal{Y}$ any $r \in \mathcal{R}$, $P_j(\cdot, r)$ has all partial derivatives up to 3rd order. Let $\nabla^l P_j(v,r) \equiv \frac{\partial^l P_j(v,r)}{(\partial v)^l}$ where $l = 1, 2, 3$. $\nabla^l P_j(\cdot, r)$ is uniformly bounded and Lipschitz continuous on $\mathbb{R}$: for all $v, \widetilde{v} \in \mathbb{R}$, $|\nabla^l P_j(v,r) - \nabla^l P_j(\widetilde{v}, r)| \leqslant C||v - \widetilde{v}||$, for some constant $C > 0$, where $|| \cdot ||$ is the Euclidean norm.

A-A.4. Kernel. For some integer ν, the univariate symmetric kernel function $k : \mathbb{R} \to (0,1)$, satisfies i). $\int u^i k(u)du = \delta_{i0}$, for $i = 0, 1, \cdots, \nu - 1$, where δ_{ij} is the Kronecker's delta. ii). $\int u^\nu k(u)du < \infty$. iii). $k(u) = O((1 + u^{1+u+\varepsilon})^{-1})$ for some $\varepsilon > 0$.

A-A.5. Bandwidth. As $N \to \infty$, then $h_i \to 0$, $Nh_i^4 \to \infty$, for $i = 1, 2$, $\sqrt{N}h_1^6 \to 0$ and $\sqrt{Nh_2}h_2^4 \to 0$.

A-A.6. Parameter space. $\theta_0 \in \Theta_0$, where Θ_0 is the interior of the compact support Θ.

A-A.1 reiterates the data generating process. We do not need X and R to be compactly supported as the trimming indicator will guarantee the density denominators away from 0. A-A.2 and A-A.3 are regularity conditions usually appearing in nonparametric estimators. They indicate that densities and conditional expectations are smooth enough and have partial derivatives up to 3rd order with respect to the index V. A-A.4 is standard in kernel estimation. In this paper, the second-order kernels, $\nu = 2$, mostly suffices to reduce the asymptotic bias. A-A.5 concerns bandwidths and window parameters. A-A.6 restricts support of the finite and infinite-dimensional parameters to be compact given point identification.

Theorem B.2.1 (Consistency)**.** *Under Assumption I.1-I.2 and A.1-A.6, then for any*

$\epsilon > 0$, *as* $N \to \infty$, *it follows that*

$$a). \qquad N^{1/2-\epsilon}|\widehat{\theta} - \theta_0| = o_p(1)$$

$$b). \qquad (Nh_2)^{1/2-\epsilon}|\widehat{\Delta}_{j,j-1}(r) - \Delta_{j,j-1}(r)| = o_p(1), \quad j = 0, 1 \cdots, J-2$$

To conserve space, we do not reiterate the proof of Theorem B.2.1 which can be found in Klein and Sherman (2002). The proof of b). is based on the functional Delta method approach. Our proof resembles Theorem 5.1 in Altonji and Matzkin (2005).

$$\widehat{\Delta}_{j,j-1}(r) = \frac{1}{N}\sum_{i=1}^{N}\left[\widehat{P}_{j-1}^{-1}\left(\widehat{P}_j(V_i(\widehat{\beta}), r), r\right) - V_i(\widehat{\beta})\right], \quad j \in \{1, \cdots, J-1\}$$

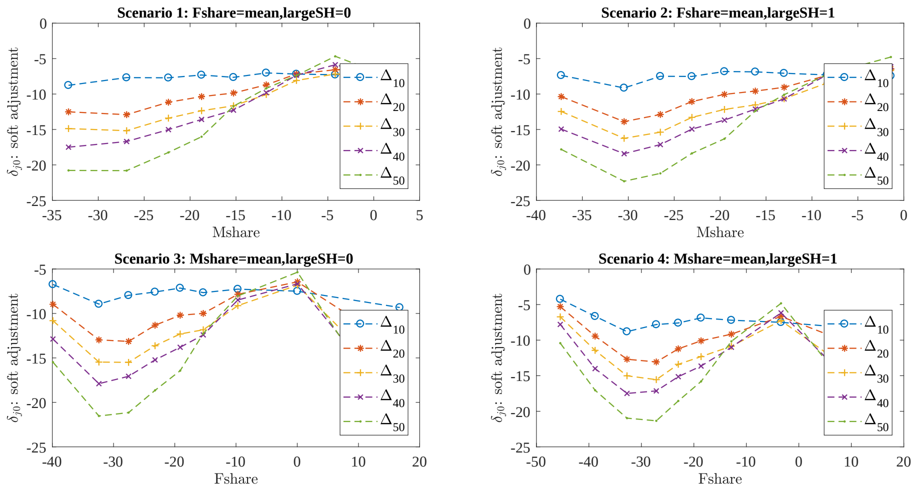
A remark on bandwidth selection. As in Assumption A-A.5, bandwidths are allowed to be different for estimating $\widehat{\theta}$ and $\widehat{\Delta}(r)$. In order to eliminate the bias in the theorem, bias-reducing techniques can be apply (Klein and Shen, 2010; Shen and Klein, 2017, etc.). A nice finding shows that the limiting variances of the relative thresholds estimators do not depend on the variability of the first-stage index estimators because the latter converge at a faster $\sqrt{N}$ rate than the nonparametric rate $\sqrt{Nh_2}$ for the second-stage estimators. This property permits us to analyze the second stage variability separately from the index estimators.

For single index control index, $R = R_i'\alpha_0$, the above asymptotic results would still apply when replacing R_i with the estimated $R_i'\widehat{\alpha}$, given the fact that $\widehat{\alpha}$ is root-N consistent.

B.3 Pattern of Thresholds over Shareholding Relations

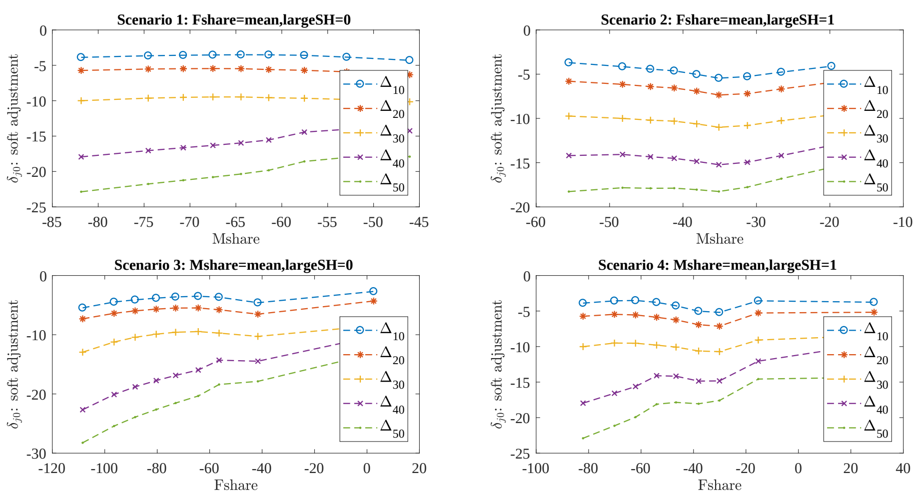
In the upper panels of Figure B.2 and Figure B.3, we plot the estimated relationship between the investment liaison with Moody or Mshare, and the soft adjustment, holding Fshare at mean, under situations of no large shareholder (in left panel) and at least one large shareholder (in right panel). Then we switch Mshare and Fshare in the lower panels. We first look at Figure B.2. For scenario 1, after fixing the investment structure in firms and removing large shareholders, we lose the pattern completely. For example, for the thresholds Δ_{10}, it almost stays the same and invariant with respect to Mshare. While in scenario 2, the reductions of thresholds start to emerge if at least one large shareholder exists. This highlights the importance and contribution of influential shareholders to the conflict-of-interest effect. Scenario 3 and 4 illustrate that the dominant factor in the relationship is the total investment shares in bond-issuers since Fshare drives the primary shapes of soft adjustment. The difference in Fshare might reflect the hidden unobserved qualities of the bonds. Furthermore, this effect is nonlinear. In particular, bonds of around median investment by common shareholders obtains the most favorable treatment from the CRA. Note that this effect is further magnified by having some influential shareholders in presence, according to scenario 4. In the upper panels of Figure B.3, the pattern is completely gone. It indicates that even for a fairly strong connection with the CRA, no obvious soft adjustment is given after the Dodd-Frank Act. Moreover, the partial effect of having large shareholders is only minimal. In the lower panels, we can also conclude that the primary factor driving the soft adjustment is Fshare, though the pattern is obvious for the top four notches. An interesting observation is the increasing relationship for the lower two categories, Δ_{40} and Δ_{50}. This may be due to the speculative nature of those bonds. They might have undergone stricter evaluations and required more tightened criteria if more investors want to speculate on their stocks.

Figure B.2: Estimated Relationship in Various Scenarios before the Dodd-Frank Act



Note: 1. Sample period: 2000-2010. 2. Y-axis plots the estimated soft adjustment as conditional mean thresholds relative to Aaa level. 3. X-axis plots various percentiles of Mshare or Fshare. 4. largeSH=0: no large shareholders; largeSH=1: at least one large shareholders. Fshare or Mshare=mean: fixed at mean.

Figure B.3: Estimated Relationship in Various Scenarios after the Dodd-Frank Act



Note: 1. Sample period: 2011-2016. 2. Y-axis plots the estimated soft adjustment as conditional mean thresholds relative to Aaa level. 3. X-axis plots various percentiles of Mshare or Fshare. 4. largeSH=0: no large shareholders; largeSH=1: at least one large shareholders. Fshare or Mshare=mean: fixed at mean.

Bibliography

ABREVAYA, J., HSU, Y.-C. and LIELI, R. P. (2015). Estimating conditional average treatment effects. *Journal of Business & Economic Statistics*, **33** (4), 485–505.

ADAMS, R. B. and FERREIRA, D. (2007). A theory of friendly boards. *The Journal of Finance*, **62** (1), 217–250.

ADMATI, A. R. and PFLEIDERER, P. (2009). The wall street walk and shareholder activism: Exit as a form of voice. *Review of Financial Studies*, p. hhp037.

AHN, H., ICHIMURA, H., POWELL, J. L. and RUUD, P. A. (2017). Simple estimators for invertible index models. *Journal of Business & Economic Statistics*, (just-accepted).

AI, C. and CHEN, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, **71** (6), 1795–1843.

ALESSI, C., WOLVERSON, R. and SERGIE, M. A. (2013). The credit rating controversy. *Council on Foreign Relations*.

ALLEN, F. (1990). The market for information and the origin of financial intermediation. *Journal of financial intermediation*, **1** (1), 3–30.

ALP, A. (2013). Structural shifts in credit rating standards. *The Journal of Finance*, **68** (6), 2435–2470.

ALTMAN, E. I. and RIJKEN, H. A. (2004). How rating agencies achieve rating stability. *Journal of Banking & Finance*, **28** (11), 2679–2714.

ALTONJI, J. G. and MATZKIN, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, pp. 1053–1102.

BAGHAI, R. and BECKER, B. (2016). Non-rating revenue and conflicts of interest. *Swedish House of Finance Research Paper*, (15-06).

BAHADUR, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, **37** (3), 577–580.

BARCLAY, M. J. and HOLDERNESS, C. G. (1989). Private benefits from control of public corporations. *Journal of financial Economics*, **25** (2), 371–395.

BATES, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of financial studies*, **9** (1), 69–107.

BECKER, B. and MILBOURN, T. (2011). How did increased competition affect credit ratings? *Journal of Financial Economics*, **101** (3), 493–514.

BHATTACHARYA, P. (1967). Estimation of a probability density function and its derivatives. *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 373–382.

BLUME, M. E., LIM, F. and MACKINLAY, A. C. (1998). The declining credit quality of us corporate debt: Myth or reality? *The journal of finance*, **53** (4), 1389–1413.

BLUNDELL, R. and POWELL, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. *Econometric Society Monographs*, **36**, 312–357.

BLUNDELL, R. W. and POWELL, J. L. (2004). Endogeneity in semiparametric binary response models. *The Review of Economic Studies*, **71** (3), 655–679.

BOGLE, J. C. (2005). *The battle for the soul of capitalism*. Yale University Press.

BOLTON, P., FREIXAS, X. and SHAPIRO, J. (2012). The credit ratings game. *The Journal of Finance*, **67** (1), 85–111.

CAMPBELL, J. Y. and TAKSLER, G. B. (2003). Equity volatility and corporate bond yields. *The Journal of Finance*, **58** (6), 2321–2350.

CANTOR, R. and MANN, C. (2006). Analyzing the tradeoff between ratings accuracy and stability. *Journal of Fixed Income, September*.

—, PACKER, F. *et al.* (1994). The credit rating industry. *Quarterly Review*, (Sum), 1–26.

CHEN, S. (1999). Semiparametric estimation of a location parameter in the binary choice model. *Econometric Theory*, **15** (01), 79–98.

— (2000). Rank estimation of a location parameter in the binary choice model. *Journal of Econometrics*, **98** (2), 317–334.

CHENG, G., HUANG, J. Z. *et al.* (2010). Bootstrap consistency for general semiparametric m-estimation. *The Annals of Statistics*, **38** (5), 2884–2915.

CORNAGGIA, J. and CORNAGGIA, K. J. (2013). Estimating the costs of issuer-paid credit ratings. *Review of Financial Studies*, **26** (9), 2229–2269.

CRAWFORD, V. P. and SOBEL, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pp. 1431–1451.

CREMERS, M., DRIESSEN, J., MAENHOUT, P. and WEINBAUM, D. (2008). Individual stock-option prices and credit spreads. *Journal of Banking & Finance*, **32** (12), 2706–2715.

DAVID, H. (1981). Order statistics. *J. Wiley*.

EDDY, W. F. and HARTIGAN, J. (1977). Uniform convergence of the empirical distribution function over convex sets. *The Annals of Statistics*, pp. 370–374.

EDERINGTON, L. H. (1985). Classification models and bond ratings. *Financial review*, **20** (4), 237–262.

EDMANS, A. (2009). Blockholder trading, market efficiency, and managerial myopia. *The Journal of Finance*, **64** (6), 2481–2513.

FENGLER, M. R. (2006). *Semiparametric modeling of implied volatility*. Springer Science & Business Media.

FLORENS, J.-P., HECKMAN, J. J., MEGHIR, C. and VYTLACIL, E. (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica*, **76** (5), 1191–1206.

GOEL, A. M. and THAKOR, A. V. (2015). Information reliability and welfare: A theory of coarse credit ratings. *Journal of Financial Economics*, **115** (3), 541–557.

GRIFFIN, J. M. and TANG, D. Y. (2012). Did subjectivity play a role in cdo credit ratings? *The Journal of Finance*, **67** (4), 1293–1328.

HANSEN, B. E. (2009). Lecture notes on nonparametrics. *Lecture notes*.

HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, **84** (408), 986–995.

HAUSMAN, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, pp. 1251–1271.

—, LO, A. W. and MACKINLAY, A. C. (1992). An ordered probit analysis of transaction stock prices. *Journal of financial economics*, **31** (3), 319–379.

HE, J., QIAN, J. and STRAHAN, P. E. (2015). Does the market understand rating shopping? predicting mbs losses with initial yields. *The Review of Financial Studies*, **29** (2), 457–485.

HODERLEIN, S. and SHERMAN, R. (2015). Identification and estimation in a correlated random coefficients binary response model. *Journal of Econometrics*.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, **58** (301), 13–30.

HOROWITZ, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica: journal of the Econometric Society*, pp. 505–531.

— and HÄRDLE, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association*, **91** (436), 1632–1640.

HORRIGAN, J. O. (1966). The determination of long-term credit standing with financial ratios. *Journal of Accounting Research*, pp. 44–62.

HULL, J. and WHITE, A. (1987). The pricing of options on assets with stochastic volatilities. *The journal of finance*, **42** (2), 281–300.

ICHIMURA, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, **58** (1), 71–120.

— and LEE, L.-F. (1991). Semiparametric least squares estimation of multiple index models: single equation estimation. In *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. Cambridge*, pp. 3–49.

IMBENS, G. W. (2007). Nonadditive models with endogenous regressors. *Econometric Society Monographs*, **43**, 17.

— and NEWEY, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, **77** (5), 1481–1512.

JIANG, J. X., STANFORD, M. H. and XIE, Y. (2012). Does it matter who pays for bond ratings? historical evidence. *Journal of Financial Economics*, **105** (3), 607–621.

JORION, P. (1988). On jump processes in the foreign exchange and stock markets. *Review of Financial Studies*, **1** (4), 427–445.

—, SHI, C. and ZHANG, S. (2009). Tightening credit standards: the role of accounting quality. *Review of Accounting Studies*, **14** (1), 123–160.

KAMENICA, E. and GENTZKOW, M. (2011). Bayesian persuasion. *The American Economic Review*, **101** (6), 2590–2615.

KAPLAN, R. S. and URWITZ, G. (1979). Statistical models of bond ratings: A methodological inquiry. *Journal of Business*, pp. 231–261.

KEDIA, S., RAJGOPAL, S. and ZHOU, X. (2014). Did going public impair moody s credit ratings? *Journal of Financial Economics*, **114** (2), 293–315.

—, — and — (2017). Large shareholders and credit ratings. *Journal of Financial Economics*, **124** (3), 632–653.

KLEIN, R. and SHEN, C. (2010). Bias corrections in testing and estimating semiparametric, single index models. *Econometric Theory*, **26** (06), 1683–1718.

— and — (2016). Difference-in-difference bias control estimator.

—, — and VELLA, F. (2015). Estimation of marginal effects in semiparametric selection models with binary outcomes. *Journal of Econometrics*, **185** (1), 82–94.

— and SHERMAN, R. P. (2002). Shift restrictions and semiparametric estimation in ordered response models. *Econometrica*, **70** (2), 663–691.

— and SPADY, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, pp. 387–421.

— and VELLA, F. (2009). A semiparametric model for binary response and continuous outcomes under index heteroscedasticity. *Journal of Applied Econometrics*, **24** (5), 735–762.

KRAFT, P. (2014). Rating agency adjustments to gaap financial statements and their effect on ratings and credit spreads. *The Accounting Review*, **90** (2), 641–674.

— (2015). Do rating agencies cater? evidence from rating-based contracts. *Journal of Accounting and Economics*, **59** (2), 264–283.

LEE, L.-F. (1995). Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. *Journal of Econometrics*, **65** (2), 381–428.

LEWBEL, A. (1997). Semiparametric estimation of location and other discrete choice moments. *Econometric Theory*, **13** (01), 32–51.

— (2003). Ordered response threshold estimation. *Unpublished working paper, Boston College*.

LÜHRMANN, M. and MAURER, J. (2008). Who wears the trousers? a semiparametric analysis of decision power in couples. *MEA discussion papers*, **168**.

MANSKI, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of econometrics*, **27** (3), 313–333.

MATHIS, J., MCANDREWS, J. and ROCHET, J.-C. (2009). Rating the raters: are reputation concerns powerful enough to discipline rating agencies? *Journal of monetary economics*, **56** (5), 657–674.

MAURER, J., KLEIN, R. and VELLA, F. (2011). Subjective health assessments and active labor market participation of older men: evidence from a semiparametric binary choice model with nonadditive correlated individual-specific effects. *Review of Economics and Statistics*, **93** (3), 764–774.

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

MILLON, M. H. and THAKOR, A. V. (1985). Moral hazard and information sharing: A model of financial information gathering agencies. *The Journal of Finance*, **40** (5), 1403–1422.

MULLER, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *The Annals of Statistics*, pp. 766–774.

NADARAYA, E. (1965). On non-parametric estimates of density functions and regression curves. *Theory of Probability & Its Applications*, **10** (1), 186–190.

NADARAYA, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, **9** (1), 141–142.

PAKES, A. and POLLARD, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, pp. 1027–1057.

PARTNOY, F. (2009). Rethinking regulation of credit rating agencies: An institutional investor perspective. *Council of Institutional Investors, April*, pp. 09–014.

PETERSEN, M. A. (2004). Information: Hard and soft.

PINCHES, G. E. and MINGO, K. A. (1973). A multivariate analysis of industrial bond ratings. *The journal of Finance*, **28** (1), 1–18.

— and — (1975). The role of subordination and industrial bond ratings. *The Journal of Finance*, **30** (1), 201–206.

POWELL, J. L. (1994). Estimation of semiparametric models. *Handbook of econometrics*, **4**, 2443–2521.

—, STOCK, J. H. and STOKER, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, pp. 1403–1430.

RAMAKRISHNAN, R. T. and THAKOR, A. V. (1984). Information reliability and a theory of financial intermediation. *The Review of Economic Studies*, **51** (3), 415–432.

RENAULT, E. and TOUZI, N. (1996). Option hedging and implied volatilities in a stochastic volatility model1. *Mathematical Finance*, **6** (3), 279–302.

ROBINSON, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pp. 931–954.

SANGIORGI, F., SOKOBIN, J. and SPATT, C. (2009). *Credit-rating shopping, selection and the equilibrium structure of ratings*. Tech. rep., Working Paper, Stockholm School of Economics and Carnegie Mellon University.

SERFLING, R. J. (2009). *Approximation theorems of mathematical statistics*, vol. 162. John Wiley & Sons.

SHEN, C. and KLEIN, R. (2017). Recursive differencing for estimating semiparametric models.

SHLEIFER, A. and VISHNY, R. W. (1986). Large shareholders and corporate control. *The Journal of Political Economy*, pp. 461–488.

SIMAR, L. and WILSON, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, **136** (1), 31–64.

SKRETA, V. and VELDKAMP, L. (2009). Ratings shopping and asset complexity: A theory of ratings inflation. *Journal of Monetary Economics*, **56** (5), 678–695.

STOKER, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica: Journal of the Econometric Society*, pp. 1461–1481.

WEST, R. R. (1970). An alternative approach to predicting corporate bond ratings. *Journal of Accounting Research*, pp. 118–125.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pp. 1–25.

WHITE, L. J. (2002). The credit rating industry: An industrial organization analysis. In *Ratings, rating agencies and the global financial system*, Springer, pp. 41–63.