Semiparametric Estimation of a Credit Rating Model

Yixiao Jiang*

Abstract

This paper develops 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 choice 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 use the estimated model to 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. To obtain the statistical significance of these results, I prove a U-statistics equivalence result that implies asymptotic inference for a large class of semiparametric models.

Keywords: Credit Rating Agencies, Conflicts of Interest, Semiparametric-index Models, U-statistics

JEL Classification: C25, G24, G32

*Department of Economics, Rutgers University, 75 Hamilton St, New Brunswick, NJ 08854. Email: yixiao@economics.rutgers.edu. I thank my advisor, Roger Klein, for his advice and comments. I am also grateful to Simi Kedia, Diana Knyazeva (FMA discussant), Bruce Mizrach, John Landon-Lane, Hilary Sigman, Zhutong Gu, Shuyang Yang, Xiye Yang, Bin Xie, Celso Brunetti, Yuliyan Mitkov, Xu Tong (discussant), Geoff Clarke, Yuan Liao, Daniel Millimet, Tom Fomby, Saltuk Ozerturk, Lu Wang, and conference participants for their helpful discussions. I gratefully acknowledge financial support provided by the graduate school and economics department of Rutgers University. All errors remain mine.

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 to explicitly take into account 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 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,

[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 (He et al., 2015, Jiang et al., 2012, Mathis et al., 2009)

employed by Campbell and Taksler (2003), Jiang et al. (2012), Kraft (2015), 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 (Blume et al., 1998, Horrigan, 1966, Kaplan and Urwitz, 1979, 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 Härdle and Stoker (1989), Horowitz and Härdle (1996), Ichimura (1993), Klein and Sherman (2002), Klein and Spady (1993), Manski (1985), Powell et al. (1989). 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 $MFOI$. The contribution

[2] Higher order kernels (Ichimura and Lee, 1991, Lee, 1995, Muller, 1984) 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.

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 benefit (Admati and Pfleiderer, 2009, Edmans, 2009). 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 (Becker and Milbourn, 2011, Bolton et al., 2012, White, 2002): 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. In terms of predictive performance, I also find that the semiparametric model outperforms comparable parametric models at predicting initial ratings, particularly at significantly the high and low rating tails.

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 a 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.

[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.

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 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:

$$y^* = y^*(V, m, b) \tag{1}$$

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 (14) reflects that credit risk is driven by both

hard information and *soft information* in a potentially non-separable form[4]. For estimation, I treat the level of soft information, represented by m, as a regression error term.

3. Since the conflicts of interest measure b is neither observed nor directly measurable, I proceed as follows. Assuming that firms with a closer liaison with the CRA are exposed to larger conflicts of interest, I use an unknown function of the MFOI, a shared-ownership index that I introduce later in the study, for b.

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.

To empirically study the model in (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 3.1. Using the institutional shareholding data (13f) from Thomson Reuters, I construct a variable, MFOI, to characterize a bond issuer's shared-ownership relation with Moody's. Similarly, I assume that the conflict of interest b is an unknown function of $MFOI$, for example, $b = C(MFOI)$. This variable is formally defined in Section 3.2.

3.1 Firm and bond characteristics

Table 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 (Blume et al., 1998, Horrigan, 1966, Jiang et al., 2012, Kaplan and Urwitz, 1979, West, 1970). The explanatory variables are: (1) Firm leverage, defined as the ratio of long-term debt to total assets (LEVERAGE). (2) Operating

[4]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.

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 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

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 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.

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[5], and any subset of those shareholders can invest in an issuing firm. The key variable of interest, the MFOI, is defined as follows:

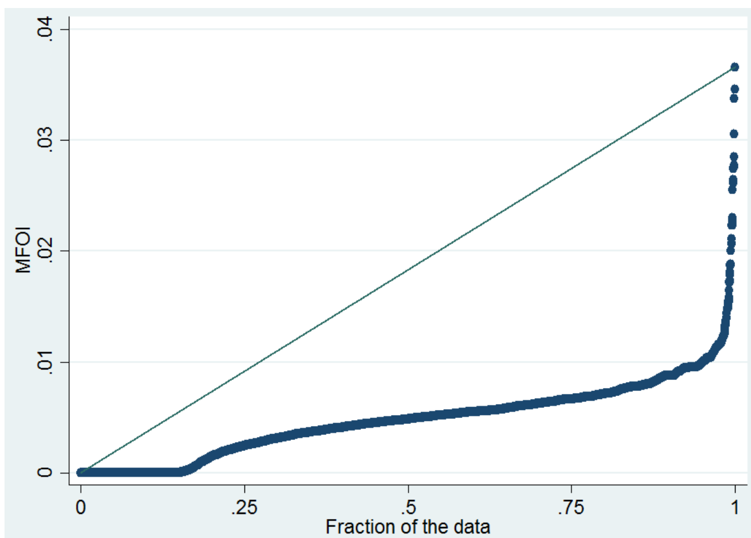
$$MFOI = \sum_{j=1}^{M} p_j \lambda_j \tag{2}$$

[5]Since all of the variable are time-specific, I drop the time t subscript for notational 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.

I plot the distribution of the MFOI in Figure 1. Since institutional investors hold diverse portfolios, most p_j and λ_j take on small values[6], 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).

Figure 1: Distribution of MFOI



4 Econometric Strategy

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

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

$Y_i = 1, 2, 3 \dots L$ based on y_i^* and a series of cutoff points c_j[7] between rating categories:

$$\begin{aligned} Y_i &= \sum_{j=1}^{L} j\mathbf{1}\{c_{j-1} < y_i^* < c_j\}, \\ y_i^* &= y^*(X_i, U_i) \\ \mathbf{1}\{E\} &: \text{ an indicator function of the event } E \end{aligned} \tag{3}$$

where U_i is a potentially multidimensional disturbance term representing the soft information. Motivated by the theoretical framework in (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,

$$Pr(Y_i = j|X_i) = P_j(X_i), \text{ for } j = 1, 2, \cdots, L \tag{4}$$

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:

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:*

$$Prob(Y_i = j|X_i) = Prob(Y_i = j|F_i\beta_0^F, B_i\beta_0^B, MFOI_i) \equiv H_j(V_F, V_B, MFOI_i) \tag{5}$$

[7] These 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.

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 CRA-issuer relation proxy $MFOI_i$. 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[8].

In the bond rating context, our consideration of a multiple-index formulation is motivated by the institutional evidence suggesting F_i and B_i can enter the rating model in a non-additive way. In a research manual, Moody analysts (Crosbie and Bohn, 2003) state that an investment's default risk y_i^* is related to its "expected loss," which is the product of two components: *probability of default (PD)* and *loss given default (LGD)*. Note that PD is only driven by the issuer fundamentals F_i (all bonds within the same firm have identical PD), whereas LGD is only driven by the issue-specific characteristics B_i, such as whether the investment is a senior debt. This product-form decomposition implies that F_i and B_i are likely to affect y^* in a non-additive fashion, which motivates the potential non-separability of $H_j(\cdot,\cdot,\cdot)$ and implies that a single index formulation may not be adequate.

4.2 Quantile marginal effect

Given the above model, it is convenient to define the impact of $MFOI$ on the rating as the cumulative change in (5) from a marginal increase in $MFOI_i$: for example, the probability of obtaining a better rating from a counterfactual change in the shareholding relation:

$$\begin{aligned} ME(F_i, B_i, m_b; \Delta, K) &\equiv Prob(Y_i < K|F_i, B_i, m_b + \Delta) - Prob(Y_i < K|F_i, B_i, m_b) \\ &= \sum_{j=1}^{K-1} H_j(V_F, V_B, m^b + \Delta) - H_j(V_F, V_B, m^b), \end{aligned} \tag{6}$$

[8]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 (5). 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 $m^b + \Delta$.

One object of empirical interest is the pattern of $ME(F_i, B_i, m_b; \Delta, K)$ for different values of m_b. 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

$$QME(\mathcal{Z}_q; K) \equiv E[ME(F_i, B_i, m_b; \Delta, K)|m_b \in \mathcal{Z}_q] \tag{7}$$

That is, the unit-level marginal effects defined in (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.

4.3 Estimation

Note that the function H_j in (5) 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 (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: More formally, the estimator is obtained by maximizing the following (log-) "quasi-likelihood:"

$$\hat{\theta} \in argmax\ Q(\theta) \equiv \sum_{i=1}^{N} \tau_i \{ \sum_{k=1}^{L} Y_i^k Ln(P_i^k(\theta)) \} \tag{8}$$

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:

$$\widehat{P^k(\theta)} = \frac{N^{-1} \sum_j (Y_i^k - \delta_j(V_i)) K_h(V_j - V_i)}{N^{-1} \sum_j K_h(V_j - V_i)} \tag{9}$$

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: 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

$$\widehat{QME(\mathcal{Z}_q; K, \hat{\theta})} \equiv \frac{\sum_{i=1}^{N} \hat{t}_{qi} \widehat{ME_i}(F_i, B_i, m_b; \Delta, K, \hat{\theta})}{\sum_{i=1}^{N} \hat{t}_{qi}} \tag{10}$$

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:

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

4.4 Inference

I also compute the large sample distribution of both $\hat{\theta}$ and $\widehat{QME(\mathcal{Z}_q; K, \hat{\theta})}$. 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 $\hat{\theta}$. I also defer the asymptotic theorems along with the full poofs in Appendix B.

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:

$$\begin{aligned} \sqrt{N}\hat{G}(\theta_0) &= \underbrace{N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{L}\tau_i[Y_i^k - E_i^k(\theta_0)]\nabla_\theta E_i^k(\theta)|_{\theta=\theta_0}\alpha_i}_{\text{A}} \\ &+ \underbrace{N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{L}\tau_i[E_i^k(\theta_0) - \widehat{E_i^k(\theta_0)}]\nabla_\theta E_i^k(\theta)|_{\theta=\theta_0}\alpha_i}_{\text{B}} + o_p(1) \end{aligned} \tag{11}$$

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 (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[9]. In Theorems 2 and 3, I derive the large sample distribution of $\hat{\theta}$ and $QM\widehat{E(\mathcal{Z}_q}; K, \hat{\theta})$.

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.

Table 2: 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.

5.1 Index parameter estimates and average marginal effects

In Table 2, 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 semi-parametric model is obtained based on the asymptotic covariance matrix derived in this paper (Theorem 2 in the Appendix)[10]. 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,

[9]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.

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

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).

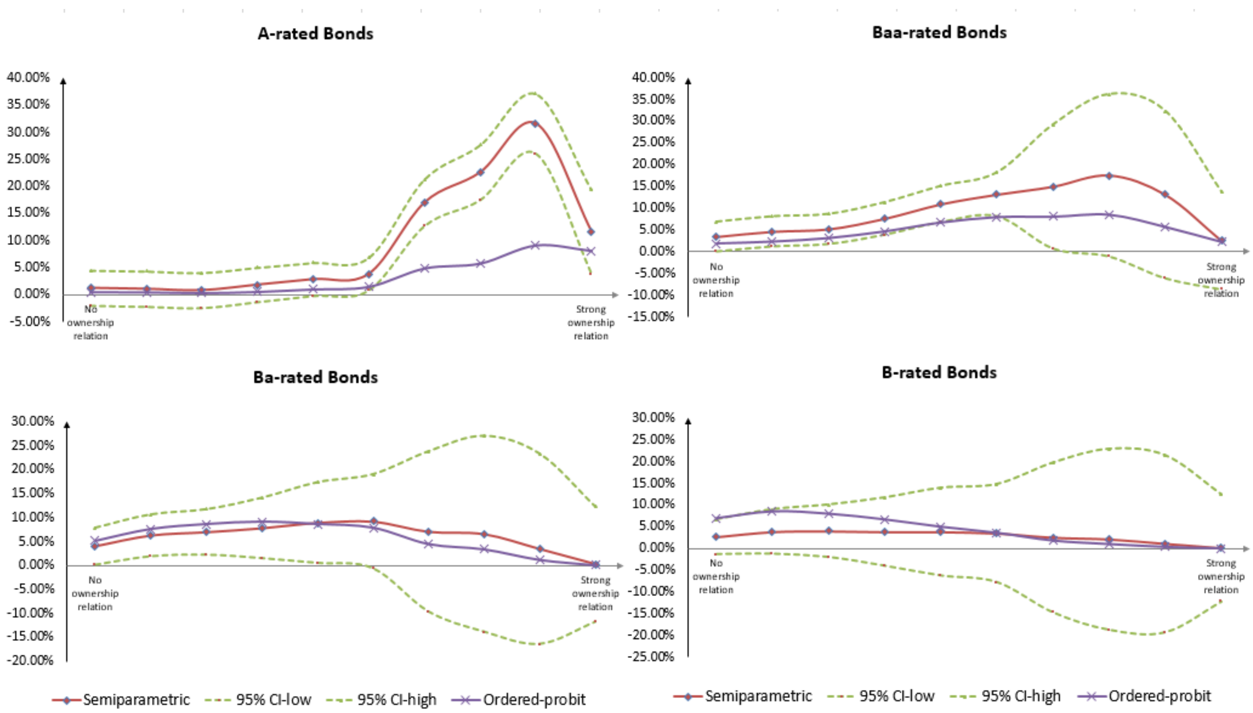
5.2 Quantile effects of MFOI

In Figure 2, 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 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

[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: 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.

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 2); 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 2, 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.

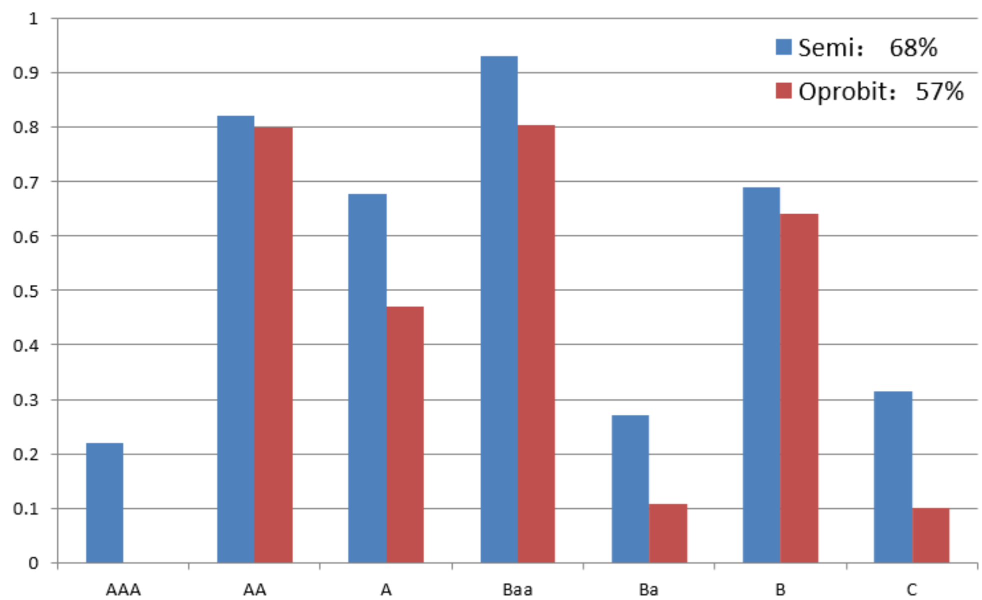
Capturing such heterogeneity is difficult in a parametric setting because of the constrained functional form. By comparing the ordered-probit estimates (purple line) and semiparametric estimates (red lines), the ordered-probit estimates are more "homogeneous": they vary between zero and 10 % (whereas 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 distributed random variable. In contrast, the semiparametric model allows the rating probability function in (5) to be category-specific.

5.3 Predictive results

In Figure 3, I compare the semiparametric model and ordered-probit model in terms of prediction accuracy for each rating category. Collectively, the semiparametric model correctly predicts 68 % of initial ratings, which is 10 percentage points higher than the ordered-probit model with the

[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 3: Predictive Results (Semiparametric vs Ordered-probit)



Note: - In this Figure, I compare the prediction performance of the semiparametric model and ordered-probit model. The blue (red) bar represents the percentage of correct prediction in each category for the semiparametric (ordered-probit) model.

same set of explanatory variables. In addition, the semiparametric model outperforms the other models in every rating category, particularly the Aaa, A, and Ba categories. This is not surprising; after assuming the latent default risk is a linear function of covariates, the ordered-probit model tends to perform poorly at the tails of the distribution. Despite a notable improvement by the semiparametric model, neither approach performs well in the *Ba* category, the category directly below the investment-grade/high-yield brink. In fact, most bonds rated as *Ba* in our data have a predicted rating of *Baa*. One explanation is that Moody's is being conservative on the boundary and is choosing to be harsh rather than overly optimistic.

In Table 3, I further compare the semiparametric model with other models in the bond rating literature[13]. In terms of the percentage of correct predictions, the semiparametric model outperforms West (1970), Horrigan (1966), and Blume et al. (1998). The semiparametric model has approximately the same overall predictive accuracy as Kaplan and Urwitz (1979). However, Kaplan and Urwitz (1979) performed poorly for *Aa*, *Ba*, and *B* bonds (all less than 25% correct)

[13]Note that for each previous work, the statistics on the percentage of correct predictions is directly imported from the corresponding paper. Therefore, I consider the comparison suggestive because the dataset and the explanatory variables used are different.

while the semiparametric model shows more robust predictive power across all rating categories.

Table 3: Comparison with Previous Models

	% of correct prediction							
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 investment grade bonds using S%P's ratings.
** - In Pinches and Mingo (1975), the authors use multiple discriminant analysis (MDA) instead of regular regression

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.

Appendix A 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[14] 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 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 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:

$$U^R(y, \pi) = -(y - \pi)^2 \tag{12}$$

[14] 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.

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

$$U^S(y, \pi, b) \;\; = \;\; -(y - \pi + b)^2. \tag{13}$$

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 1. 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[15]. 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.

A.2 Equilibrium

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

$$y^* \;\; = \;\; V + \sum_{i=1}^{N(b)} \frac{a_i(b) + a_{i+1}(b)}{2} 1\{a_i(b) \leq m < a_{i+1}(b)\} \tag{14}$$

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

$$a_i = \frac{i}{N(b)} + 2bi(N(b) - i), \;\; i = 0, 1, \cdots, N(b), \;\;\; a_0 = 0, a_N = 1 \tag{15}$$

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 $U \in (a_i, a_{i+1})$[16].

[15] 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.

[16] In fact, CS shows that the model has multiple equilibira for every $1 \leq N \leq N(b)$. 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

A.3 Implication

Looking at the equilibrium action in (14), 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[17]. 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 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 (15) 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 (14) and (15), b affects y^* through two aspects: the set of cutoff points $a_i(b)$ and the discontinuous mapping $N(b)$. In regions where a increasing 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.

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.

[17] 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

Appendix B 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 $MFOI$.

B.1 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:

$$\begin{aligned} Y_j^k &= P^k(V_j) + \epsilon_j \text{ with } \epsilon_j = Y_j^k - E[Y_j^k = 1|V_j] \\ &= P^k(v) + \underbrace{[P^k(V_j) - P^k(v)]}_{\Delta_j(v)} + \epsilon_j \end{aligned} \tag{16}$$

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:

$$\hat{I}^k(v) = argmin_\alpha \sum_j (Y_j^k - \alpha)^2 K_h(V_j - v) \tag{17}$$

$$\implies \hat{I}^k(v) = \frac{N^{-1}\sum_j Y_j^k K_h(V_j - v)}{N^{-1}\sum_j K_h(V_j - v)} \tag{18}$$

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:

$$\widehat{P}^k(v) \equiv \frac{N^{-1}\sum_j [Y_j^k - \hat{\Delta}_j(v)]K_j(v)}{N^{-1}\sum_j K_j(V_j - v)} = \hat{f}_1(v)/\hat{g}(v) \tag{19}$$

has a "better" convergence property than $\hat{I}^k(v)$ from Lemma B.1.

Lemma B.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, 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.

B.2 "Residual Property" of $\nabla_\theta E[Y_i^k = 1|V_i(\theta)]|_{\theta=\theta_0}$

Lemma B.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 $V(\hat{\theta}_1)$.

Appendix C Asymptotic results

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[18]. Based on this important result, in Theorem 2,3, I derive the large sample distribution of $\hat{\theta}$ and $QM\widehat{E(\mathcal{Z}_q; K, \hat{\theta})}$.[19]

Theorem C.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 (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 (11).

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

$$\widehat{I_i^k(\theta_0)} = \frac{N^{-1} \sum_j Y_i^k K_h(V_j - v_i)}{N^{-1} \sum_j K_h(V_j - v_i)} \equiv f_i^0 / g_i \tag{20}$$

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

[18] 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.

[19] A more detailed version can be accessed at https://www.dropbox.com/s/8huis14f871cpus/econometrics%20appendix.pdf?dl=0

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:

$$\begin{aligned}
\hat{g}(v_i,\theta)[B^* - B] &= \qquad (21) \\
N^{-1/2}\sum_i [\widehat{f^0(v_i,\theta_0)} - \widehat{f^1(v_i,\theta_0)}]\tau_i w_i &\leq \sqrt{N}\sup_v |[\hat{f}^0(v,\theta_0) - \hat{f}^1(v,\theta_0)]\tau_i w_i| = o_p(1)
\end{aligned}$$

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

$$\begin{aligned}
[\hat{f}^0(v) - \hat{f}^1(v)]\tau_i w_i &= \hat{f}^0(v)\tau_i w_i - E[\hat{f}^0(v)\tau_i w_i] - \hat{f}^1(v)\tau_i w_i + E[\hat{f}^1(v)\tau_i w_i] \\
&= \int_{V_j} f_{0j}(v,\theta_0)\tau_i w_i d[G_n(v) - G(v)] - \int_{V_j} f_{1j}(v,\theta_0)\tau_i w_i d[G_n(v) - G(v)] \\
&= \int_{V_j} [\tau_i(f_{0j}(v,\theta_0) - f_{1j}(v,\theta_0))w_i][dG_n(v) - dG(v)] \qquad (22) \\
&= \int_{V_j} [\tau_i \delta_j(v) K_h(V_j - v) w_i][dG_n(v) - dG(v)] \qquad (23)
\end{aligned}$$

Via integrating-by-parts, the above integral equals

$$\tau_i\delta_j(v)K_h(V_j - v)w_i[dG_n(v) - dG(v)]|_{V_j\in\partial\Omega} - \int_{V_j}[G_n(v) - G(v)]d[\delta_j(v)K_h(V_j - v)w(v)] \qquad (24)$$

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 $\sup_v |G_n(v) - G(v)|$[20] 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 (21) follows. That is, $sup_v|(\hat{f}^0(v,\theta_0) - \hat{f}^1(v,\theta_0))\tau_i w_i|| = o_p(N^{-1/2})$.

□

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

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

Proof. A taylor expansion of the gradient around θ_0 gives:

$$\sqrt{N}(\hat{\theta} - \theta_0) \quad = \quad -\hat{H}^{-1}(\theta^+)\sqrt{N}\hat{G}(\theta_0) \tag{25}$$

The average estimated Hession $\hat{H}(\theta^+)$ will converge to its expectation as N goes to infinity (*Law of large number*), which is a fix matrix that we termed H_0. It can be shown that with the recursive differencing estimator $\hat{P}_i^k(\theta)$ defined in (9), the estimated firm gradient converges to the sum of (i) a "true gradient" A, (ii) a bias term B and (iii) a series of $o_p(1)$ terms that will vanish asymptotically:

$$\sqrt{N}\hat{G}(\theta_0) = \underbrace{N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{L}\tau_i(Y_i - P_i^k(\theta_0))w_i}_{\text{A: can be handled by CLT}} + \underbrace{N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{L}\tau_i(\widehat{P_i^k(\theta_0)} - P_i^k(\theta_0))w_i}_{\text{B: cause bias}} + o_p(1)$$

where the weight function $w_i = \nabla_\theta P_i^k|_{\theta=\theta_0}/P_i^k(\theta_0)$.

From the established U-statistics equivalence result in Theorem 1, we have $g(v_i,\theta)[B - B^*] = o_p(1)$. Since estimated density $\hat{g}$ is bounded away from zero by the trimming function, $sup_v|\hat{g}(v)[B - B^*]| > c * sup_v|B - B^*| = o_p(1)$, for some constant $c = inf_v|\hat{g}(v)|$. Therefore we have $B = o_p(1)$. With H_0 as the probability limit for the hession, $\sqrt{N}(\hat{\theta} - \theta_0)$ has the following asymptotic linear form:

$$\sqrt{N}(\hat{\theta} - \theta_0) \quad = \quad H_0^{-1}\sqrt{N}\sum_{i=1}^{N} G_i(\theta_0)/N + o_p(1) \tag{26}$$

so the asymptotic normality follows from applying the Lindberg CLT. □

Theorem C.3 (**Normality for Quantile Marginal effects**). *Under A.1-A.5, we have*

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

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

$$\psi_{1j}^k \equiv \frac{E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z+\delta; \theta_0)] - E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z; \theta_0)]}{E[t_{qj}]} H_0^{-1} G(\theta_0) \tag{27}$$

$$\psi_{2j}^k \equiv \{\nabla_q E[t_{qj} m_j(\theta_0)] - \nabla_q E[t_{qj} QME_q^k]\} \frac{\mathcal{B}_j}{E[t_{qj}]}$$

$$\psi_{3j}^k \equiv \frac{E[t_{qj}|V_F, V_B, Z+\delta]\epsilon_j^\delta - E[t_{qj}|V_F, V_B, Z]\epsilon_j}{E[t_{qj}]}$$

$$\psi_{4j}^k \equiv \frac{t_{qj} m_j(\theta_0) - E[t_{qj}] QME_q^k}{E[t_{qj}]} - \frac{t_{qj} - E[t_{qj}]}{E[t_{qj}]} QME_q^k \tag{28}$$

Proof. 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

$$\sqrt{N}(QME_q^K - \widehat{QME_q^K}) = \sqrt{N}\sum_{k=1}^{K-1}(\widehat{M_q^k} - M_q^k) \tag{29}$$

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

$$\widehat{m_j^k} \equiv \widehat{P_i^k}(V_F, V_B, Z+\delta; \hat{\theta}) - \widehat{P_i^k}(V_F, V_B, Z; \hat{\theta}) \tag{30}$$

$$m_j^k \equiv P_i^k(V_F, V_B, Z+\delta; \theta_0) - P_i^k(V_F, V_B, Z; \theta_0) \tag{31}$$

$$\hat{N}_q \equiv \sum \hat{t}_{qj} \tag{32}$$

and we proceed with the following decomposition:

$$\begin{aligned}
\sqrt{N}(\widehat{M_q^k} - M_q^k) &\equiv \sqrt{N}(\frac{\sum_{j=1}^N \hat{t}_{qj}\hat{m}_j^k(\hat{\theta})}{\sum_{j=1}^N \hat{t}_{qj}} - \frac{\sum_{j=1}^N t_{qj}m_j^k}{\sum_{j=1}^N t_{qj}}) \\
&= \sqrt{N}(\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4) \qquad (33) \\
\text{where}\quad \Delta_1 &= (N/\hat{N}_q)\frac{1}{N}\sum_j \hat{t}_{qj}(\hat{m}_j^k(\hat{\theta}) - m_j^k(\theta_0)) \\
\Delta_2 &= (N/\hat{N}_q)\frac{1}{N}\sum_j m_j^k(\theta_0)(\hat{t}_{qj} - t_{qj}) \\
\Delta_3 &= (N/\hat{N}_q)(\frac{1}{N}\sum_j t_{qj}m_j^k(\theta_0) - E[t_{qj}]M_q^k) \\
\Delta_4 &= (N/\hat{N}_q)\frac{1}{N}\sum_j (E[t_{qj}] - \hat{t}_{qj})M_q^k
\end{aligned}$$

Loosely speaking, Δ_1 reflects the estimation uncertainly from the parameter θ_0. I show that Δ_1 can be characterized as the sum of one term that related to the limiting distribution of $\hat{\theta}$ and another third order U-statistic that vanishes asymptotically. Both Δ_2 and Δ_4 deal with estimation uncertainty from quantiles. In light of the work from Bahadur (1966), Pakes and Pollard (1989), Shen and Klein (2017), I derive their probability limits respectively. There is no estimated components in Δ_3, so a central limited theorem can be applied directly.

After incorporating the convergence results of all four Δs[21] and the asymptotic linear structure of $\sqrt{N}(\hat{\theta} - \theta_0)$ in (26), the vector of $\widehat{M_q^k} - M_q^k$ has an asymptotic linear form:

$$\begin{aligned}
\sqrt{N}(M_q^k - \widehat{M_q^k}) &= \frac{\sqrt{N}}{N}\sum_{g=1}^N [\Psi_{1j}^k + \Psi_{2j}^k + \Psi_{3j}^k + \Psi_{4j}^k] + o_p(1) \qquad (34) \\
\Psi_{1j}^k &\equiv \frac{E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z+\delta;\theta_0)] - E[t_{qj}\nabla_\theta P_j^k(V_F, V_B, Z;\theta_0)]}{E[t_{qj}]} H_0^{-1}G(\theta_0) \\
\Psi_{2j}^k &\equiv \{\nabla_q E[t_{qj}m_j(\theta_0)] - \nabla_q E[t_{qj}QME_q^k]\}\frac{\mathcal{B}_j}{E[t_{qj}]} \\
\Psi_{3j}^k &\equiv \frac{E[t_{qj}|V_F, V_B, Z+\delta]\epsilon_j^\delta - E[t_{qj}|V_F, V_B, Z]\epsilon_j}{E[t_{qj}]} \\
\Psi_{4j}^k &\equiv \frac{t_{qj}m_j(\theta_0) - E[t_{qj}]QME_q^k}{E[t_{qj}]} - \frac{t_{qj} - E[t_{qj}]}{E[t_{qj}]}QME_q^k
\end{aligned}$$

[21] A detailed proof is available upon request

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

$$\sqrt{N}\begin{bmatrix}\widehat{M_q^1}-M_q^1\\ \widehat{M_q^2}-M_q^2\\ \vdots\\ \widehat{M_q^L}-M_q^L\end{bmatrix}\xrightarrow{d} Z\sim N(0,\Omega) \tag{35}$$

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 column vector of length L 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 row vector of length L with the first K-1 component equals one and the rest equals zero, we have

$$\sqrt{N}(Q\hat{M}E_q^K - QME_q^K) = \sqrt{N}A\begin{bmatrix}\widehat{M_q^1}-M_q^1\\ \widehat{M_q^2}-M_q^2\\ \vdots\\ \widehat{M_q^L}-M_q^L\end{bmatrix}\xrightarrow{d} AZ\sim N(0,A'\Omega A) \tag{36}$$

Then the normality result follows. $\square$

References

Jason Abrevaya, Yu-Chin Hsu, and Robert P Lieli. Estimating conditional average treatment effects. *Journal of Business & Economic Statistics*, 33(4):485–505, 2015.

Renee B Adams and Daniel Ferreira. A theory of friendly boards. *The Journal of Finance*, 62(1): 217–250, 2007.

Anat R Admati and Paul Pfleiderer. The wall street walk and shareholder activism: Exit as a form of voice. *Review of Financial Studies*, page hhp037, 2009.

Hyungtaik Ahn, Hidehiko Ichimura, James L Powell, and Paul A Ruud. Simple estimators for invertible index models. *Journal of Business & Economic Statistics*, (just-accepted), 2017.

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

Bo Becker and Todd Milbourn. How did increased competition affect credit ratings? *Journal of Financial Economics*, 101(3):493–514, 2011.

Marshall E Blume, Felix Lim, and A Craig MacKinlay. The declining credit quality of us corporate debt: Myth or reality? *The journal of finance*, 53(4):1389–1413, 1998.

Patrick Bolton, Xavier Freixas, and Joel Shapiro. The credit ratings game. *The Journal of Finance*, 67(1):85–111, 2012.

John Y Campbell and Glen B Taksler. Equity volatility and corporate bond yields. *The Journal of Finance*, 58(6):2321–2350, 2003.

Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982.

Peter Crosbie and Jeff Bohn. Modeling default risk. 2003.

William F Eddy and JA Hartigan. Uniform convergence of the empirical distribution function over convex sets. *The Annals of Statistics*, pages 370–374, 1977.

Alex Edmans. Blockholder trading, market efficiency, and managerial myopia. *The Journal of Finance*, 64(6):2481–2513, 2009.

Wolfgang Härdle and Thomas M Stoker. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408):986–995, 1989.

Jie He, Jun Qian, and Philip E Strahan. Does the market understand rating shopping? predicting mbs losses with initial yields. *The Review of Financial Studies*, 29(2):457–485, 2015.

Joel L Horowitz and Wolfgang Härdle. Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association*, 91(436):1632–1640, 1996.

James O Horrigan. The determination of long-term credit standing with financial ratios. *Journal of Accounting Research*, pages 44–62, 1966.

Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, 58(1):71–120, 1993.

Hidehiko Ichimura and Lung-Fei Lee. 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*, pages 3–49, 1991.

John Xuefeng Jiang, Mary Harris Stanford, and Yuan Xie. Does it matter who pays for bond ratings? historical evidence. *Journal of Financial Economics*, 105(3):607–621, 2012.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *The American Economic Review*, 101(6):2590–2615, 2011.

Robert S Kaplan and Gabriel Urwitz. Statistical models of bond ratings: A methodological inquiry. *Journal of Business*, pages 231–261, 1979.

Simi Kedia, Shivaram Rajgopal, and Xing Alex Zhou. Large shareholders and credit ratings. *Journal of Financial Economics*, 124(3):632–653, 2017.

Roger Klein and Chan Shen. Bias corrections in testing and estimating semiparametric, single index models. *Econometric Theory*, 26(06):1683–1718, 2010.

Roger Klein, Chan Shen, and Francis Vella. Estimation of marginal effects in semiparametric selection models with binary outcomes. *Journal of Econometrics*, 185(1):82–94, 2015.

Roger W Klein and Robert P Sherman. Shift restrictions and semiparametric estimation in ordered response models. *Econometrica*, 70(2):663–691, 2002.

Roger W Klein and Richard H Spady. An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, pages 387–421, 1993.

Pepa Kraft. Rating agency adjustments to gaap financial statements and their effect on ratings and credit spreads. *The Accounting Review*, 90(2):641–674, 2014.

Pepa Kraft. Do rating agencies cater? evidence from rating-based contracts. *Journal of Accounting and Economics*, 59(2):264–283, 2015.

Lung-Fei Lee. Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. *Journal of Econometrics*, 65(2):381–428, 1995.

Melanie Lührmann and Jürgen Maurer. Who wears the trousers? a semiparametric analysis of decision power in couples. *MEA discussion papers*, 168, 2008.

Charles F Manski. Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of econometrics*, 27(3):313–333, 1985.

Jerome Mathis, James McAndrews, and Jean-Charles Rochet. Rating the raters: are reputation concerns powerful enough to discipline rating agencies? *Journal of monetary economics*, 56(5): 657–674, 2009.

Hans-Georg Muller. Smooth optimum kernel estimators of densities, regression curves and modes. *The Annals of Statistics*, pages 766–774, 1984.

EA Nadaraya. On non-parametric estimates of density functions and regression curves. *Theory of Probability & Its Applications*, 10(1):186–190, 1965.

Ariel Pakes and David Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, pages 1027–1057, 1989.

Mitchell A Petersen. Information: Hard and soft. 2004.

George E Pinches and Kent A Mingo. The role of subordination and industrial bond ratings. *The Journal of Finance*, 30(1):201–206, 1975.

James L Powell. Estimation of semiparametric models. *Handbook of econometrics*, 4:2443–2521, 1994.

James L Powell, James H Stock, and Thomas M Stoker. Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, pages 1403–1430, 1989.

Francesco Sangiorgi, Jonathan Sokobin, and Chester Spatt. Credit-rating shopping, selection and the equilibrium structure of ratings. Technical report, Working Paper, Stockholm School of Economics and Carnegie Mellon University, 2009.

Chan Shen and Roger Klein. Recursive differencing for estimating semiparametric models, 2017.

Vasiliki Skreta and Laura Veldkamp. Ratings shopping and asset complexity: A theory of ratings inflation. *Journal of Monetary Economics*, 56(5):678–695, 2009.

Thomas M Stoker. Consistent estimation of scaled coefficients. *Econometrica: Journal of the Econometric Society*, pages 1461–1481, 1986.

Richard R West. An alternative approach to predicting corporate bond ratings. *Journal of Accounting Research*, pages 118–125, 1970.

Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.

Lawrence J White. The credit rating industry: An industrial organization analysis. In *Ratings, rating agencies and the global financial system*, pages 41–63. Springer, 2002.