Online Appendices: Identification and Estimation of Soft Adjustment in Structural Bond Rating Models: Before and After the Dodd-Frank Act

Zhutong Gu*[1], Yixiao Jiang[2], and Shuyang Yang[3]

[1]*Peking University, HSBC Business School*

[2]*Rutgers University, Department of Economics*

[3]*Amazon Inc.*

The online appendices are supplemental to our main text. In the first section, we give an example to illustrate the consequences due to the presence of unobserved soft adjustment. In section 2, we provide empirical methods and results to examine such likely consequences. Section 3 presents asymptotic results and properties for the proposed two-stage semiparametric estimator. The last section contains additional empirical results of interest.

1 An Example of Endogenous Bond Characteristics

In what follows, we use a simple example to illustrate the potential endogeneity problem due to the presence of bond-level soft adjustment. Now consider firms' issuing decision. In particular, a typical firm decides how much debt to issue in order to maximize the discounted

*Corresponding author, zgu@phbs.pku.edu.cn.

expected return on bond capital given its private information including the shareholder connection with the CRA.[1] For simplicity, it is also assumed that firms invest all what they can finance from the issuance in some businesses that on average pays the discounted return of ROI_i per dollar investment.

$$x^B = \underset{x^B \geq 0}{\arg\max}\, E\left[ROI_i - C(Y_i)\right] | private_i]\, x^B$$

where the expectation is taken with respect to the return of investment, ROI_i as well as the categorical rating, Y_i. $C(j)$ denotes the per dollar interest payment to investors for a bond being rated in notch j with a predetermined subordination status.[2] The investment decisions of common shareholders, R_i, in the CRA and firm i are outside the model and taken as given. Note that the *ex post* information on the shareholding structure is known to both firm i and the CRA before the endogenous determination of the "hard" and "soft" assessment.

The borrowing cost faced by the firm is largely determined by the CRA's reported rating. It can be conjectured that the interest cost is strictly decreasing with the ratings, e.g. $C(0) < \cdots < C(J-1)$. Firms make their issuing decisions based on its own *ex ante* interest cost which can be written as the average category-specific cost weighted by the rating distribution.

$$E[C(Y_i)|private_i] = \sum_{j=0}^{J-1} C(j) \Pr(T_{j-1,i} < X_i'\beta_0 \leq T_{ji}|private_i) \equiv F(X_i'\beta_0, \boldsymbol{\delta}(R_i))$$

While, as reflected above, the rating distribution is purely driven by the firm's default risk index, $V_i = X_i'\beta_0$, and the set of firm i-specific soft adjustment, $\boldsymbol{\delta}(r) \equiv \{\delta_0(r), \cdots, \delta_{J-1}(r)\}$. Note that $F(\cdot)$ is a potentially nonseparable function of the default risk index V_i and the conditional thresholds, $\boldsymbol{\delta}(R_i)$. After solving the model, the optimal amount x_1^B will be an implicit function of $\boldsymbol{\delta}(R_i)$. In short, bond-issuing firms could update their beliefs of

[1] Among other characteristics that the firm may choose upon issuance, prior studies (Pinches and Mingo, 1973; Kaplan and Urwitz, 1979; Blume et al., 1998, .etc) have shown that issuing amount and subordination status are the two dominant factors that affect the borrowing cost.

[2] We abstract away any dynamic issues to keep the illustration simple but the implications hold in more general settings.

the CRA's soft adjustment (or thresholds distribution) given the private knowledge on investment relationship while making financing decisions accordingly. With a closer liaison, the issued bonds may enjoy a lower interest cost because of the higher probability of receiving favorable adjustment and being rated into better notches, holding other factors the same. This therefore motivates the firm to issue more debt, declare subordination status more often and undertake higher leverage ratios.

Moreover, it implies that endogeneity of bond characteristics would arise unless the dependency between x_1^B and $\boldsymbol{\delta}(R_i)$ is properly taken care of. This simple model also suggests that controlling for the nonlinear effect of shareholder structure R_i could solve the omitted variable problem in terms of empirical strategy. In contrast, most previous bond rating models have taken financial variables, including bond characteristics, as purely exogenous, from CRAs' point of view. This simplification can only be justified when there is no private information of the firm-specific soft adjustment or all firms have common beliefs on the distribution of rating thresholds. However, our empirical evidence does not support this view.

A similar argument is applied to other choices such as the subordinate status. Loosely speaking, firms choose to declare subordinate status if the net-payoff from declaring outweighs not declaring. For simplicity, one may assume that firms make issuing choices in a sequential way—first declaring the subordination status and then determining the issuing amount. Then it would be natural that the subordination status is correlated with the conditional thresholds through its correlation with X_1^B. As a consequence, other financial variables such as leverage ratios also become endogenous as they are functions of the amount of debt that need to be paid off later. For example, consider a firm with multiple bonds, the debt-to-asset ratio increases as more bonds are issued. Since the optimal issuing amount is correlated with the individual threshold through the CRA-issuer liaison, so would be the leverage ratio.

The last step is rather mechanical. Given the firm's choices X_i, the CRA calculates the *ex-post* default risk measure V_i and adopts the pre-specified bond-specific soft adjustment in the thresholds. Then an independent rating noise, u_{ij}, is randomly drawn from its distribution and being added to the threshold at each category. Finally, an ordered rating is generated and released to all investors and the general public.

2 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.

2.1 Structural and Nonstructural Probabilistic Rating Functions

We first define the non-structural conditional rating probability function in Eq. (2.1) for any $v \in \mathbb{R}$

$$P_j^n(v) \equiv \Pr(Y_i \leq j | V_{0i} = v) = \int \{v \leq T_j\} dF_{T_j|V_0=v}(t), \quad j = 0, 1, \cdots, J-1 \tag{2.1}$$

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. (2.2) given $V_{0i} = v$,

$$P_j^s(v) \equiv \Pr(v \leq T_j) = \int \{v \leq T_j\} dF_{T_j}(t), \quad j = 0, 1, \cdots, J-1 \tag{2.2}$$

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.[3] 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. $\mathcal{R} = \mathcal{R}^v, \ \ \forall v \in \mathbb{R},$ *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 adjustment alone, A-I.3 is not necessary. Proposition A states that $P_j(v)$ in Eq. (2.2) 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.

Proposition A. *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.

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

$\square$

Finally, consider estimators of the structural and non-structural conditional probability functions defined in Eq. (2.1) and Eq. (2.2). 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. (2.3).

$$\widehat{P}_j^s(v) = \frac{1}{N}\sum_{i=1}^{N}\widehat{P}_j(v, R_i(\widehat{\alpha})) \tag{2.3}$$

In contrast, the nonstructural conditional probability functions can be straightforwardly estimated as the conditional expectation function in Eq. (2.4) in which $\widehat{E}$ denotes the local constant estimator

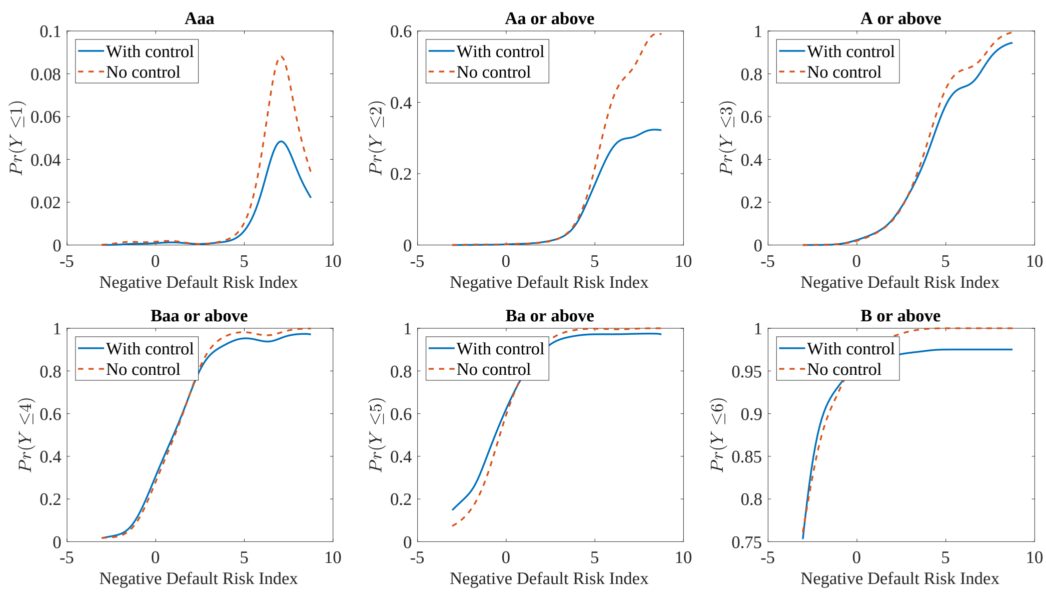
$$\widehat{P}_j^n(v) = \widehat{E}(\{Y_i \leq j\}|V_i(\widehat{\beta}) = v) \tag{2.4}$$

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.

2.2 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 1, we plot the cumulative structural (in solid) and non-structural (in dash) rating probability functions, defined in Eq. (2.2) and (2.1), 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 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 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.

3 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 $v, \widetilde{v} \in \mathbb{R}$, $|\nabla^l P_j(v, r) - \nabla^l P_j(\widetilde{v}, r)| \leq C||v - \widetilde{v}||$, 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. Silverman's rule of thumb bandwidth, e.g. $h_i = 1.06 \times std \times N^{-r_i}$, for $i = 1, 2$, is being used. A-A.6 restricts support of the finite and infinite-dimensional parameters to be compact given point identification.

Theorem 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 conserver space, we do not reiterate the proof of Theorem 3 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).

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

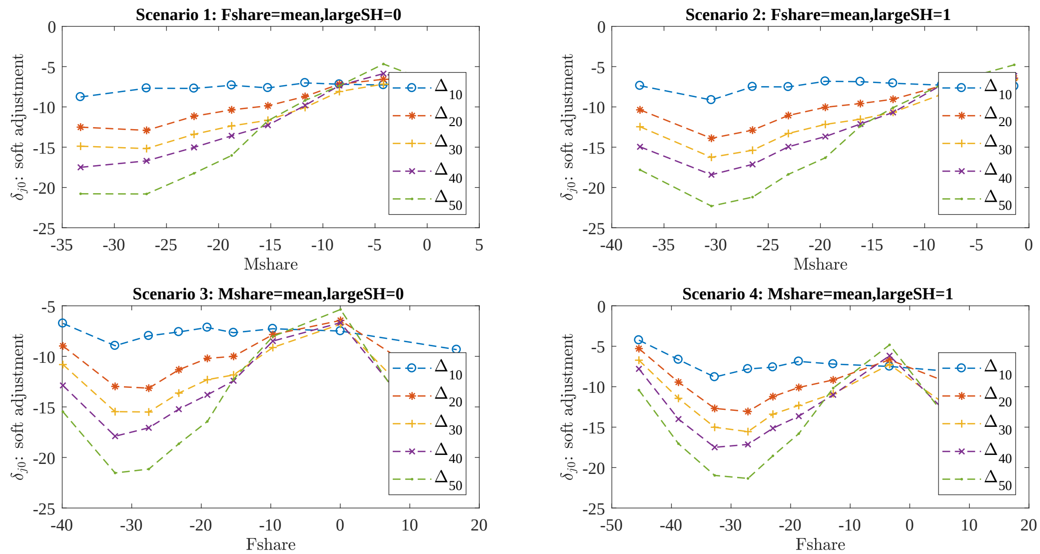
A remark on bandwidth selection. As in Assumption A-A.5, bandwidths are allowed to be different for estimating $\widehat{\theta}$ and $\widehat{\Delta}(r)$. In order to eliminate the bias in the theorem, bias-reducing techniques can be apply (Klein and Shen, 2010; Shen and Klein, 2017, etc.). A nice 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.

4 Pattern of Thresholds over Shareholding Relations

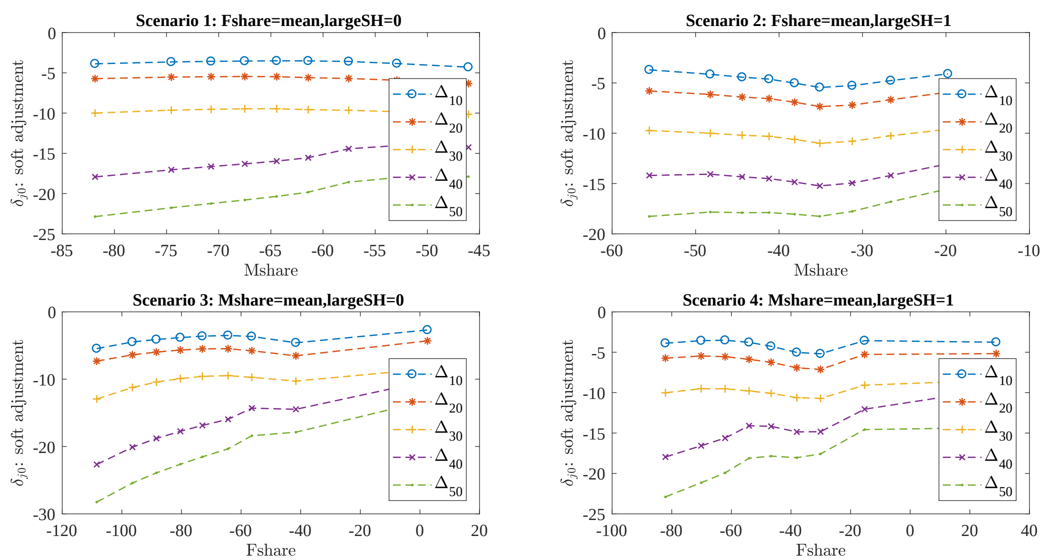
In the upper panels of Figure 2 and Figure 3, we plot the estimated relationship between the investment liaison with Moody or Mshare, and the soft adjustment, holding Fshare at mean, under situations of no large shareholder (in left panel) and at least one large shareholder (in right panel). Then we switch Mshare and Fshare in the lower panels. We first look at Figure 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 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 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 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.

References

Joseph G Altonji and Rosa L Matzkin. Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, pages 1053–1102, 2005. doi: 10.1111/j.1468-0262.2005.00609.x.

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. doi: 10.1111/0022-1082.00057.

Richard Blundell and James L Powell. Endogeneity in nonparametric and semiparametric regression models. *Econometric Society Monographs*, 36:312–357, 2003.

Richard W Blundell and James L Powell. Endogeneity in semiparametric binary response models. *The Review of Economic Studies*, 71(3):655–679, 2004. doi: 10.1111/j.1467-937X.2004.00299.x.

Guido W Imbens and Whitney K Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512, 2009. doi: 10.3982/ECTA7108.

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

Roger Klein and Chan Shen. Bias corrections in testing and estimating semiparametric, single index models. *Econometric Theory*, 26(06):1683–1718, 2010. doi: 10.1017/S0266466609990764.

Roger W Klein and Robert P Sherman. Shift restrictions and semiparametric estimation in ordered response models. *Econometrica*, 70(2):663–691, 2002. doi: 10.1111/1468-0262.00299.

George E Pinches and Kent A Mingo. A multivariate analysis of industrial bond ratings. *The Journal of Finance*, 28(1):1–18, 1973. doi: 10.1111/j.1540-6261.1973.tb01341.x.

Chan Shen and Roger Klein. Market recursive differencing: Bias reduction with regular kernels. Technical report, Rutgers University, Department of Economics, 2017. Working paper.