

Non-Linearity in Corporate Finance Studies: Bank Size Matters When It Comes to Leverage Components

David E. Booth
Kent State University

Nonna Sorokina
The Pennsylvania State University

We propose a regression-based procedure to account for non-linearity, while maintaining robustness to multicollinearity and endogeneity. To illustrate our method, we focus on capital structure, a fundamental topic in corporate finance, and examine excess, non-deposit and short-term components of leverage. We model those leverage measures as functions of competition, diversification, and liquidity, interacting with bank size. Our analysis begins with fully specified second-order models for each determinant in combination with bank size, revealing significant non-linear effects. To address potential multicollinearity among second-order terms, we apply lasso regression for variable selection and initial coefficient estimation. We then augment the selected models with bank and time fixed effects to control for unobserved heterogeneity and address endogeneity concerns. Our findings indicate that linear specifications often misattribute significance to less relevant variables by failing to account for underlying non-linearities and interactions. This omission leads to biased estimates and distorted interpretations. Our approach provides a practical and reliable method for obtaining unbiased and interpretable estimates in the presence of complex variable relationships.

Keywords: bank, capital structure, leverage, lasso, fixed effects, multicollinearity, endogeneity, non-linearity

INTRODUCTION

We decompose bank leverage into three components—short-term, non-deposit, and excess (defined as the portion of leverage exceeding that of a comparable non-financial firm)—and examine their determinants. Our analysis identifies competition, diversification, and liquidity as significant drivers of all three components. However, the relationships are demonstrably non-linear and vary with the financial institution's size.

To capture these complexities, we employ fully specified second-order models incorporating both non-linear and interaction terms between the three key determinants and bank size. Given the high variance inflation factors (VIFs) observed, multicollinearity is a concern that renders ordinary least squares (OLS) estimates unreliable. To address this, we utilize lasso—a variable selection technique robust to multicollinearity—which allows us to uncover and confirm the non-linear and interaction effects more effectively.

While the initial results from the lasso estimation are revealing, our goal is to construct comprehensive leverage models that incorporate control variables and unify all relevant effects. To this end, we estimate models for each leverage component (short-term, non-deposit, and excess leverage) by including all variables from the complete second-order models for competition, diversification, and liquidity into previously established base specifications. Lasso is then reapplied to identify the most informative and parsimonious set of predictors.

Recognizing that these lasso-selected models may still be vulnerable to endogeneity—a common concern in corporate finance—we incorporate bank and time fixed effects in the final estimation stage. By comparing lasso and fixed effects estimates, we ensure robustness to both multicollinearity and endogeneity.

This study advances the understanding of the determinants of bank leverage components and introduces a methodological framework of broad relevance to empirical corporate finance. In particular, it underscores the value of explicitly modeling non-linearities and interactions—an area often overlooked in standard linear approaches—thereby revealing important, hidden dynamics.

LITERATURE AND MOTIVATION

Most models in empirical corporate finance literature are linear and the most common statistical concern is endogeneity. Baker and Wurgler (2002), Frank and Goyal (2003), Jung et al. (1996), Fischer et al. (1989), Leary and Roberts (2010), Rajan and Zingales (1995), and Titman and Wessels (1988) all utilize linear models for various studies of leverage determinates. Similarly, Gropp and Heider (2010) whose approach served as a basis for Sorokina et al. (2017), a source for our base models, also employ linear regressions with fixed effects in their research.

One notable example of a simple non-linear model may be found in Lemmon and Zender (2010). Lemmon and Zender (2010) estimate a firm's debt capacity in a two-step procedure, based on the probability of a firm's having a bond rating in a particular year. Lemmon and Zender use the presence/absence of a firm's bond ratings in COMPUSTAT as a dependent variable in the logit estimation model. In the classic pecking order model test based on financing deficit, the deficit is included in first and second order form and the second order term improve explanatory power for low debt capacity firms. Some more complex models allow non-linearity such as Berger et al. (2008) and Berger et al. (2009). However, none perform a systematic assessment of the interactions and non-linearity in all major determinants of interest.

A treatment of multicollinearity is also rather rare in the corporate finance studies, which makes Frank and Goyal (2009) another methodological exception. The scope of Frank and Goyal (2009) study of leverage ratios determinants is uniquely broad. They survey existing literature and identify significant determinants of capital structure based on major theories. Frank and Goyal use several measures of firm leverage - ratios of total debt to market value of assets, total debt to book value of assets, long-term debt to market and to book value of assets. Frank and Goyal select a number of leverage determinants categories in their tests, including firm-specific, industry-specific, time-specific and macro-economic factors, and identify proxies for each category. With a couple dozens of dozen determinants, many of which are correlated, Frank and Goyal decide to use a novel for the discipline variable selection model – lasso and cite Hastie et al. (2001) as their source of variable selection method. Originally, Tibshirani (1996) develops lasso, a regression method that combines both the benefits of information on criterion-based variable selection and ridge regression. We use a similar, yet more powerful, modern version of the model—adaptive lasso for linear regression models with the weighted approach by Zou (2006). The adaptive lasso combines the benefits of greater variable selection accuracy and estimation precision. The procedure uses SBC criterion and other information criterion measures, such as BIC and AIC, reported in Frank and Goyal.

One indicator of the potential multicollinearity and a need for an advanced variable selection model is the Variance Inflation Factors (VIFs) in estimated models. The commonly accepted threshold of VIF=10 was defined by Marquardt (1970). However, there is a substantial debate in statistical literature regarding

the methodology of multicollinearity diagnostics, mostly between Marquardt and Belsley - see for example Snee and Marquardt (1984) for details.

Roberts and Whited (2011) explain that concern regarding endogeneity, a correlation between error term in the regression and explanatory variables, is present nearly in every empirical corporate study due to complexity and opacity of firm-level decision processes. While many endogeneity treatments are available, panel estimators remain most popular when dataset size permits. Although fixed effects sometimes strip a study of the most interesting and important finding.

Sorokina et al. (2022) find significant non-linearity in relationship between bank regulatory capital and important determinants of bank capital structure – competition, diversification, liquidity, and their interaction with bank size. They don't explore non-linearity in similar explanatory models for leverage components – excess, non-deposit, and short-term leverage. Our study extends the existing literature in two ways. First, the estimation of a sophisticated non-linear model is non-trivial though due to correlations between first and second order terms and interactions. We propose a multi-step procedure to build a non-linear model with appropriate variable selection and estimate robust to multicollinearity and endogeneity. While we only focus on capital structure literature in this study, the adherence to linear models is not limited to this area. Anywhere, from executive compensation to mergers and acquisitions, non-linearity is regularly left out of the scope of study. Our method of non-linear model development is useful for all empirical corporate finance studies. Second, we develop a non-linear model of robust explanatory models for bank leverage components and shed extra the light on bank capital structure drivers in various size categories.

DATA AND EMPIRICAL DESIGN

We use COMPUSTAT North America and COMPUSTAT Bank as sources of bank-specific information, CRSP for the stock-related data, and the Bank Regulatory Database for detailed bank-specific data. We obtain macro-variables from various governmental agencies, following Frank and Goyal (2009). The details of the underlying data elements for each variable are provided in Table 1. We begin with a broad sample that covers forty years of quarterly data between 1973 and 2012. The data period is further limited because some data are not available for the early years (for example, risk-weighted capital for Pre-Basel years, debt rating variable and other regulatory data are unavailable before 1990). There are 16,459 bank-quarter observations with valid values of leverage variables along with respective independent variables over the 1993 to 2012 period. There are 609 unique banks in our sample. The size of the banks varies greatly between \$31.5 million and \$1.5 trillion with average bank's book assets at \$1.68 billion and median bank's assets worth \$11 billion in book value terms.

TABLE 1
VARIABLE DESCRIPTION

Name	Description	Source
<u>Leverage</u>	<p><i>Short Term Leverage</i> The ratio of short-term borrowing = Total Assets (ATQ) – Shareholder's Equity (SEQQ) - Long-Term Debt (DLTTQ) – Deposits (DPTCQ) to assets (ATQ)</p> <p><i>Non-Deposit Leverage</i> The ratio of borrowing except deposits = total debt (debt in current liabilities (DLC) + long-term debt (DLTT)) to assets (ATQ).</p> <p><i>Excess Leverage</i> A calculated variable - details are explained in methodology section.</p>	Frank, Goyal (2009)

<u>MTB</u>	Market-to-Book ratio. MVA to COMPUSTAT item 6, assets. MVA is obtained as the sum of the market value of price-close (PRCC) × shares outstanding (CSHPRI)+ short-term debt (DLC) + long-term debt (DLTT) + preferred-liquidation value (PSTKL) – deferred taxes and investment tax credit (TXDITC)	Frank, Goyal (2009)
<u>Profit</u>	Profitability, the ratio of operating income before depreciation(OIBDP) and assets (AT)	Frank, Goyal (2009)
<u>Ln(Size)</u>	Log of assets (ATQ)	Frank, Goyal (2009)
<u>Collateral</u>	(total securities + treasury bills + other bills + bonds + CDs + cash and due from banks + land and buildings + other tangible assets)/book value of assets	Gropp, Heider (2010)
<u>Dividend</u>	Dummy variable, equal to 1 when firm pays dividends and 0 otherwise. (based on COMPUSTAT item cash dividends declared on common stock DVCY for banks or CDVCY for all firms)	Frank, Goyal (2009)
<u>Risk (LogRiskM)</u>	Annual variance of stock returns (based on CRSP stock returns)	Frank, Goyal (2009)
<u>MatureDummy</u>	Mature firm dummy, equal to 1 if firm has been listed in COMPUSTAT for more than five years	Frank, Goyal (2009)
<u>Maturity</u>	Firm maturity Retained Earnings / Total Equity (Compustat items REQ, retained earnings, SEQQ, shareholders' equity, total)	DeAngelo, DeAngelo, Stultz (2006)
<u>Δ Ln(Size)</u>	Year-to-year change in log book assets (see ln(size) variable above)	Frank, Goyal (2009)
<u>MedIndLeverage (MedSampleLeverage)</u>	Median industry leverage. The median of total debt to market value of assets (see definition of leverage above) by SIC code and by year. Industry is defined at the four-digit SIC code level for firms. For banks, the variable represents median leverage of the sample by year.	Frank, Goyal (2009)
<u>MedIndGrowth (MedSampleGrowth)</u>	Median industry growth, median year-to-year change in log assets (ATQ) for the entire sample of banks.	Frank, Goyal (2009)
<u>SGA</u>	The ratio of selling, general and administration expenses (XSGA) to sales (SALE).	Frank, Goyal (2009)

<u>DebtRating</u>	The dummy variable that is equal to 1 when senior debt rating (SPLTICRM) subordinate debt rating (SPSDRM), is less than 13 (investment grade) Actual: letter grade ratings, investment grade to BBB-per S&P: http://www.standardandpoors.com/ratings/definitions-and-faqs/en/us	Frank, Goyal (2009)
<u>Debt capacity</u>	Measured as propensity to have a bond rating calculated for each firm using estimates obtained from entire sample based on COMPUSTAT item senior debt rating (SPLTICRM)	Lemmon, Zender (2010)
<u>MktReturn</u>	Cumulative quarterly CRSP stock returns	Frank, Goyal (2009)
<u>CRSPVW</u>	Annual Returns on value weighted stock portfolio, CRSP	Frank, Goyal (2009)
<u>Inflation</u>	expected change in the consumer price index over the coming year using data from the Livingston Survey available at http://www.phil.frb.org/econ/liv/index.html	Frank, Goyal (2009)
<u>Term spread</u>	difference between the 10-year interest series and the one-year interest series (Source: Federal Reserve files at http://www.federalreserve.gov/releases/)	Frank, Goyal (2009)
<u>Macro Growth</u>	log of aggregate annual corporate profits after tax for nonfinancial firms. (Source: US Department of Commerce, Bureau of Economic Analysis.) Actual: Federal Reserve of St Louis, PctAnnualChange	Frank, Goyal (2009)
<u>CapEx</u>	Capital Expenditures to assets, ratio of compustat capital expenditure (CAPXQ), to assets (ATQ)	Frank, Goyal (2009)
<u>GDPGrowth</u>	differences in log of real gross domestic product (Source: US Department of Commerce, Bureau of Economic Analysis) Actual: PctAnnualChange	Frank, Goyal (2009)

<p><u>Competition</u></p>	<p><i>Lerner index</i> A price mark-up over marginal costs, where marginal costs are derived using translog function from total assets (proxy for output) COMPUSTAT item ATQ, the ratios of personnel expenses to total assets (XLRQ – salaries and wages + XPRBQ – pensions and benefits + STLCOQ = stock compensation expense to ATQ), interest expenses to total deposits (XINDCQ, interest on deposits to DPTCQ, deposits total) and other operating and administrative expenses to total assets (FEOENQ, fixed expense (occupancy and equipment) + XADQ, advertising and marketing expenses +XCOMC, communications expenses +OCOEQ , all other current operating expenses to ATQ), proxies for prices of labor, funds, and fixed capital</p>	<p>Berger et al. (2009)</p>
<p><u>Diversification</u></p>	<p>HHI for each bank based on different types of loans: Bank Regulatory database items 1410 – real estate loans, 1590 – agricultural loans, 1600-commercial and industrial loans, 1975 – loans to individuals, 2165-lease financing receivables</p>	<p>Lang and Stulz (1994) Further in bank literature: Demsetz and Strachan (1997), Acharya et al. (2006)</p>
<p><u>Liquidity</u></p>	<p>Loan-to-deposit shortfall. COMPUSTAT items [Loans (LNTALQ) - Deposits (DPTCQ)]/ assets (ATQ)</p>	<p>Acharya, Mora (2015)</p>

The letter-coded names in parenthesis, such as (AT), (DLC), etc. in this table are COMPUSTAT field names.

The summary statistics of this dataset are presented in Table 2. The short-term leverage of the average bank in our sample is 5.45%, while median short-term leverage is somewhat higher, at 7.50%. The range of all values is very wide, and spans from 0% to 62.44%. Excess book leverage for both the mean and median institution is 10%, varying between 0% and 23.50% in our sample. The leverage based on debt that does not include deposits is, on average, 12.00% while the median is quite close, at 13.60%. However, some banks carry much higher non-deposit debt obligations, up to 70% of their total book assets. The respective measures for all independent variables are also provided in Table 2, for reference.

TABLE 2
SUMMARY STATISTICS

Variable	Mean	Median	Std Dev	Minimum	Maximum
BookLevShortLead1Y	0.0544631	0.0749952	0.0717466	0.000191542	0.6244194
BookLevExcessLead1Y	0.1001787	0.0989188	0.0390539	0.000186763	0.2355122
BookLevNoDepLead1Y	0.1200777	0.1360723	0.0968502	-0.00062888	0.6981946
CapEx	0	0.00053837	0.0020502	-0.000985203	0.0770392
CRSPVW	0.1620248	0.1102018	0.1909313	-0.3821115	0.356723
Collateral	1.0793356	1.0949503	0.176684	0.3883865	2.1678599
CompetLerner	0.3611186	0.3790954	0.1489373	0.0015895	0.8886656
DebtCapacity	0.9897972	0.8825352	0.2564886	0.0001762	1
DebtRating	0	0.1201167	0.3251078	0	1
Diversification	5.2266894	5.496733	1.7479083	0.000928731	12.2470948
Dividend	1	0.818458	0.3854783	0	1
GDPGrowth	0.049	0.0494895	0.0264036	-0.078	0.102
Growth	0.086995	0.3805142	1.2939472	-0.7063579	9.7635323
Inflation	0.0334386	0.0551683	0.0420557	0.0228515	0.1566376
LiquidBank	-0.1284042	-0.140115	0.1393252	-0.6725266	0.5654228
LogSize	7.4266324	7.6865355	1.5728791	3.4524607	14.1865298
MacroGrowth	0.0845101	0.1055027	0.1800778	-0.4889244	1.0586833
MatureDummy	1	0.8786682	0.3265224	0	1
Maturity	0.5212358	0.4889962	0.523275	-18.81925	5.7662733
MedIndGrowth	0.0937532	0.4158928	1.2467662	-0.0168607	5.8097208
MedIndLeverage	0.8635792	0.8464989	0.1270685	0	0.929969
MktReturn	0.1875202	0.147984	0.2494891	-0.4320149	0.7259837
MTB	0.2708462	0.2858583	0.1186168	0.000610571	1.0992567
Profit	0.0066028	0.0067966	0.0037223	-0.0347138	0.100467
LogRiskM	-5.4340478	-5.3967417	0.9850531	-10.3390199	-0.2127349
SGA	0.2752722	0.2792044	0.0702451	0.0290105	0.9882821
Size	1680.14	11066.12	47242.19	31.578	1449211
TermSpread	0.0093	0.0128083	0.0107737	-0.0038	0.0333

In this study, we explore non-linear relationship and interaction between bank size and competition, diversification, and liquidity as they explain bank leverage components—excess, non—deposit, and short-term leverage. The details of the non-deposit and short-term variable construction are provided in Table 1. We estimate the excess leverage of banks in three-step procedure. First, we compute coefficients for the Frank and Goyal (2009) core leverage model based on the sample of non-financial firms. Second, we calculate the predicted leverage of banks using coefficients estimated for the non-financial firms' sample. Third, we compute the difference between the actual and predicted leverage (excess leverage).

Sorokina et. al (2017) find that competition, diversification, and liquidity are important determinants of the three leverage components we have defined above. There is, however, a substantial variation in relations between variables for different size groups. The effect appears to be non-linear.

First, we use complete second order models with excess, non-deposit, and short-term leverage as dependent variables, confirming non-linearity in the relationship.

$$\text{Projected Leverage}_t = \beta_0 + \beta_1 \text{MTB}_t + \beta_2 \text{Profit}_t + \beta_3 \text{Ln}(\text{Size})_t + \beta_4 \text{Tangibility/Collateral}_t + \beta_5 \text{MedIndLeverage}_t + \varepsilon \quad (0a)$$

$$\text{Excess Leverage}_t = \text{Actual Bank Leverage}_t - \text{Projected Leverage}_t \quad (0b)$$

$$\text{Excess Leverage}_t = \beta_0 + \beta_1 \text{Profit}_{t-1} + \beta_2 \text{Ln}(\text{Size})_{t-1} + \beta_3 \text{Ln}(\text{Size})_{t-1}^2 + \beta_4 \text{Collateral}_{t-1} + \beta_5 \Delta \text{Ln}(\text{Size})_{t-1} + \beta_7 \text{MedIndLeverage}_{t-1} + \beta_9 \text{Inflation}_{t-1} + \beta_{10} \text{GDPGrowth}_{t-1} + \beta_{16} \text{Diversification}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{17} \text{Liquidity}_{t-1} + \beta_{18} \text{Liquidity}_{t-1}^2 + u \quad (1a)$$

$$\text{Excess Leverage}_t = \beta_0 + \beta_1 \text{MTB}_{t-1} + \beta_2 \text{Ln}(\text{Size})_{t-1} + \beta_3 \text{Ln}(\text{Size})_{t-1}^2 + \beta_4 \text{Collateral}_{t-1} + \beta_5 \text{Dividend}_{t-1} + \beta_6 \text{MatureDummy}_{t-1} + \beta_7 \text{MedIndLeverage}_{t-1} + \beta_8 \text{MktReturn}_{t-1} + \beta_9 \text{CRSPVW}_{t-1} + \beta_{10} \text{Competition}_{t-1} + \beta_{11} \text{Diversification}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{12} \text{Liquidity}_{t-1} + \beta_{13} \text{Liquidity}_{t-1}^2 + u \quad (1b)$$

$$\text{Non-Deposit Leverage}_t = \beta_0 + \beta_1 \text{Profit}_{t-1} + \beta_2 \text{Ln}(\text{Size})_{t-1} + \beta_3 \Delta \text{Ln}(\text{Size})_{t-1} + \beta_4 \text{CapEx}_{t-1} + \beta_5 \text{SGA}_{t-1} + \beta_6 \text{Inflation}_{t-1} + \beta_7 \text{GDPGrowth}_{t-1} + \beta_8 \text{Competition}_{t-1} + \beta_9 \text{Diversification}_{t-1} + \beta_{10} \text{Diversification}_{t-1}^2 + \beta_{11} \text{Diversification}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{12} \text{Liquidity}_{t-1} + \beta_{13} \text{Liquidity}_{t-1}^2 + u \quad (2a)$$

$$\text{Non-Deposit Leverage}_t = \beta_0 + \beta_1 \text{MTB}_{t-1} + \beta_2 \text{Ln}(\text{Size})_{t-1}^2 + \beta_3 \text{Collateral}_{t-1} + \beta_4 \text{Dividend}_{t-1} + \beta_5 \text{MedIndLeverage}_{t-1} + \beta_6 \text{MktReturn}_{t-1} + \beta_7 \text{CRSPVW}_{t-1} + \beta_8 \text{Competition}_{t-1}^2 + \beta_9 \text{Competition}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{10} \text{Diversification}_{t-1}^2 + \beta_{11} \text{Diversification}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{12} \text{Liquidity}_{t-1} + \beta_{13} \text{Liquidity}_{t-1}^2 + u \quad (2b)$$

$$\text{Short-term Leverage}_t = \beta_0 + \beta_1 \text{Profit}_{t-1} + \beta_2 \text{Ln}(\text{Size})_{t-1} + \beta_3 \text{Collateral}_{t-1} + \beta_4 \Delta \text{Ln}(\text{Size})_{t-1} + \beta_5 \text{Inflation}_{t-1} + \beta_6 \text{GDPGrowth}_{t-1} + \beta_7 \text{Competition}_{t-1} + \beta_8 \text{Competition}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_9 \text{Diversification}_{t-1}^2 + \beta_{10} \text{Diversification}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{11} \text{Liquidity}_{t-1} * \text{Ln}(\text{Size})_{t-1} + u \quad (3a)$$

$$\text{Short-term Leverage}_t = \beta_0 + \beta_1 \text{MTB}_{t-1} + \beta_2 \text{Ln}(\text{Size})_{t-1} + \beta_3 \text{Collateral}_{t-1} + \beta_4 \text{MedIndLeverage}_{t-1} + \beta_5 \text{DebtCapacity}_{t-1} + \beta_6 \text{Competition}_{t-1}^2 + \beta_7 \text{Competition}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_8 \text{Diversification}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_9 \text{Liquidity}_{t-1}^2 + \beta_{10} \text{Liquidity}_{t-1} * \text{Ln}(\text{Size})_{t-1} + u \quad (3b)$$

$$\text{Leverage component}_t^* = \beta_0 + \beta_1 \text{Profit}_{t-1} + \beta_2 \text{Ln}(\text{Size})_{t-1} + \beta_3 \text{Ln}(\text{Size})_{t-1}^2 + \beta_4 \text{Collateral}_{t-1} + \beta_5 \Delta \text{Ln}(\text{Size})_{t-1} + \beta_6 \text{CapEx}_{t-1} + \beta_7 \text{MedIndLeverage}_{t-1} + \beta_8 \text{SGA}_{t-1} + \beta_9 \text{Inflation}_{t-1} + \beta_{10} \text{GDPGrowth}_{t-1} + \beta_{11} \text{Competition}_{t-1} + \beta_{12} \text{Competition}_{t-1}^2 + \beta_{13} \text{Competition}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{14} \text{Diversification}_{t-1} + \beta_{15} \text{Diversification}_{t-1}^2 + \beta_{16} \text{Diversification}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{17} \text{Liquidity}_{t-1} + \beta_{18} \text{Liquidity}_{t-1}^2 + \beta_{19} \text{Liquidity}_{t-1} * \text{Ln}(\text{Size})_{t-1} + u \quad (4a-c)$$

$$\text{Leverage component}_t^* = \beta_0 + \beta_1 \text{MTB}_{t-1} + \beta_2 \text{Ln}(\text{Size})_{t-1} + \beta_3 \text{Ln}(\text{Size})_{t-1}^2 + \beta_4 \text{Collateral}_{t-1} + \beta_5 \text{Dividend}_{t-1} + \beta_6 \text{MatureDummy}_{t-1} + \beta_7 \text{MedIndLeverage}_{t-1} + \beta_8 \text{DebtCapacity}_{t-1} + \beta_9 \text{MktReturn}_{t-1} + \beta_{10} \text{CRSPVW}_{t-1} + \beta_{11} \text{Competition}_{t-1} + \beta_{12} \text{Competition}_{t-1}^2 + \beta_{13} \text{Competition}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{14} \text{Diversification}_{t-1} + \beta_{15} \text{Diversification}_{t-1}^2 + \beta_{16} \text{Diversification}_{t-1} * \text{Ln}(\text{Size})_{t-1} + \beta_{17} \text{Liquidity}_{t-1} + \beta_{18} \text{Liquidity}_{t-1}^2 + \beta_{19} \text{Liquidity}_{t-1} * \text{Ln}(\text{Size})_{t-1} + u \quad (5a-c)$$

*Leverage component – excess, non-deposit or short-term leverage

The independent variables in the complete second order models are highly correlated by definition. We are concerned about the potential multicollinearity issue in the models. Multicollinearity in a multivariable regression model leads to misattribution of the effect between variables. An entire model may retain its explanatory power, however, as a result of multicollinearity, size, sign and/or statistical significance of

individual estimated coefficients may not be reliable. Traditionally, variance inflation factors (VIFs) analysis shows how much coefficient's variance increases due to multicollinearity; the VIFs-based adjustments help determine the correct statistical significance of the estimated coefficients. We check VIFs in the estimated models and find that VIFs, provided in Table 3, in most cases, are high. Lasso, a regression method that combines a well-established variable selection procedure based on the information criteria and a robust coefficient estimation method, appears to be the most appropriate and reliable in our study. We use a modern version of the method - adaptive lasso method of Zou (2006) that produces properly estimated coefficients, adjusted for multicollinearity, and provides the best predictive variables selection.

The results are further confirmed by including time and bank fixed effects. The fixed effects mitigate endogeneity issues resulting from the usual presence of unobserved explanatory variables correlated with independent variables included in the model, simultaneity, and measurement errors. While there is a number of methods that help alleviating endogeneity – see excellent summary review in Roberts and Whited (2012) – including instrumental variables, difference-in-differences estimators and others, panel estimation with fixed effects gained enormous popularity in corporate finance. It became, perhaps, a discipline standard. When dataset is sufficiently large, the method may be implemented by running OLS regression with added intercepts for each firm and each period on the data panel. Roberts and Whited point out the deficiencies of the method, as well. Sometimes, the most important part of the variation is taken away with fixed effects. Therefore, we do not suggest solely resorting to the fixed effects estimates. We view a combined analysis of lasso and fixed effects estimates as the most relevant.

We use two base specifications defined in Sorokina et, al (2017) and add the components of complete second order models to the base.

We apply lasso and select the most relevant variables in two models for each component of leverage. Then we run those models with time and bank fixed effects for the full sample and in size quintiles. The resulting estimates retain great explanatory power, incorporate non-linearity in important determinants, and avoid issues of the effect misattribution.

TABLE 3
COMPLETE SECOND ORDER MODELS FOR COMPETITION, DIVERSIFICATION AND LIQUIDITY AS DETERMINANTS OF NON-DEPOSIT, SHORT-TERM AND EXCESS LEVERAGE OF BANKS

	Excess leverage			Non-deposit leverage			Short-term leverage		
	Estimate	VIF	Estimate	Estimate	VIF	Estimate	Estimate	VIF	VIF
Intercept	9.29E-02	***	-7.39E-03	*	0.00	0.00	1.37E-02	***	0.00
Competition	8.99E-02	***	5.75E-01	***	17.38	17.38	1.34E-01	***	17.38
Competition ²	-1.44E-01	***	-4.85E-01	***	18.07	18.07	3.67E-02	**	18.07
Size	-8.08E-07	***	2.95E-07	***	36.14	36.14	4.38E-07	***	36.14
Size ²	2.74E-13	***	-5.01E-13	***	4.66	4.66	-2.68E-13	***	4.66
Competition*Size	9.29E-02	***	6.96E-07	***	46.07	46.07	5.09E-08	***	46.07
Intercept	5.95E-02	***	2.99E-01	***	0.00	0.00	2.62E-01	***	0.00
Diversification	1.31E-05	***	-6.87E-05	***	36.22	36.22	-6.03E-05	***	36.22
Diversification ²	-8.53E-10	***	6.21E-09	***	34.33	34.33	4.23E-09	***	34.33
Size	-8.93E-09	***	1.25E-07	***	38.78	38.78	4.23E-07	***	38.78
Size ²	3.40E-13	***	-5.46E-13	***	3.66	3.66	-1.77E-13	***	3.66
Diversification*Size	-1.12E-10	***	1.69E-10	***	35.74	35.74	-1.25E-11	***	35.74
Intercept	1.07E-01	***	1.80E-01	***	0.00	0.00	7.81E-02	***	0.00
Liquidity	3.18E-03	***	4.83E-01	***	3.54	3.54	7.61E-02	***	3.54
Liquidity ²	-7.43E-02	***	4.54E-01	***	3.36	3.36	2.51E-02	***	3.36
Size	-4.92E-07	***	4.20E-07	***	4.22	4.22	4.99E-07	***	4.22
Size ²	4.02E-13	***	-3.78E-13	***	3.49	3.49	-4.46E-13	***	3.49
Liquidity*Size	2.64E-07	***	-2.03E-06	***	1.96	1.96	-2.42E-06	***	1.96

EMPIRICAL RESULTS

The results of the complete second order models for all three dependent variables and three key determinants combined with bank's size are presented in Table 3. The non-linearity is present in all models. We visualize the effects by graphing the estimated relationship for the full range of values of independent variables – presented in Figures 1-3. The variable construction for all three determinants requires “flipped sign” interpretation: the greater value of the variable means less competition, diversification and liquidity, respectively. Excess leverage is the highest for the largest banks with the greatest market power and the smallest for the largest banks with the least market power. The profile is very flat for small and medium banks across all levels of competitiveness. The relationship with diversification is the other way around. Large diversified banks have the lowest excess leverage, while large, concentrated banks carry more excess leverage. Similarly to competition, the profile is flat for small and medium banks. The liquidity's impact on excess leverage is almost as competition's, just less pronounced for most liquid largest banks – they carry somewhat less excess leverage. The least liquid largest banks have much more excess leverage than other groups. There seems to be not much difference in excess leverage across the range of liquidity values for small and medium banks.

FIGURE 1
NON-LINEARITY IN RELATIONS OF EXCESS BANK LEVERAGE WITH COMPETITION, DIVERSIFICATION AND LIQUIDITY

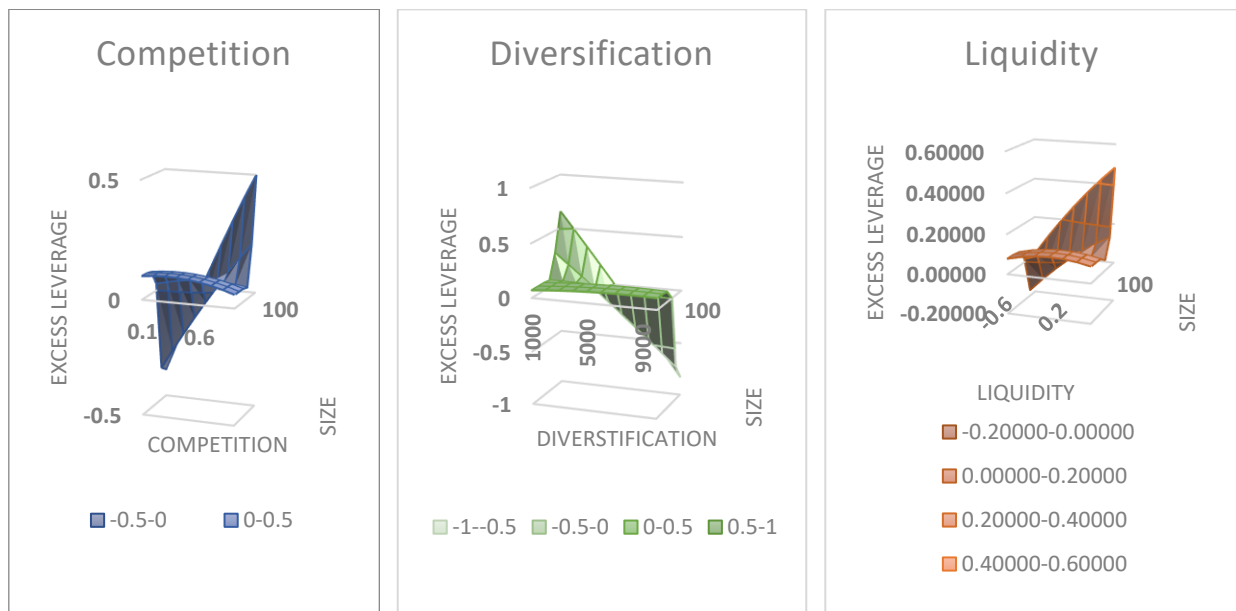


FIGURE 2
NON-LINEARITY IN RELATIONS OF NON-DEPOSIT BANK LEVERAGE WITH
COMPETITION, DIVERSIFICATION AND LIQUIDITY

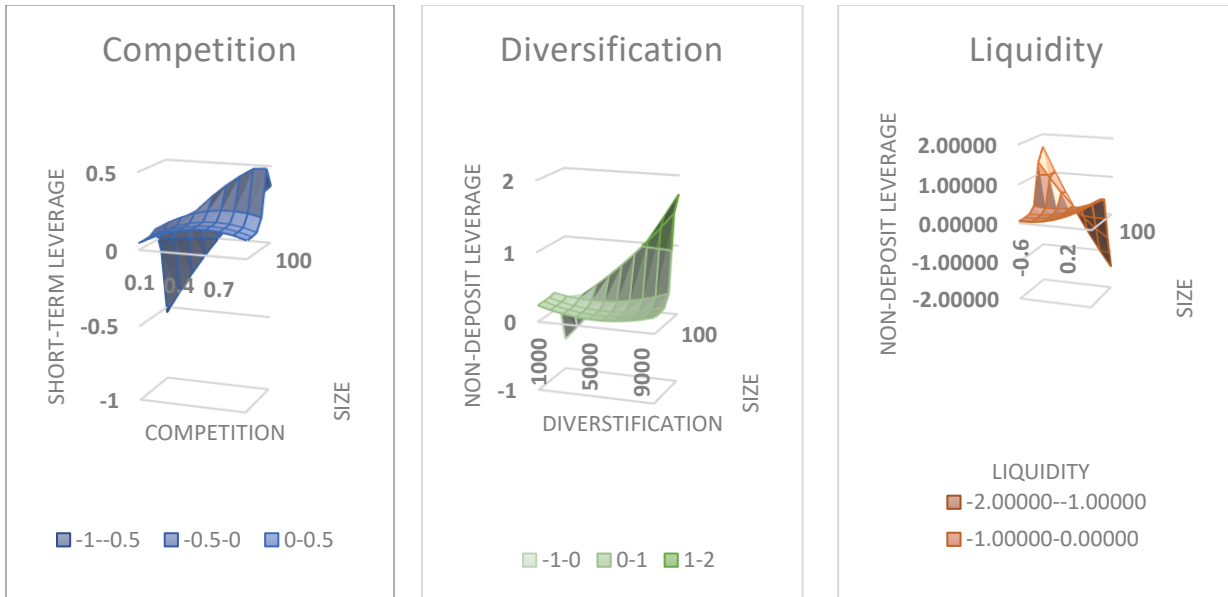
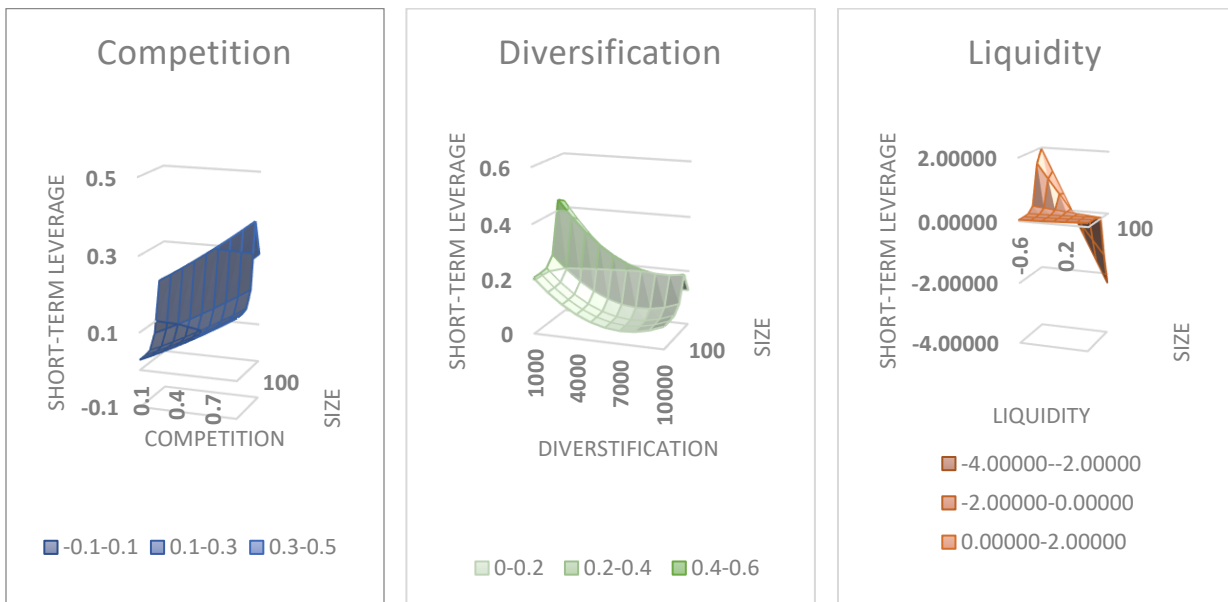


FIGURE 3
NON-LINEARITY IN RELATIONS OF SHORT-TERM BANK LEVERAGE WITH
COMPETITION, DIVERSIFICATION AND LIQUIDITY



The non-deposit leverage is lowest for the largest banks with the least market power, it is the highest for the large banks with higher market power. It declines substantially for the banks with the greatest market power in all sizes. Least diversified largest banks hold greatest non-deposit leverage just as the most liquid largest banks; least liquid largest banks carry least amounts of non-deposit leverage. For small and medium-sized banks, the amount of non-deposit leverage increases slightly with the decline in liquidity.

The short-term leverage generally increases as competition declines and size increases, but it drops for the largest banks. It declines with a decrease in diversification and is fairly flat for smaller-to-medium banks. However, short-term leverage is greater for large, but not the largest banks. The relationship of short-term leverage with liquidity is not evident for the small and medium banks in our sample. Interestingly, large and liquid banks borrow a lot short-term, which is not the case for the large illiquid banks.

The results of the lasso variable selection for the full leverage models are presented in Table 4 and further analysis of the models selected by lasso with fixed effects is in Table 5. The non-linear variables are selected as significant for all models. High order liquidity variable and interaction of diversification with size are important for excess leverage; second order diversification and liquidity, and interactions of competition and diversification with size are significant for non-deposit leverage; all three second-order variables and three interactions appear in two specifications explaining short-term leverage. The majority of results differ in the analysis with fixed effects. The coefficients of the same sign and similar in size to lasso estimates include squared liquidity and size in the second specification of excess leverage, squared liquidity and diversification in the first specification for non-deposit leverage; all non-linear coefficients for short-term leverage. We deem these results most reliable. However, the remaining results should not be dismissed, but rather carefully analyzed from the economic point of view. It is worth mentioning that the R-squared are notably high: for the excess leverage models, they are around 80%, while for short-term leverage - almost 70%.

TABLE 4
DETERMINANTS OF LEVERAGE COMPONENTS – LASSO VARIABLE
SELECTION AND ESTIMATES

Panel A: Specification 1					
Excess leverage		Non-deposit leverage		Short-term leverage	
Profit	-0.7149	Profit	-2.1407	Profit	-0.1931
LogSize	-0.0120	LogSize	0.0174	LogSize	0.0167
lnsize2	-0.0001	Growth	-0.0012	Collateral	-0.0350
Collateral	-0.0200	CapEx	0.2804	Growth	-0.0004
Growth	-0.0013	SGA	-0.1470	Inflation	-0.0299
MedIndLeverage	-0.0261	Inflation	-0.0920	GDPGrowth	0.0931
Inflation	-0.0443	GDPGrowth	0.1577	CompetLerner	0.0389
GDPGrowth	0.2922	CompetLerner	0.0775	Diversification2	5.53E-10
LiquidBank	0.0201	Diversification	-2.26E-05	LiquidBank	0.0273
LiquidBank2	-0.0506	Diversification2	2.65E-09	CompLnSize	0.0074
DivLnSize	-9.44E-08	LiquidBank	0.3299	DivLnSize	-1.49E-06
		LiquidBank2	0.2247	LiqLnSize	-0.0047
		DivLnSize	-5.17E-07		

Panel B: Specification 2					
Excess leverage		Non-deposit leverage		Short-term leverage	
MTB	0.0445	MTB	0.5483	MTB	0.2780
LogSize	-0.0119	Collateral	0.0360	LogSize	0.0071
Collateral	-0.0131	Dividend	-0.0006	Collateral	-0.0189
Dividend	-0.0056	MedIndLeverage	-0.0050	MedIndLeverage	-0.0064
MatureDummy	-0.0085	MktReturn	0.0273	DebtCapacity	-0.0122
MedIndLeverage	-0.0236	CRSPVW	-0.0446	CompetLerner2	0.0697
MktReturn	0.0181	CompetLerner2	0.0400	LiquidBank2	-0.0236
CRSPVW	0.0104	Insize2	0.0004	CompLnSize	0.0007
Insize2	-0.0001	Diversification2	1.13E-09	DivLnSize	-7.51E-07
LiquidBank	0.0109	LiquidBank	0.1648	LiqLnSize	-0.0099
LiquidBank2	-0.0620	LiquidBank2	0.0816		
DivLnSize	-1.35E-07	CompLnSize	0.0016		
		DivLnSize	-1.14E-06		

TABLE 5
THREE TYPES OF LEVERAGE WITH NON-LINEAR DETERMINANTS: FIXED EFFECTS MODELS

Panel A: Specification 1		Excess leverage		Non-deposit leverage		Short-term leverage		
Profit	0.2021	***	Profit	-0.8899	***	Profit	0.4542	***
LogSize	-1.0022	***	LogSize	0.0287	***	LogSize	0.0028	
Insize2	0.0047	***	Growth	0.0015	***	Collateral	-0.0292	***
Collateral	-0.0012	***	CapEx	0.3848	**	Growth	-0.0009	**
Growth	-0.015	***	SGA	-0.0454	***	Inflation	-0.0265	
MedIndLeverage	0.0012	***	Inflation	-0.1525		GDPGrowth	-0.0073	
Inflation	-0.0223	***	GDPGrowth	-0.0135		CompetLerner	-0.2044	***
GDPGrowth	-0.4178	***	CompetLerner	0.0783	***	Diversification2	7.70E-10	***
LiquidBank	0.0381	***	Diversification	-9.16E-06	**	LiquidBank	0.1596	***
LiquidBank2	0.0152	***	Diversification2	8.09E-10	***	CompLnSize	0.031	***
DivLnSize	-0.0418	***	LiquidBank	0.1805	***	DivLnSize	-1.54E-06	***
			LiquidBank2	0.0046		LiqLnSize	-0.0197	
			DivLnSize	1.11E-08			0.4542	
R2	0.8127			0.8203			0.6828	
Adj R2	0.8052			0.8131			0.67	
No obs	16459			16459			16459	

Panel B: Specification 2

	Excess leverage	Non-deposit leverage	Short-term leverage
MTB	-0.0076 ***	MTB 0.3553 ***	MTB 0.1986 ***
LogSize	0.0095 ***	Collateral 0.0165 ***	LogSize 0.0059 ***
Collateral	-0.0146 ***	Dividend -0.0013	Collateral -0.0058
Dividend	-0.0057 ***	MedIndLeverage -0.0052	MedIndLeverage -0.0084 **
MatureDummy	-0.004 ***	MktReturn 0.0422 *	DebtCapacity -0.0434 ***
MedIndLeverage	-0.0272 ***	CRSPVW -0.0674	CompetLerner2 0.083 ***
MktReturn	-0.0219 **	CompetLerner2 0.0385 ***	LiquidBank2 -0.2226 ***
CRSPVW	0.0844 ***	Insiz2 0.0011 ***	CompLnSize -0.0038 **
Insiz2	-0.0013 ***	Diversification2 -2.52E-10 **	DivLnSize -3.44E-07 ***
LiquidBank	0.0192 ***	LiquidBank 0.0858 ***	LiqLnSize -0.0128 ***
LiquidBank2	-0.0386 ***	LiquidBank2 -0.0114	
DivLnSize	1.64E-07 ***	CompLnSize 0.0026 *	
		DivLnSize 7.08E-07 ***	
R2	0.8102	0.8502	0.7023
Adj R2	0.8026	0.8442	0.6903
No obs	16459	16459	16459

TABLE 6
NON-LINEAR DETERMINANTS OF BANK LEVERAGE IN SIZE QUINTILES

Panel A: Excess Leverage - Specification 1					
	Quantile1	Quantile2	Quantile3	Quantile4	Quantile5
Profit	-0.5576 ***	-0.8207 ***	-1.5891 ***	-0.9746 ***	-0.2274 *
LogSize	0.0745 ***	0.0551 *	0.0607	-0.0087	0.0041
Insize2	-0.0073 ***	-0.005 **	-0.0054 **	0.0001	-0.0007 **
Collateral	-0.0248 ***	-0.034 ***	-0.0461 ***	-0.0196 ***	0.0039
Growth	0.001 ***	0.0018 ***	0.0013 ***	0.001 ***	0.0019
MedIndLeverage	-0.0388 ***	-0.0225	-0.0405	-0.0419 **	-0.0273 ***
Inflation	-0.3076 **	-0.397 ***	-0.5693 ***	-0.4112 ***	-0.3268 **
GDPGrowth	0.0181	0.0343 **	0.051 ***	0.0345 **	0.046 ***
LiquidBank	0.0024	-0.0471 ***	-0.0275 ***	0.0233 ***	0.0413 ***
LiquidBank2	-0.0536 ***	-0.1306 ***	-0.0327 ***	0.0333 **	-0.0223
DivLnSize	-1.75E-07 *	3.74E-07 ***	-7.68E-08	3.65E-07 ***	1.84E-07 ***
R2	0.8023	0.8208	0.8112	0.7822	0.7903
Adj R2	0.783	0.8027	0.7955	0.7664	0.7797
No obs	3263	3308	3305	3308	3275
Panel B: Excess Leverage - Specification 2					
	Quantile1	Quantile2	Quantile3	Quantile4	Quantile5
MTB	-0.0027	-0.0222 ***	-0.0156 **	-0.0076	-0.0046
LogSize	0.0792 ***	0.0699 **	0.0697 *	-0.0116	0.0031
Collateral	-0.0272 ***	-0.0344 ***	-0.0457 ***	-0.0172 ***	0.0061
Dividend	-0.004 ***	-0.01 ***	-0.0049 ***	-0.0036 ***	-0.0017
MatureDummy	-0.0058 ***	-0.0043 ***	-0.0027 *	-0.0023	0.0308 ***
MedIndLeverage	-0.0407 ***	-0.056 *	-0.0439	-0.0441 **	-0.0249 ***
MktReturn	-0.0136	-0.0205	-0.0209	0.004	-0.0268
CRSPVW	0.0557	0.0776	0.1005 *	0.0378	0.082 *

Insize2	-0.0076	***	-0.0056	**	-0.0058	**	0.0004	**	-0.0006	**
LiquidBank	-0.0011	***	-0.0406	***	-0.0225	**	0.0346	***	0.045	***
LiquidBank2	-0.0509	***	-0.1438	***	-0.0336	*	0.0534	***	-0.0108	***
DivLnSize	-1.63E-07	*	3.07E-07	***	-2.92E-08		3.72E-07	***	2.13E-07	***
R2	0.8026		0.8245		0.8035		0.7782		0.7919	
Adj R2	0.7834		0.8068		0.7871		0.7621		0.7813	
No obs	3263		3308		3305		3308		3275	

Panel C: Non-Deposit Leverage - Specification 1

	Quantile1	Quantile2	Quantile3	Quantile4	Quantile5					
Profit	-0.8768	***	-1.4663	***	-1.3366	***	-0.1563	**	0.3884	
LogSize	-0.0067		0.0139		0.075	***	0.0653	***	0.0063	***
Growth	0.0011		0.001		0.0029	***	-0.0002	**	0.0035	**
CapEx	-1.2138	***	1.6928	***	0.4429		-1.6317	***	0.2788	
SGA	0.0088		0.0246		-0.0699	**	-0.0998	***	0.0285	*
Inflation	-0.2977		-0.212		-0.6789	**	-0.3238		0.3624	
GDPGrowth	-0.0425		0.0077		0.0225		-0.0344		0.0412	
CompetLerner	0.071	***	0.1145	***	0.0198		0.0079		0.0149	***
Diversification	-2.65E-05	***	-1.31E-05	***	0.0001	***	-1.80E-05	***	1.51E-05	***
Diversification2	1.45E-10		-1.41E-09	***	2.00E-10		1.69E-09	***	6.44E-10	
LiquidBank	0.1787	***	0.1111	***	0.0987	***	0.1197	***	0.0171	***
LiquidBank2	0.1749	***	-0.0114		-0.0928	**	-0.0308		0.045	***
DivLnSize	3.90E-06	***	4.62E-06	***	-1.05E-05	***	-7.98E-07	***	1.17E-06	***
R2	0.8705		0.8642		0.8535		0.8722		0.7874	
Adj R2	0.8579		0.8504		0.8412		0.8628		0.7765	
No obs	3263		3308		3305		3308		3275	

Panel D: Non-Deposit Leverage - Specification 2

	Quantile1	Quantile2	Quantile3	Quantile4	Quantile5
MTB	0.2952 ***	0.2923 ***	0.3124 ***	0.3068 ***	0.303 ***
Collateral	0.0316 ***	-0.0083	0.0168	0.0329 ***	-0.0377 ***
Dividend	0.0017	-0.0091 ***	-0.0014	0.0009	0.0106 ***
MedIndLeverage	-0.0485 ***	0.2907 ***	0.3581 ***	-0.0151	-0.0017
MktReturn	0.0728	0.0402	-0.0575	0.0517	0.0193
CRSPVW	-0.1178	-0.0427	0.1816 *	-0.0573	0.001
CompetLerner2	0.0632	0.1188 ***	0.0531	0.1154 ***	0.1236 ***
Insize2	0.0007	0.0029 ***	0.0011 **	0.004 ***	0.0013 ***
Diversification2	-7.01E-10 *	-7.08E-10 *	5.90E-10	6.52E-10	8.50E-10 **
LiquidBank	0.1186 ***	0.0021	-0.0057	0.0857 ***	0.1168 ***
LiquidBank2	0.1443 ***	-0.0655 *	-0.1496 ***	0.0511	0.0516
CompLnSize	-0.0026	-0.0072 *	-0.0092 **	-0.0099 ***	-0.0052 *
DivLnSize	1.22E-06	1.37E-06 *	3.30E-07	-1.25E-06 **	-1.19E-06 ***
R2	0.8828	0.8772	0.8668	0.8879	0.8242
Adj R2	0.8713	0.8647	0.8557	0.8797	0.8152
No obs	3263	3308	3305	3308	3275

Panel E: Short Term Leverage - Specification 1

	Quantile1	Quantile2	Quantile3	Quantile4	Quantile5
Profit	-0.83 ***	0.266	-0.1413	1.0154 ***	0.7815 **
LogSize	-0.0144	0.0204 **	-0.0283 ***	0.0613 ***	0.0028
Collateral	-0.0119	-0.0558 ***	-0.0284 **	-0.0135	-0.0283 **
Growth	4.25E-05	-0.0005	0.0011	-0.0023 **	-0.0023
Inflation	0.4531	-0.2191	-0.2715	-0.2783	-0.3497
GDPGrowth	-0.0132	0.0274	-0.0161	-0.0011	-0.0119
CompetLerner	-0.1522	0.0511	-0.077	0.4674 ***	-0.0917
Diversification2	-6.67E-10	7.37E-10 *	-1.40E-09 **	3.77E-09 ***	3.22E-09 ***

LiquidBank	-0.0738	0.0699	0.508	0.6741	0.7616
CompLnSize	0.0335 *	-0.0018	0.0112	-0.0572	0.0151 **
DivLnSize	2.80E-07	-2.35E-06 ***	2.47E-06 ***	-5.13E-06 ***	-3.32E-06 ***
LiqLnSize	0.0137	-0.0154	-0.0689 ***	-0.0774 ***	-0.0726 ***
R2	0.4991	0.6298	0.7598	0.6624	0.7416
Adj R2	0.4502	0.5922	0.7397	0.6377	0.7284
No obs	3263	3308	3305	3308	3275

Panel F: Short Term Leverage - Specification 2

	Quantile1	Quantile2	Quantile3	Quantile4	Quantile5
MTB	0.1566 ***	0.0656 ***	0.1195 ***	0.1948 ***	0.251 ***
LogSize	-0.0066	0.0108 **	-0.0067	0.0105 **	0.0171 ***
Collateral	-0.0063	-0.0589 ***	-0.0228 **	0.0077	0.005
MedIndLeverage	-0.057 ***	0.0258	0.081	-0.0738	0.0024
DebtCapacity	-0.1388	0.2492 **	-0.0117	0.0341	-0.0001
CompetLerner2	-0.1382 ***	0.071 *	-0.0908 **	0.0953 ***	0.163 ***
LiquidBank2	-0.1164 ***	0.0082	-0.2605 ***	-0.2942 ***	-0.4 ***
CompLnSize	0.0165 ***	-0.0019	0.0069 *	-0.0089 **	-0.012 ***
DivLnSize	-1.06E-06 ***	-9.70E-07 ***	2.88E-07 **	3.04E-07 *	-4.15E-07 **
LiqLnSize	-0.0112 ***	-0.007 **	-0.0147 ***	-0.0131 ***	-0.0124 ***
R2	0.5157	0.6322	0.7668	0.6715	0.7653
Adj R2	0.4687	0.5949	0.7474	0.6477	0.7534
No obs	3263	3308	3305	3308	3275

The results of the same fixed effects models in size quintiles yield some interesting observations. The explanatory power of the models is quite consistent in the case of excess leverage. For non-deposit leverage, it deteriorates for the largest banks. The explanatory power of the models for short-term leverage consistently increases with size.

For excess leverage, the squared liquidity's and squared size's coefficients in the first specification are negative and significant only for the first three quintiles. The interaction of diversification and size is significant and positive in the last two quintiles. The results for the second specification are fairly similar.

For the non-deposit leverage, squared diversification is negative and significant for the second quintile, but positive and significant for the fourth quintile; squared liquidity is positive and significant for the largest and smallest banks, while negative and significant for the mid-sized third quintile. The interaction of diversification and size is positive and significant for smaller and for the largest banks; it is negative and significant for the mid-sized banks. All results are similar, but somewhat weaker in the second specification, except squared size, which is consistently positive and significant, except for the smallest banks' group.

For the short-term leverage, squared diversification is negative and significant for the mid-sized banks, while positive and significant for the larger banks. Interaction of competition with size is only consistently negative for the fourth quintile in both specifications, interaction of liquidity and size is negative and significant for most categories, interaction of diversification with size is negative for the smaller and larger banks, while positive for the mid-sized group. Squared diversification and competition are positive for larger banks; squared liquidity is negative, where significant. As we can see, there is a substantial variability in the relationship between non-linear determinants and leverage components in various size groups of banks.

CONCLUSIONS

We contribute to the empirical corporate finance literature in two key dimensions: advancing understanding of bank capital structure decisions and introducing a novel methodological approach. Our findings demonstrate that competition, diversification, and liquidity play an important role in explaining banks' borrowing behavior as financial intermediaries—specifically, their excess leverage, the non-deposit portion of their borrowings, and reliance on short-term debt.

The study supports our initial hypothesis regarding the non-linear nature of these relationships and reveals significant interactions between leverage determinants and bank size. Furthermore, we uncover non-linearity in the relationship between bank size and leverage components. In the new models derived through a two-step estimation process, interaction and second-order terms replace several variables that appeared significant under conventional linear model specifications.

Compared to the results obtained using fixed effects regressions alone (Sorokina et al., 2017)—a standard approach in corporate finance—our non-linear models achieve comparable or superior explanatory power while utilizing a more parsimonious set of predictors. We find that our method helps avoid the misattribution of effects that are better captured through interaction and non-linear terms. In fixed effects models, such effects are often erroneously attributed to unrelated variables. Notably, model results vary substantially across bank size quintiles, even though bank size is explicitly accounted for.

We also emphasize the value of benchmarking fixed effects models against those generated via lasso selection, rather than relying solely on fixed effects estimation. This comparison enables us to identify and interpret variables selected by lasso that may not appear significant once bank and time fixed effects are included. In doing so, we highlight potentially important determinants that could otherwise be obscured. Our approach aligns with the caution that Roberts and Whited (2012) advocated regarding the indiscriminate use of fixed effects in empirical research.

ACKNOWLEDGEMENTS

We appreciate comments and suggestions received from Elyas Elyasiani and the participants of the session at MWDSI'17 conference. We are also thankful for the excellent research assistantship provided

by Liam Connors. Supported by the Penn State Institute for Computational and Data Sciences (RRID: SCR_025154), and Penn State University Center for Applications of Artificial Intelligence and Machine Learning to Industry Core Facility Fellowship (RRID: SCR_022867).

REFERENCES

- Acharya, V.V., Hasan, I., & Saunders, A. (2006). Should Banks Be Diversified? Evidence from Individual Bank Loan Portfolios. *The Journal of Business*, 79(3), 1355–1412.
- Acharya, V.V., & Mora, N. (2015). A crisis of banks as liquidity providers. *The Journal of Finance*, 70(1), 1–43.
- Baker, M., & Wurgler, J. (2002). Market timing and capital structure. *Journal of Finance*, 57, 1–32
- Berger, A.N., Flannery, M.J., Lee, D., & Öztekin, Ö. (2008). How do large banking organizations manage their capital ratios? *Journal of Financial Services Research*, 34(2–3), 123–149.
- Berger, A.N., Klapper, L.F., & Turk-Ariss, R. (2009). Bank Competition and Financial Stability. *Journal of Financial Services Research*, 35, 99–118.
- Demsetz, R.S., & Strahan, P.E. (1997). Diversification, Size, and Risk at Bank Holding Companies. *Journal of Money, Credit & Banking*, 29(3). Ohio State University Press.
- Fischer, E.O., Heinkel, R., & Zechner, J. (1989). Dynamic capital structure choice: Theory and tests. *Journal of Finance*, 44(1), 19–40.
- Frank, M.Z., & Goyal, V.K. (2009). Capital structure decisions: Which factors are reliably important? *Financial Management*, 38(1), 1–37.
- Gropp, R., & Heider, F. (2010). The Determinants of Bank Capital Structure. *Review of Finance*, 14(4), 587–622.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, Springer-Verlag.
- Jung, K., Kim, Y-C., & Stulz, R.M. (1996, October). Timing, Investment opportunities, managerial discretion, and the security issue decision. *Journal of Financial Economics*, 42(2), 159–185
- Lang, L.H.P., & Stulz, R.M. (1994, December). Tobin's Q, Corporate Diversification and Firm Performance. *Journal of Political Economy*, 102(6), 1248–1280.
- Leary, M.T., & Roberts, M.R. (2010). The pecking order, debt capacity an information asymmetry. *Journal of Financial Economics*, 95, 332–355.
- Lemmon, M.L., & Zender, J.F. (2010, October). Debt capacity and tests of capital structure theories. *Journal of Financial and Quantitative Analysis*, 45(5), 1161–1187.
- Marquardt, D.W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12, 591–256.
- Rajan, R.G., & Zingales, L. (1995, December). What Do We Know about Capital Structure? Some Evidence from International Data. *Journal of Finance*, 50(5), 1421–1460.
- Snee, R.D., & Marquardt, D.W. (1984). Comment: Collinearity diagnostics depend on the domain of prediction, the model, and the data. *The American Statistician*, 38(2), 83–87.
- Sorokina, N.Y., Thornton Jr., J.H., & Patel, A. (2017). *Bank Leverage Components*. Working paper.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.
- Titman, S., & Wessels, R. (1988). The determinants of capital structure choice. *The Journal of Finance*, 43(1), 1–19.
- Roberts, M.R., & Whited, T.M. (2012). Endogeneity in empirical corporate finance. *The Bradley Policy Research Center Financial Research and Policy Working Paper No. FR 11-29*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1748604 accessed 03/31/2017
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429.