Determining the probability of default for Agrifinancial's loan portfolio

by

Kaci E. Ferch

B.S., Iowa State University, 2012

A THESIS

Submitted in partial fulfillment of the requirements

for the degree

MASTER OF AGRIBUSINESS

Department of Agricultural Economics

College of Agriculture

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2023

Approved by:

Major Professor
Dr. Allen Featherstone

ABSTRACT

Credit risk rating models are tools used by nearly every financial organization across the world. These models allow for organizations to measure risk more accurately within their credit portfolios, ensuring more confident credit approvals and capital requirements in the event default may occur. AgriFinancial is proposing the use of a credit risk rating system to make funding decisions using capital from its parent company in addition to participation funding that is currently the sole source of funds.

This thesis evaluates AgriFinancial's portfolio risk of originated and serviced loans, using ten years of historical data to determine the probability of default for the portfolio of originated loans and if the variables considered are statistically significant. The evaluation includes the estimation of a binomial logit regression model of default defined as loans that are ninety days or more past due. Credit factors measured at the time of loan origination are independent variables that predict default. Regression results were applied to the portfolio of loans to calculate their respective probability of default.

Overall, the model has explanatory power in predicting the probability of default. Originally, eleven variables were considered; two were removed due to multicollinearity and a ratio used more predominately for loss given default, a default risk calculation not considered. It was concluded that four of the nine independent variables were statistically significant in predicting that a loan will default at some time during the life of that loan. One variable was statistically significant at the 1% level, two variables at the 5% level, and one at the 10% level. Credit score has the highest statistical significance at the 1% level for determining probability of default. Signs for the statistically significant independent variables were correctly predicted; pro forma debt/asset positive resulting in a higher

probability of default as the variable increases, three year average total debt coverage negative resulting in a lower probability of default as the variable increases, credit score negative resulting in a lower probability of default as the variable increases, and the original principal positive resulting in a higher probability of default as the variable increases. Based on the results, recommendations were made regarding implementation of the model and the advantage of proceeding with the new business opportunity.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The past two and a half years as a master's student have taught me more about time management, dedication, and my capabilities than I ever expected. The reality of having multiple titles in my professional and personal life made for long, difficult days. Without the help and support of my husband, in-laws, parents, and siblings, completing this degree would not have been possible, especially with the travel to Manhattan for on-campus sessions. A heart-felt thank you to them all. At the end of the day, I dedicate this thesis to my two daughters. One day they can look back and understand the time I spent dedicated to this and the characteristics I gained to also be a better mother.

Relationships gained with the MAB staff and cohorts will continue to be reflected on for years to come. The friendships and time spent talking through readings, assignments, and projects made the late nights easier. I wish everyone well and the best of luck in their future endeavors!

A huge thank you to my company for supporting me through this journey and allowing me the opportunity to grow my knowledge for the betterment of the organization. I look forward to continuing to apply new concepts and theories to expand and improve our business.

Last, but certainly not least, my major professor and committee members. Thank you for your time and knowledge shared with me throughout this process. I could not have completed this without your help and confidence along the way.

CHAPTER I: INTRODUCTION

AgriFinancial "AgFi", based in Louisville, Kentucky, is a mortgage company that originates and services agricultural loans and leases throughout the United States. Currently, AgFi's servicing asset portfolio is over $2 billion that includes AgFi originated loans from direct customer and correspondent relationships. Funds for originated loans and leases are sourced by participations from the Farm Credit System, the Federal Agricultural Mortgage Company, the Farm Service Agency, and banks. Given the company has only used funds from participations, credit risk models for the company's originated loans and leases have not been a part of the origination and servicing process. Currently, the five "Cs" of credit, Character, Capital, Capacity, Collateral, and Conditions, are considered to make a decision for each individual request. The five "Cs" identify key ratios, scores, and other identifiers lenders have historically benchmarked and used to make credit decisions. Data from approved requests has been kept; however, it has not been evaluated or used to develop credit risk models.

Consolidated Grain and Barge (CGB), AgriFinancial's parent company, has discussed opportunities for AgFi to begin funding loans using capital from CGB's shareholders. Further evaluation and understanding of AgFi's portfolio risk of originated and serviced loans has the potential to open this new business opportunity, as well as others, for the organization. Identifying the current levels of risk will give leaders an idea of agricultural credits the organization prefers to fund, pricing structures, their overall level of risk tolerance, and capital reserves if losses may occur, as credit risk models go beyond origination.

The objective for this thesis includes creating a probability of default model using historical loan data to determine risk assessment for each loan submitted for a credit decision through AgFi's current and future funding programs. In addition to the primary objective, the thesis objective is to determine what loan portfolio information, such as underwriting ratios needed to make credit decisions, are statistically significant in predicting if loans will or will not default at the time of origination.

For this thesis, historical loan data from AgFi originated loans and servicing portfolio, years 2012 through 2021, is used to predict an originated loan's probability of default. Key financial variables considered include both historical and pro forma data collected from documents received at application. These include credit score, solvency, liquidity, and repayment capacity measures for traditional underwriting programs; scored or seasoned programs were not considered. Binary logit regression is used to determine coefficients for the variables to calculate probability of default and determine statistical significance.

Relationships with AgFi's participants and research indicates that nearly every agricultural lender has some sort of credit risk model for their organization (Featherstone, Roessler and Barry 2006). Having a probability of default model for future AgFi originated loans will allow the organization to capitalize on risk assessment for a more competitive advantage in the agricultural finance marketplace. Competitive advantages include more profitability from direct funding, pricing based on credit quality, and quicker credit decisions.

CHAPTER II: LITERATURE REVIEW

Agricultural producers throughout the United States use different financing entities for various capital needs in their operations. Production agriculture is capital intensive with more than 90% of total assets consisting of farm real estate and machinery (Katchova and Barry 2013). Financial organizations that offer capital to the agricultural industry include, but are not limited to, banks, mortgage companies, insurance companies, private investors, and the Farm Credit System. Over the years as competition in the agricultural finance industry has emerged, applicants typically desire quick turnaround, low interest rates, and specific products that fit their marketing plans and goals.

Farm real estate and equipment continues to increase in price year after year. Increases in these purchase prices and values lead to increases in capital required by agricultural producers. While, from a risk perspective, many agricultural lenders would prefer to do in-depth analyses for operations as they continue to grow and add debt, quicker turnaround times are desired by applicants. The process from application to close can be lengthy due to appraisal and title searches from third parties that occur post-approval. Considering the stages of a loan request, the credit decision and underwriting process is where a lender can improve efficiency. Risk rating models focused on credit underwriting ratios and other originating factors are an option that organizations can turn to for efficiency and better understanding of the request's overall credit risk, as well as portfolio risk and capital reserves necessary for such risks should default occur.

As noted in the introduction, nearly every agricultural lender has implemented some sort of credit risk model for their organization. Good risk rating models should improve process efficiency by reducing costs and freeing time for origination teams (Yu, Garside and Stoker 2001). Typically, these models are based on The Basel Committee on Banking Supervision's (BCBS) minimum standards, that include the following published accords on capital adequacy: Basel I: the Basel Capital Accord, Basel II: the new capital framework, and Basel III: responding to the 2007-09 financial crisis. After the BCBS's inception in 1975 to supervise financial organizations, these three series of accords have been the standard set for international bank regulation and continue to be revised as events occur throughout the financial industry (History of the Basel Committee n.d.).

Basel II: the new capital framework, led financial organizations throughout the world to re-evaluate their portfolio risks and measures for such risks, including agricultural lenders in the United States. One of the limiting factors for measuring a portfolio's credit risks is data. Under Basel II, probabilities of default and loss given default can be measured using internal or external data. Internal data must include loss and non-loss events from wide cross-section and lengthy time-series to produce reliable default measures. Specifically, at least five years of data history with a preference of longer series. External data must portray similar borrower characteristics to the organization's current loan portfolio (Barry and Katchova 2005). Also, the data "must reflect economic conditions in the past that are expected to generally prevail in the future. If economic conditions have changed significantly, then historically based measures of loss rates may have limited applicability in the future" (Barry 2001, 107).

The fundamental goal of a credit risk rating model is to estimate the credit risk of a given transaction or portfolio of transactions/assets (Yu, Garside and Stoker 2001). It is important that the model is intuitive to understand. By providing the organization with an integrated view of its credit risk, it should allow for useful decisions to be made by management based on solid estimation of the credit risk being taken. The overall goal is to provide a measure of the loss expected for funding a request and the capital required in the event of default. Most systems use probability of default (PD) and the loss given default (LGD) to achieve these goals, using separate quantified measures which is consistent with the Basel II guidelines (Yu, Garside and Stoker 2001). Both PD and LGD are usually expressed in percentage terms (Zech and Pederson 2004). One important aspect that organizations need to consider with a credit risk rating model is the level of granularity within the data being measured. Too small or large of granularity within the data used for the model will not yield accurate or useful decision making for the organization to use (Yu, Garside and Stoker 2001).

According to Barry, two major approaches to measuring credit risks include the default-mode and mark-to-market approach. Of these two approaches, the default-mode is more widely used by organizations as it focuses on the specific year of origination for determining probability of default and the severity of loss given default from historical data. The mark-to-market approach reviews credit risk variables and how any changes in the variables from year to year may affect the loan's current market or present value. Evaluation each year of a customer's financials is not currently consistent for portfolios, which is why it is not more widely used. The loan's risk may be upgraded, decreasing the loan rate, or downgraded, increasing the loan rate, from one year to the next depending on if the variables improve or not. If upgraded, the market or present value increases and alternatively decreases if downgraded. "Changes in the risk premium in the interest rate portray a clear linkage between the effects of credit risk and interest risks on the change in market value" (Barry 2001, 108).

After an organization has determined the probability of default model, the next step is to implement a risk rating scale. “A risk rating system uses an objective scale to rank credits according to risk” (Yu, Garside and Stoker 2001, 40). Scales can vary from letter to number ratings, depending on how the organization chooses to define the risks. The most effective ratings, however, should be applicable for an entire portfolio using a “master scale” to which all credit programs or products are mapped (Yu, Garside and Stoker 2001). Organizations need to outline how and who assigns the ratings, ensuring consistency to the approaches. Ratings must be validated to ensure that the scale is accurate and meets expectations for the organization’s goals it set out for the credit risk rating models. If reviewing the customers in the portfolio each year, credits may be upgraded or downgraded if the variables determining the risk improve or deteriorate, respectively.

In general, profitability, growth, and risk revolve around uniform capital distribution throughout a financial organization. The primary source of risk to financial organizations are credit risks and the capital reserves held in the event of loan losses are the reactions to such risks. Organizations must consider the following three questions when determining the relationship between the portfolio’s credit risk and adequate capital for these risks: How much risk is present? What is the institution’s tolerance to risk? How much capital should be held? Insight regarding the amount of capital needed to manage risks appropriately and safely can be answered with these questions (Barry 2001). Once these questions are answered from an overall credit risk rating scale, efficiencies can be applied to the underwriting process to improve speed and accuracy of the decision-making process. Overall, decreasing days from origination to closing are favorable for an organization and customer due to reduced costs and increased customer experience.

CHAPTER III: METHODS

The Basel Committee on Banking Supervision (BCBS) defines expected losses for a portfolio of loans as the composite formula, for each loan *i*, where *PD* is the probability of default measured as a percentage, *LGD* is the loss given default measured as a percentage, *EAD* is the exposure at default measured in currency units, and *EL* is the expected loss for a loan measured in currency units (Tauer and Dressler 2016).

$PD_i(\%) \times LGD_i(\%) \times EAD_i(\$) = EL_i(\$)$

This thesis will focus on the *PD* of the formula, probability of default. The *PD* measures and focuses on the odds that a loan, based on certain variables from past portfolio performance, may default if an organization takes on the risk by making a loan (Featherstone, Roessler and Barry 2006). As previously mentioned, the objective includes creating a probability of default model using historical loan data to determine risk assessment for each loan submitted for a credit decision through AgFi's current and future funding programs. In addition to the primary objective, it is determined what loan portfolio information, such as underwriting ratios to make credit decisions, are statistically significant in predicting if loans will default based on origination information. Binary logit regression is used to determine statistical significance and coefficients for the variables to calculate probability of default.

3.1 Regression Model

The binomial logit model is used to estimate the probability of default. The binomial logit model avoids the unboundness problem of the linear probability model by using a variant of the cumulative logistic function due to estimating equations with binary dependent variables. Binary variables are variables that are equal to only 0 or 1. Below is this equation, rewritten to where the right side of the equation resembles that of the linear probability model; although, the results are those described more of a S-shaped pattern since the dependent variable is bounded by zero and one (Studenmund 2016).

$$\ln(P_i/[1-P_i]) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

In the equation, P_i is the true probability of the binary variable is equal to one, specifically that a loan will default for this thesis. The estimated model used for the AgriFinancial data determining the probability of default risk assessment:

$$\ln(P_i/[1-P_i]) = \beta_0 + \beta_1\text{*ProformaDA} + \beta_2\text{*ProformaCR} + \beta_3\text{*ProformaTDC} + \beta_4\text{*ThreeYrAvgTDC} + \beta_5\text{*CreditScore} + \beta_6\text{*OriginalPrincipal} + \beta_7\text{*OriginalLTV} + \beta_8\text{*Crops} + \beta_9\text{*Livestock} + \beta_{10}\text{*PermanentPlanting} + \beta_{11}\text{*Other}$$

Statistical summaries of the data are in Table 3.1. The mean of all the independent variables meets or exceed the underwriting standards considered when making a loan approval. Underwriting standard minimums are outlined in Table 4.1. Percentages for the commodities identify the percentage of the loans that fall in each commodity category. Pro Forma Debt/Asset ratios typically range from 1 – 100%, Credit Scores from 300 – 900 depending on the reporting agency, and Loan to Value from 1 – 100%.

Interestingly, the loans that defaulted had averages for Pro Forma Debt/Asset, Original Principal, and Loan to Value higher than the mean at 35%, $1,429,646, and 40.92%, respectively. Also, the loans that defaulted had averages for Pro Forma Current Ratio, Pro Forma Total Debt Coverage, Three-Year Average Total Debt Coverage, and Credit Score lower than the mean at 4.33, 1.69, 1.64, 709, respectively. On the surface, these calculations compared to the mean may indicate an impact on PD. However, results from the PD model will determine whether these variables are statistically significant in determining PD.

Table 3.1 Distribution of Variables Used in Regession Model

Variable	Mean	Std. Dev.	Minimum	Maximum
Pro Forma Debt/Asset	29.07%	11.63%	2.00%	63.00%
Pro Forma Current Ratio	5.04	6.77	0.02	55.34
Pro Forma Total Debt Coverage	1.91	0.90	0.02	7.46
Three-Year Average Total Debt Coverage	2.06	1.17	-0.85	9.42
Credit Score	751	44.81	536.00	886.00
Original Principal	$834,175	$1,194,626	$35,000	$13,200,000
Loan to Value	38.18	55.09	0.00	80.00%
Crops	73%	-	-	-
Livestock	14%	-	-	-
Permanent Planting	11%	-	-	-
Other	2%	-	-	-
Note: Number of Observations = 3,323				

3.2 Variable Definitions and Expected Coefficient Signs

The dependent variable, Default90Day, is a binary variable that represents “0” if the loan has not entered default and “1” if the loan has entered default. Default is defined as a payment being 90 days or more past due at least once since the origination date for the ten-year time frame studied. Of the 3,323 loans used for this thesis, 55 defaulted or 1.655% of the loans.

The independent variables include: ProformaDA (DA = Debt/Asset Ratio), ProformaCR (CR = Current Ratio), ProformaTDC (TDC = Total Debt Coverage), ThreeYrAvgTDC, CreditScore, OriginalPrincipal, OriginalLTV (LTV = Loan to Value), Crops, Livestock, Permanent Planting, and Other. Pro forma is identified as projections, based on historical and future goals, that include the proposed financing. This may impact the balance sheet ratios, ProformaDA and ProformaCR, as well as the cash flows, ProformaTDC and ThreeYrAvgTDC, depending on the credit request. The signs for each variable's coefficient will either be positive or negative for the regression output. Signs indicate if the probability will increase or decrease for each variable, when holding all the other variables constant. Prediction for the independent variable's signs is detailed under each variable below.

Pro Forma Debt/Asset Ratio – Debt/asset ratio is the organization's measure for a customer's solvency position, based on the asset's market values. This ratio is calculated by taking total debts divided by total assets. The underwriting standard for this ratio is less than or equal to 50%. A higher ratio/percentage indicates a lower equity position. Therefore, it is predicted that this sign will be positive, increasing the probability of default.

Pro Forma Current Ratio – Current ratio is the organization's measure for a customer's liquidity position. This ratio is calculated by taking total current assets divided by total current liabilities. The underwriting standard for this ratio is greater than or equal to 1.25:1 or 125%. A higher ratio/percentage indicates a higher amount of liquid assets available. Therefore, it is predicted that this sign will be negative, decreasing the probability of default.

Pro Forma Total Debt Coverage – Total debt coverage ratio is the organization's measure for a customer's capacity; cash flow available after adding back interest, amortization, and deprecation, subtracting taxes and family living, and adding/subtracting any other items, such as non-recurring income. Total debt coverage is calculated by taking total cash flow available divided by annual principal and interest payments, plus any lease payments. The underwriting standard for this ratio is greater than or equal to 1.25:1 or 125%. A higher ratio/percentage indicates a higher amount of cash flow available than obligated debts to repay. Therefore, it is predicted that this sign will be negative, decreasing the probability of default.

Three Year Average Total Debt Coverage – This ratio is calculated the same as pro forma total debt coverage, except using the previous three years of data. The three-year average is calculated by taking the average of the most recent three years cash flow available divided by the average of the most recent three years annual principal and interest payments, plus any lease payments. The underwriting standard is the same as pro forma total debt coverage and it is predicted that this sign will also be negative, decreasing the probability of default.

Credit Score – Credit score is generated by Factual Data, which scores customers from the three primary credit reporting agencies: Equifax, Experian, and Transunion. The organization uses the score that falls in the middle of the three scores given, it is a whole, three-digit number. A higher credit score indicates the customer has demonstrated adequate repayment history for past or current debts. Therefore, it is predicted that this sign will be negative, decreasing the probability of default as credit score increases.

Featherstone, Wilson, Kastens, and Jones study used the Fair Isaac credit bureau score as an independent variable. This study was based on survey results from agricultural lenders in Kansas and Indiana, regarding factors considered by those financial institutions when lending to farmer borrowers, given four hypothetical agricultural loan requests. Results concluded that the Fair Isaac credit bureau score was statistically significant for influencing both loan approvals and interest rates for borrowers (Featherstone, et al. 2007). Limited research regarding credit scores indicates that they have been included few times when studying credit risk or probability models, primarily due to the data not reported at origination.

Original Principal – The original principal is the amount borrowed by the customer, which is a whole number ranging from five to eight digits. A higher loan amount indicates more debt a customer is obligated to make payments on. Therefore, it is predicted that this sign will be positive, increasing the probability of default. Commitment amount from the Featherstone, Roessler, and Barry study, the same variable definition as original principal for this study, was not statistically significant in predicting the probability of default. It may be ambiguous due to the quantitative difference of the data studied (Featherstone, Roessler and Barry 2006).

Original Loan to Value – Original loan to value measures the amount of financing used to buy or refinance farming assets and how much equity is available post-close. This ratio is calculated by dividing the original principal by the collateral value securing the loan. The underwriting standard for this ratio is less than or equal to 70%. A higher ratio/percentage indicates a lower equity position in assets. Therefore, it is predicted that this sign will be positive, increasing the probability of default.

Crops, Livestock, Permanent Planting, and Other – These variables determine the primary commodity for a customer's operation. Just like the dependent variable, they are binary variables equal to only 0 or 1. If a customer's primary commodity is crops, 1, otherwise, 0; livestock, 1, otherwise, 0; permanent planting, 1, otherwise, 0; or other, 1, otherwise, 0. For crops, it is predicted that the sign will be negative when the binary variable is equal to 1, decreasing the probability of default for crop operation. This is predicted due to opportunities these customers have to contract prices and historical portfolio performance. For livestock, it is predicted that the sign will be positive when the binary variable is equal to 1, increasing the probability of default for livestock operations. This is predicted due to market price volatility, including feed costs, and minimal future contracting customers do. For permanent planting, it is predicted that the sign will be positive when the binary variable is equal to 1, increasing the probability of default for permanent planting operations. This is predicted due to water rights and the number of years it takes to establish production. For other, it is predicted that the sign will be positive when the binary variable is equal to 1, increasing the probability of default for other operations. This is predicted due to anomaly operations typically financed by the organization.

Table 3.2 Estimated Coefficient Signs

Coefficient	Independent Variables	Predicted Sign
β_1	Pro Forma Debt/Asset	+
β_2	Pro Forma Current Ratio	-
β_3	Pro Forma Total Debt Coverage	-
β_4	Three-Year Average Total Debt Coverage	-
β_5	Credit Score	-
β_6	Original Principal	+
β_7	Loan to Value	+
β_8	Crops	-
β_9	Livestock	+
β_{10}	Permanent Planting	+
β_{11}	Other	+

CHAPTER IV: DATA

4.1 Data Collection

The data originated from AgriFinancial's loan origination and servicing portfolio, obtained from the organization's loan accounting software. Ten years of information was requested for traditionally underwritten loans that originated between January 1st, 2012, and December 31st, 2021. Original data included 4,171 loans.

Data were for each loan transaction as of the time of origination. While the organization may update different variables of the loan information from origination to payoff, the origination data input remains the same. Loans that were paid in full and considered inactive were retained for the analysis.

4.2 Variable Selection

Data included a wide array of originating variables including: loan number, customer name, origination date, commodity, original principal, maturity date, payment frequency, county, state, pro forma debt/asset, pro forma current ratio, pro forma total debt coverage, three year average total debt coverage, credit score, appraised value, loan to value, note rate, amortization, number of months active, status, paid in full date, 30 days delinquent, 60 days delinquent, 90 days delinquent, foreclosure, and foreclosure date.

The Basel Accord suggests eight objective criteria for financial organizations to measure when implementing their risk rating systems. These criteria include measurements of evaluations of a firm's repayment, solvency, earnings, operating leverage, financial efficiency, liquidity, management, and industry standing (Featherstone, Roessler and Barry 2006).

Of the variables requested, focus on the 5 "C"s of credit criteria typically considered when approving a loan were considered: pro forma debt/asset, pro forma current ratio, pro forma total debt coverage, three year average total debt coverage, credit score, original principal, loan to value, and commodity. Three of the eight variables considered for the model are criteria the Basel Accord suggests: repayment (pro forma total debt coverage and three-year average total debt coverage), solvency (pro forma debt/asset), and liquidity (pro forma current ratio). Pro forma is identified as projections; the ratio calculations include the proposed financing for each request. For example, a real estate purchase will impact all three ratios due to increasing debts and assets (current portion of principal, long term assets, long term debts) and cash flow (principal and interest payments). While data for 30-, 60- and 90-day delinquencies were available, default is defined as a payment being 90 days or more past due at least once since the origination date for the ten-year time frame studied. Ninety days is a typical definition of default in the market (Featherstone, Roessler and Barry 2006).

Table 4.1 Underwriting Standards and Minimum Ratios

Underwriting Standard	Target Ratio
Debt/Asset	≤ 50%
Pro Forma Current Ratio	≥ 1.25:1
Pro Forma Total Debt Coverage	≥ 1.25:1
Three Year Average Total Debt Coverage	≥ 1.25:1
Credit Score	≥ 680
Loan to Value	≤ 70%

4.3 Data Eliminations and Revisions

After assessing the data, 3,323 loans were used. Eliminations were due to missing data points for individual loans, i.e., a blank or "0" entry, 636 loans, as well as loans from customers that refinanced their previous loan during the 10-year period, 118 loans. To appropriately eliminate these loans, the paid-in-full date and origination date were used. If the two dates were within five days of each other, the assumption was made that the most recent originated loan refinanced the previously originated loan. This allowed for the most recent data available for those customers. Final eliminations were for outliers, from what appeared to be input errors or ratios well above the minimums. To appropriately eliminate these outliers, three times the standard deviation calculations were completed for debt/asset, current ratio, three-year average TDC, and pro forma TDC. A total of 94 loans were eliminated for these, 17 of which had multiple outliers. These variables were eliminated in lieu of adjusting.

Table 4.2 Standard Deviation Adjustments to Data

Variable	Mean	Std. Dev.	Mean + 3x Std. Dev.
Pro Forma Debt/Asset	28.07%	11.88%	64.33%
Pro Forma Current Ratio	6.50	16.54	56.11
Pro Forma Total Debt Coverage	2.04	1.87	7.66
Three-Year Average Total Debt Coverage	2.28	2.54	9.90

The commodity data listed 81 different commodities, several recurring but also several that only occurred for one loan. This list was condensed into four commodity categories: permanent planting, crops, livestock, and other. Table A.1 in the Appendix lists the commodities under each category.

CHAPTER V: RESULTS

Results from the estimated regression model and data described in the previous chapters are evaluated in this chapter, including if the independent variables are statistically significant in predicting the probability of default for the loans analyzed. The original estimated equation used for the AgriFinancial data determining the probability of default risk is as follows:

$$\ln(P_i/[1-P_i]) = \beta_0 + \beta_1{*}ProformaDA + \beta_2{*}ProformaCR + \beta_3{*}ProformaTDC + \beta_4{*}ThreeYrAvgTDC + \beta_5{*}CreditScore + \beta_6{*}OriginalPrincipal + \beta_7{*}OriginalLTV + \beta_8{*}Crops + \beta_9{*}Livestock + \beta_{10}{*}PermanentPlanting + \beta_{11}{*}Other$$

5.1 Original Model Revisions

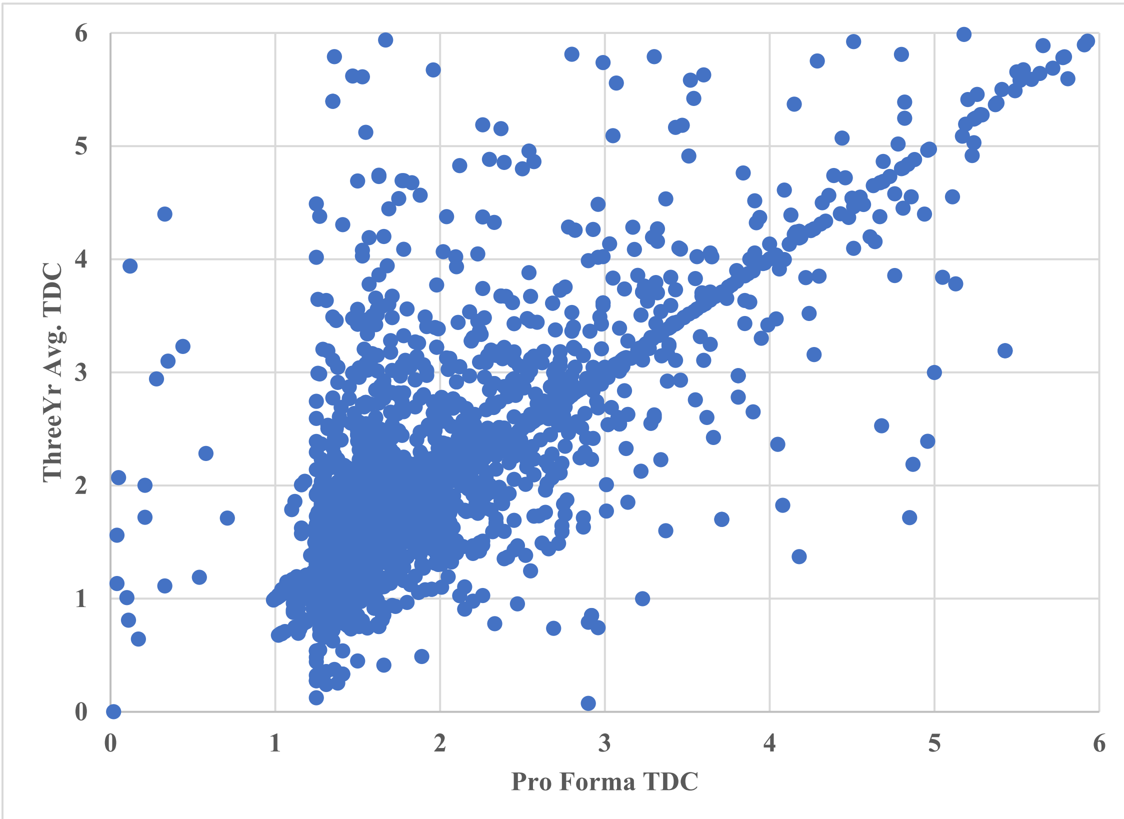
Table 5.1 Correlation Results

	ProformaDA	ProformaCR	ProformaTDC	ThreeYrAvgTDC	CreditScore	OriginalPrincipal	OriginalLTV
ProformaDA	1.00						
ProformaCR	-0.32	1.00					
ProformaTDC	-0.30	0.27	1.00				
ThreeYrAvgTDC	-0.26	0.25	0.75	1.00			
CreditScore	-0.17	0.09	0.11	0.09	1.00		
OriginalPrincipal	0.13	0.02	-0.10	-0.07	-0.11	1.00	
OriginalLTV	0.01	-0.01	-0.01	-0.02	-0.01	0.05	1.00

After estimating the model with these proposed independent variables, changes were made to the original estimated equation. A correlation matrix was completed between all the variables that were nonbinary to test for multicollinearity. Table 5.1 reports these results. It was suspected multicollinearity was present between two variables, ProFormaTDC and ThreeYrAvgTDC. Both ratios calculate capacity in the same manner, pro forma for the future of the operation and three-year average from historical. Results reported a correlation of 0.75, indicating high correlation. All other nonbinary variables reported low correlation between them. A correlation scatter plot was then created to further evaluate multicollinearity between the two variables. Figure 5.1 shows the

ThreeYearAvgTDC versus ProFormaTDC; the shape of the scatter plot indicates multicollinearity.

Figure 5.1 Correlation Scatter Plot for Total Debt Coverage Variables



The model was estimated twice, once without ProformaTDC as a variable and once without ThreeYrAvgTDC as a variable. Results indicated that ThreeYrAvgTDC was statistically significant, and ProformaTDC was not. Therefore, the final model does not include ProformaTDC as an independent variable. Once ProformaTDC was removed, evaluation of OriginalLTV and ProformaCR began. Neither indicated statistical significance. Since loan to value is not one of the suggested Basel Accord measures and more predominate in other risk assessment calculations, such as Loss Given Default, the final model does not include OriginalLTV. Lastly, the ProformaCR variable was revisited by dividing the data by 1 to normalize the variance of numbers. Results from this did not change the statistical significance. The revised estimated equation is as follows:

$$\ln(P_i/[1-P_i]) = \beta_0 + \beta_1{*}ProformaDA + \beta_2{*}1/ProformaCR + \beta_3{*}ThreeYrAvgTDC + \beta_4{*}CreditScore + \beta_5{*}OriginalPrincipal + \beta_6{*}Crops + \beta_7{*}Livestock + \beta_8{*}PermanentPlanting + \beta_9{*}Other$$

5.2 Model Results

Table 5.2 Probability of Default Regression Results

	Coefficient	Std. Error	z	p-value	
Constant	6.24073	1.95337	3.195	0.0014	***
ProformaDA	2.91367	1.31299	2.219	0.0265	**
1/ProformaCR	-0.257676	0.467194	-0.5515	0.5813	
ThreeYrAvgTDC	-0.366967	0.198124	-1.852	0.0640	*
CreditScore	-0.0141816	0.00243415	-5.826	<0.0001	***
OriginalPrincipal	1.40391E-07	6.62E-08	2.122	0.0338	**
Crops	-0.160966	0.741892	-0.217	0.8282	
Livestock	-0.547821	0.856611	-0.6395	0.5225	
PermanentPlanting	-0.674015	0.933324	-0.7222	0.4702	
McFadden R-squared	0.103203				
Adjusted R-squared	0.071073				
Likelihood ratio test: Chi-square(8)	57.817 [0.0000]				

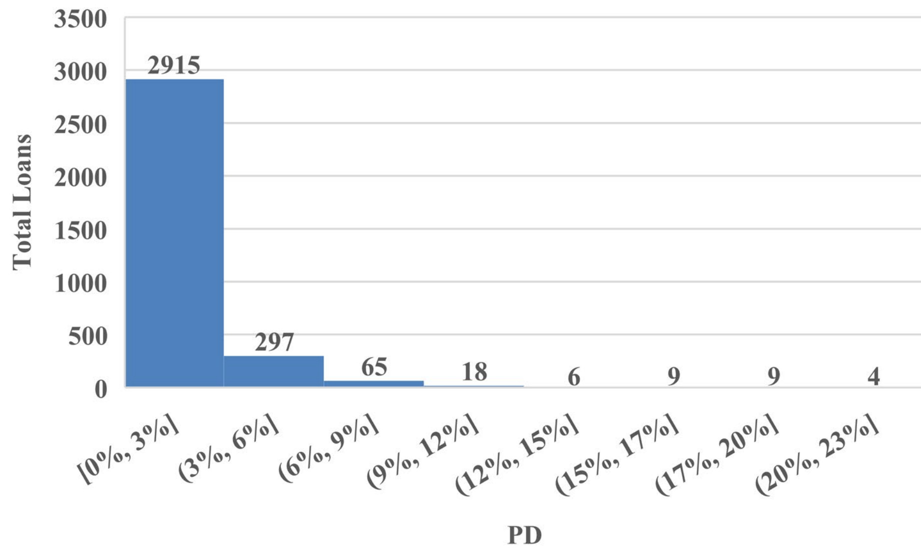
Note: Single, double and triple asterisks (*) denote statistical significance at the 10%, 5%, and 1% levels, respectively.

The regression results are outlined above in table 5.1. Of the variables tested, ProformaD/A, ThreeYrAvgTDC, CreditScore, and OriginalPrincipal resulted in statistical significance for predicting the probability of default. CreditScore is at the 1% significance level, OriginalPrincipal and ProformaDA at the 5% significance level, and ThreeYrAvgTDC at the 10% significance level. The 1/ProformaCR, Crops, Livestock, and Permanent Plantings variables were not statistically significant. The independent variable, Other, was omitted due to exact collinearity, as three other commodity independent variables were binary variables. Signs for all but two of the variables were correctly predicted; Livestock and PermanentPlanting were negative in lieu of the positive predictive, meaning probability of default decreases for these commodities relative to other. Nearly all the coefficients are small, except for ProformaDA. The chi-square value for the likelihood ratio predicts that the model has statistically significant explanatory power in predicting the probability that a loan will default.

5.3 Changes in Probability of Default for Independent Variables

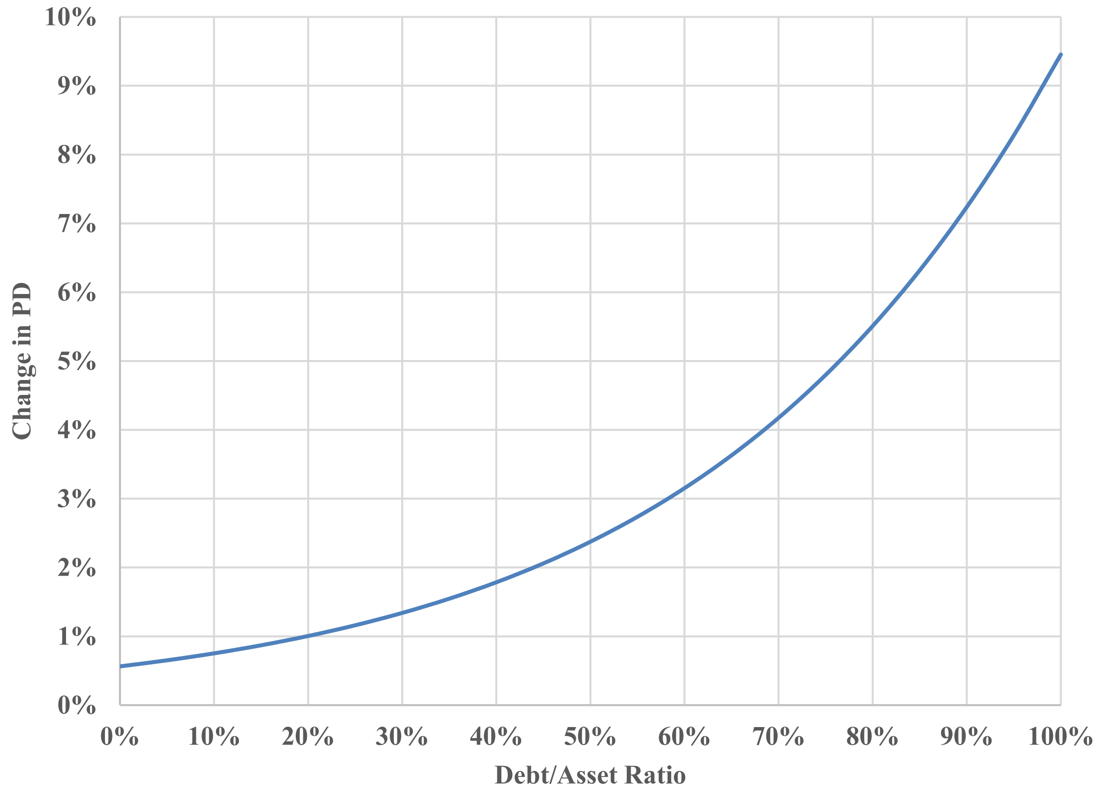
Reviewing the equation for the estimated binary logit model, the log of the dependent variable is calculated. The next step is to calculate the antilog of the dependent variable to determine the impact of a change in one of the independent variables has on the dependent variable, holding all the other independent variables constant at their means. To complete this step, the coefficient results from the model, outlined in Table 5.1, were used to calculate the probability of default for all 3,323 loans. A higher percentage indicates a higher probability of default. Results from the calculations are outlined in Figure 5.2; 87.7% of the loans had a PD between 0-3%. 8.9% between 3-6%, and 3.4% higher than a PD of 6%. The estimated value at risk for the loans with a PD between 0-3% is $25,776,650, $17,098,598 between 3-6%, and $32,160,218 higher than a PD of 6%. Total value at risk is estimated at 2.7% of the total data loan volume.

Figure 5.2 Number of Loans by Probability of Default



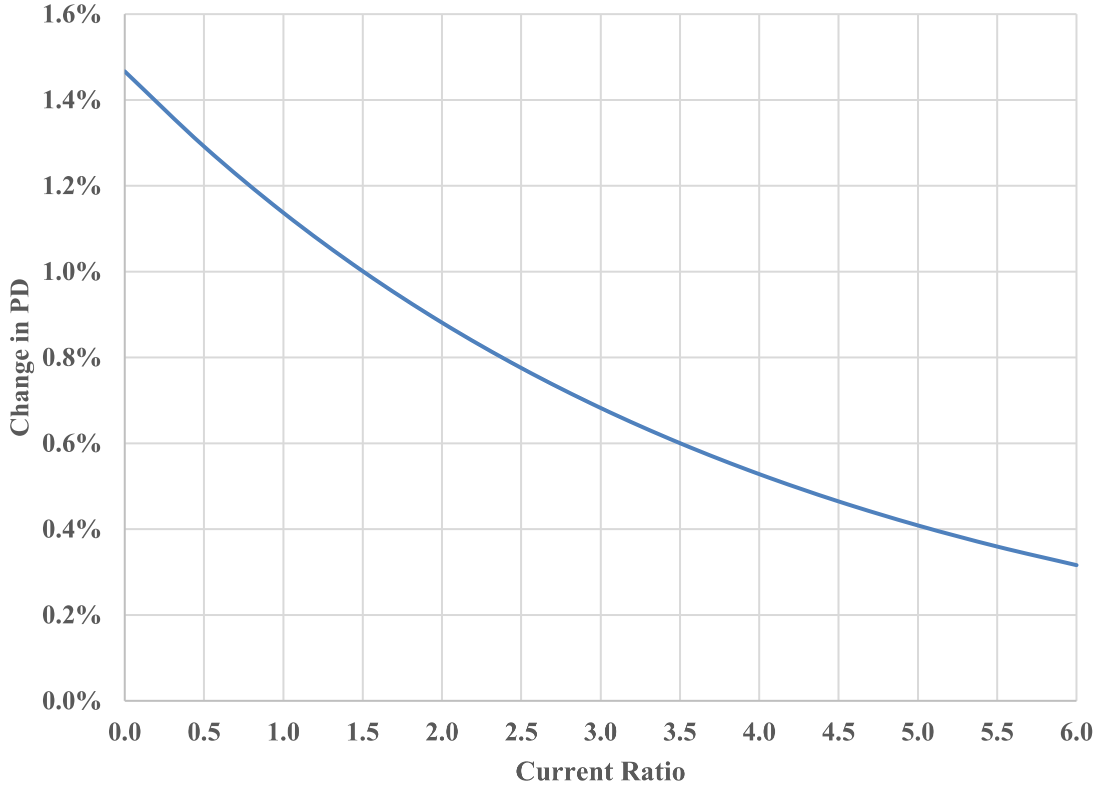
The ProformaDA coefficient had a statistically significant positive sign, indicating that as a customer's debt/asset ratio increases, the probability of default increases, holding all other independent variables at their mean. Figure 5.3 illustrates the change in PD for a change in the debt/asset ratio. Of the 55 loans that defaulted, only one had a ratio below underwriting standards of 50%. Due to this variable's statistical significance at the 5% level and increasing rate of change in PD, it is an important measure of risk.

Figure 5.3 Probability of Default Changes as Debt/Asset Increases



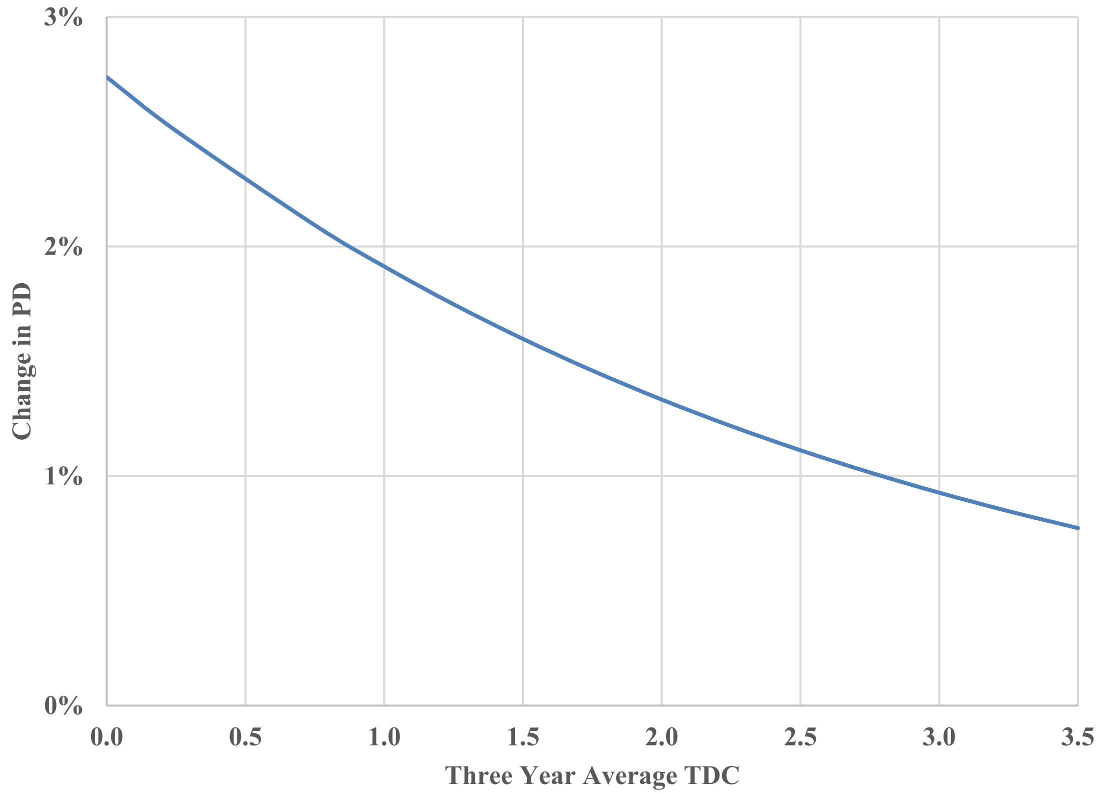
The 1/ProformaCR coefficient had the expected negative sign, indicating that as a customer's current ratio increases, the probability of default decreases, holding all other independent variables at their mean. Figure 5.4 demonstrates the change in PD for a change in the current ratio. The change is minimal; current ratio increasing from 0.50:1 to 6.00:1 only results in nearly a 1% decrease in PD. The current ratio was graphed in lieu of 1/current ratio. Model results indicate this variable was not statistically significant in determining PD. Due to these results, this variable is not a large measure of default.

Figure 5.4 Probability of Default Changes as Current Ratio Increases



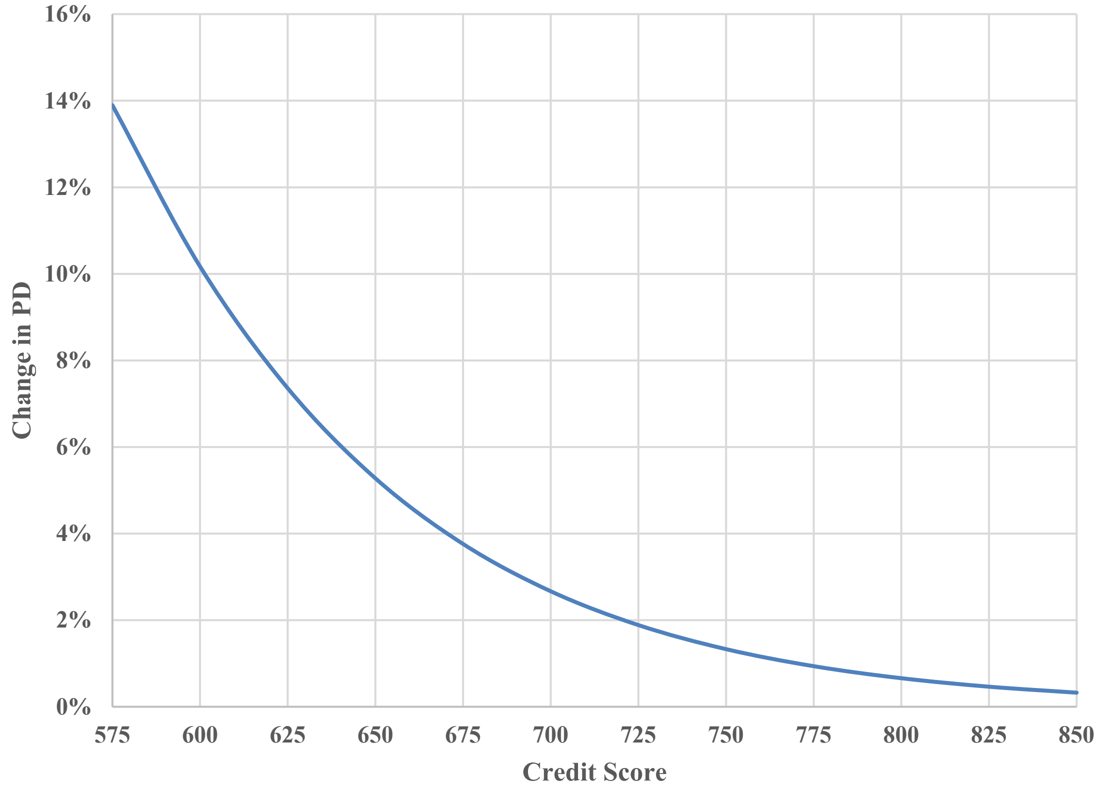
The ThreeYrAvgTDC coefficient was as predicted with a negative sign indicating that as a customer's total debt coverage ratio increases, the probability of default decreases, holding all other independent variables at their mean. Figure 5.5 demonstrates the change in PD for a change in total debt coverage ratio. The PD decreases at a decreasing rate. Of the 55 loans that defaulted, 12 had a ratio below underwriting standards of 1.25:1. Due to this variable's statistical significance at the 10% level, it is important for determining default risk.

Figure 5.5 Probability of Default Changes as Three Year Average Total Debt Coverage Increases



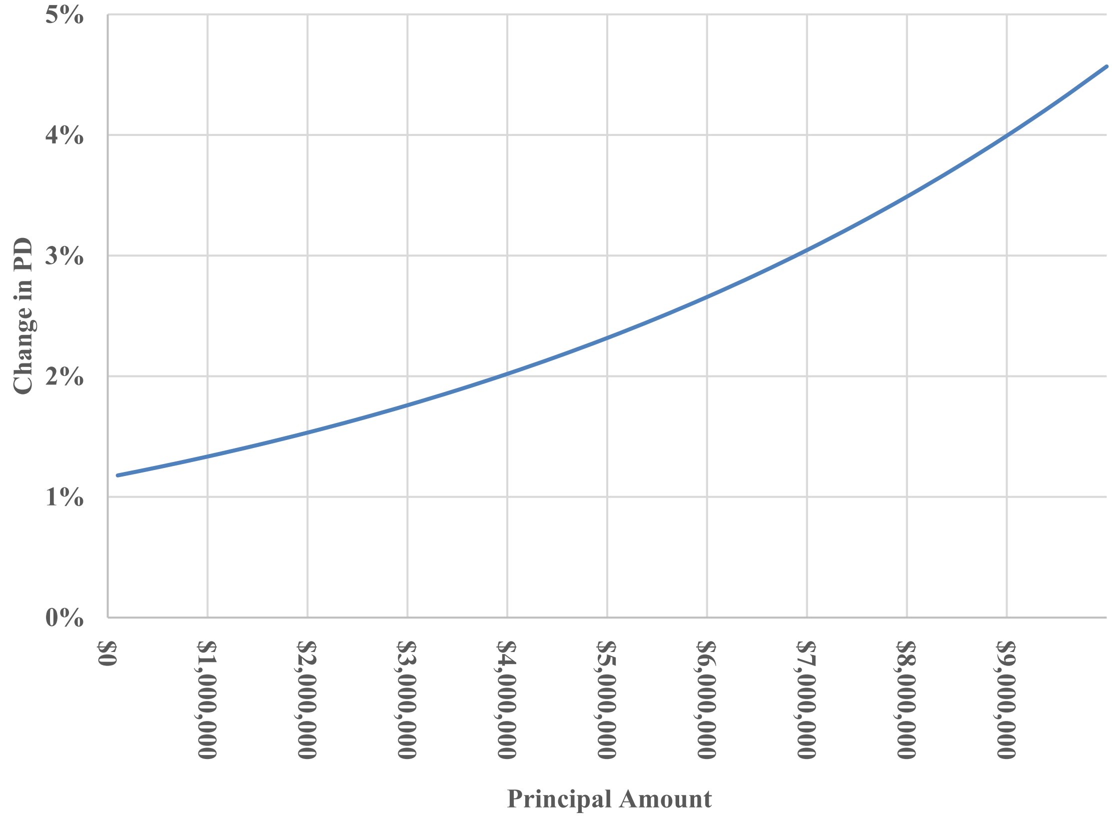
The CreditScore coefficient was as predicted with a negative sign indicating that as a customer's credit score increases, the probability of default decreases, holding all other independent variables at their mean. Figure 5.6 demonstrates the change in PD for a change in credit score. The PD decreases at a decreasing rate. This variable had the highest statistical significance at the 1% level. As the chart outlines, this variable resulted in the greatest change in PD. Of the 55 loans that defaulted, 18 had a credit score below underwriting standards of 680. Due to this variable's statistical significance at the 1% level, it should be highly considered for determining default risk.

Figure 5.6 Probability of Default Changes as Credit Score Increases



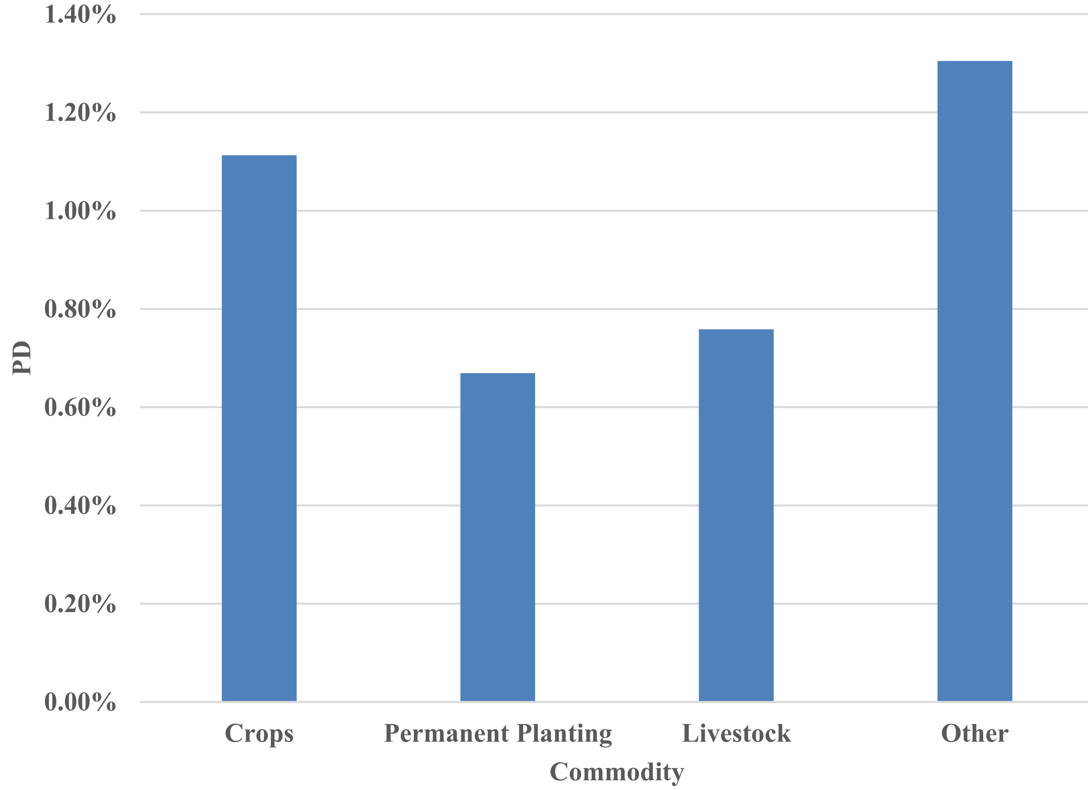
The OriginalPrincipal coefficient was correctly predicted as positive, indicating that as a loan amount increases, the probability of default increases, holding all other independent variables at their mean. Figure 5.7 demonstrates the change in PD for a change in the loan/principal amount originated. The PD increases at an increasing rate. Of the 55 loans that defaulted, the average loan amount of these loans at the time of origination was $1,429,646; 23 had a loan amount higher than the average. Due to this variable's statistical significance at the 5% level and increasing rate of change in PD, it should be considered for determining default risk.

Figure 5.7 Probability of Default Changes as Original Principal Increases



The commodity coefficients were not all predicted as expected with the Other variable omitted due to exact collinearity. Signs for all but two of the variables were correctly predicted; Livestock and PermanentPlanting were negative compared to Other in lieu of the positive predictive, meaning probability of default decreases for these commodities, just as it does for Crops, compared to the Other loan type. Figure 5.8 demonstrates the PD for when each respective commodity is "1" and the other three are "0". Model results indicated none of these variables were statistically significant in determining PD. Due to these results, these variables may need to be reconsidered for determining risk. The financial information is more important than loan type for predicting default.

Figure 5.8 Probability of Default for Commodity Variables



5.3 Summary

Revisions were made to the originally estimated model of eleven independent variables. Two of the variables, ProFormaTDC and OriginalLTV were removed. ProFormaTDC and ThreeYrAvgTDC were evaluated for multicollinearity; ThreeYrAvgTDC was kept in the model due to statistical significance. OriginalLTV was removed due to lack of statistical significance and not being one of the predominate risk assessment measures. Models were estimated using Gretl, to determine if the remaining variables had a statistical significance on the dependent variable. Overall, four of the nine variables were statistically significant at the 1%, 5% and 10% levels.

The Featherstone, Roessler, and Barry study also used a regression model to predict the probability of default. Results from the study were over seven years, broken down into different loan products, including real estate loans. Real estate loans were the most comparable to this study. Independent variables Featherstone, Roessler, and Barry studied include CDRC (capacity), OE (solvency), and WC (liquidity), comparable to ThreeYrAvgTDC, ProformaDA, and 1/ProformaCR, respectively. Default was defined the same, 90 days. Sign results were the same for capacity and liquidity measures in both studies. Solvency had opposite signs as OE and Debt/Asset are different aspects of solvency, calculated as 1 – debt/asset. As previously outlined, solvency was statistically significant at the 5% level, capacity at the 10% level, and liquidity did not report significance. The Featherstone, Roessler, and Barry study reported solvency was statistically significant at the 1% level, capacity was not significant, and liquidity at the 5% level. The graphs for each variable were similar from a quantitative perspective, this study showed a slightly large variance for the change in PD. Capacity and liquidity both decreasing at a decreasing rate and solvency decreased at a decreasing rate compared to increasing at an increasing rate given the different calculation. Overall, the signs were more similar than statistical significance. Commitment amount from the Featherstone, Roessler, and Barry study, the same variable definition as OriginalPrincipal for this study, was not statistically significant in predicting the probability of default. However, this variable was studied in a model with all loan products in lieu of the individual product models. It may be ambiguous due to the quantitative difference for data study, 157,853 loans compared to 3,323. (Featherstone, Roessler and Barry, Determining the Probability of Default and Risk-Rating Class for Loans in the Seventh Farm Credit District Portfolio 2006)

CHAPTER VI: CONCLUSION AND FUTURE ANALYSIS

It was concluded that four of the nine variables were statistically significant in determining the probability that a loan will default from the time of origination to pay off. Originally, eleven independent variables were considered; however, two were removed from the model due to multicollinearity and a ratio more predominate in other risk assessment calculations, such as Loss Given Default. All the signs for the independent variables were as predicted, except for two of the commodity variables.

Of the nine variables, four are credit measures used at origination to determine a loan decision. The results indicated that three of these four variables were statistically significant, including credit score, pro forma debt/asset ratio, and three-year average total debt coverage at the 1%, 5%, and 10% levels, respectively. The fourth variable that was not significant was pro forma current ratio. The original principal, size of the loan at origination but not a key credit measure, was also statistically significant at the 5% level.

Overall, the chi-square value for the likelihood ratio predicted that the model has explanatory power in predicting the probability that a loan will default. Credit score had the highest statistical significance for determining probability of default. The Featherstone, Wilson, Kastens, and Jones study, concluded that the Fair Isaac credit bureau score was statistically significant for influencing both loan approvals and interest rates for borrowers. Both studies indicate that credit score should be highly considered when determining default risk (Featherstone, et al. 2007).

6.1 Future Analysis

Data from AgriFinancial's servicing software had additional information not used for this thesis that could be addressed in further modeling that may better predict delinquency. Specifically, loan product, payment frequency, and collateral state.

Loan products include revolving lines of credit, fixed rate products, and variable rate products. By separating out loans by product, probability of default could be modeled by specific product. This would result in three PD models. Leading thought behind the separation is that revolving line of credit products have a different payment structure of interest only payments compared to term loans with structured principal and interest payments. Payment frequency may only be relevant to fixed and variable rate products given the outlined payment structures. Typically, lenders comment monthly payments are preferred, reducing principal each month compared to once a year. Would payment frequency be statistically significant in determining PD and confirm this idea? Given commodities were not statistically significant in the model, states could be divided into regions to determine if PD statistical significance is relevant by geographical areas.

Data not currently collected by the organization that may be considered is income diversity, such as off-farm sources of income and multiple farm commodities. Are operations with off-farm income sources statistically significant and do they increase or decrease PD? The same question can be asked regarding multiple commodities, for example, row crop and livestock operations versus only row crop.

6.2 Recommendations

Probability of default models give organizations an idea of the types of risks they are funding through loans or other financing products, and whether the risks may be appropriate. If the risk is greater, would higher interest rate pricing help mitigate the risk? If there is not enough risk, are applications being declined that should be reconsidered? Results from the thesis' estimated regression model indicate that four independent variables were statistically significant. One recommendation, based on these results, would be to consider an additional interest rate margin for loans that may have a credit ratio that falls outside of underwriting standards or a higher loan amount to account for the additional risk. This could be determined based on the PD of each loan. For example, if over a certain percentage, adding an additional 10 basis points to the interest rate margin.

Furthermore, these risks can be evaluated using a risk rating scale. Now that a probability of default model has been established, PD ratings now can be classified into risk rating categories. As of present, the organization has received a risk rating scale from one of their funding participants that is being considered. There are also companies, such as Standards & Poor's (S&P) and Moody's Investors Services, who provide a basis of rating scales. A decision regarding what risk rating scale to base the PD model on for the future needs to be made. It is recommended that once this is complete, the organization discusses how risk adverse to be on the scale for newly originated loans.

Another recommendation to further evaluate the risks is determining loss given default, LGD, to strengthen the organization's knowledge of future potential losses for newly originated loans (Tauer and Dressler 2016). PD calculations are necessary to further calculate LGD. The data for this thesis included other information pertinent to determine LGD, such as date of defaults, loan-to-value, and original collateral value. If the organization pursues opportunities to fund loans in lieu of participations currently used, LGD will help determine how much capital to set aside for potential losses that may occur.

Lastly, it is recommended the model is revisited on an annual basis, to ensure validity, accuracy, and meeting expectations for the organization's goals it set out for the credit risk rating model.

6.3 Conclusions

The organization currently funds loans through participations from other lending institutions. This thesis project was prompted to begin risk assessments from our current originated portfolio to determine future risk of current and future loan programs, specifically those that the organization may fund in lieu of participation. The probability of default is the basis to begin assessing the risk, followed by assigning risk ratings, then loss given default. Once these three are put into place, the organization can develop a level of comfort with the risk assessments by following these practices for currently originated loans, further welcoming the idea of internal funding. Internal funding may position the company to have competitive advantage such as quicker turnarounds, (using PD to support decisions in lieu of lengthy analysis when appropriate), scorecard type programs, and pricing based on overall risk per each request.

WORKS CITED

Barry, Peter J. 2001. "Modern Capital Management by Financial Institutions: Implications for Agricultural Lenders." *Agricultural Finance Review* 103-122.

Barry, Peter J., and Ani L. Katchova. 2005. "Credit Risk Models and Agricultural Lending." *American Journal of Agricultural Economics* 194-205.

Brewer, Brady E., Jason S. Bergtold, Allen M. Featherstone, and Christine A. Wilson. 2019. "Farmers' Choice of Credit among the Farm Credit System, Commercial Banks, and Nontraditional Lenders." *Journal of Agricultural and Resource Economics* 362-379.

Featherstone, Allen M., Christine A. Wilson, Terry L. Kastens, and John D. Jones. 2007. "Factors Affecting the Agricultural Loan Decision-Making Process." *Agricultural Finance Review* 13-33.

Featherstone, Allen M., Laura M. Roessler, and Peter J. Barry. 2006. "Determining the Probability of Default and Risk-Rating Class for Loans in the Seventh Farm Credit District Portfolio." *Review of Agricultural Economics* 4-23.

n.d. *History of the Basel Committee.* Accessed 08 10, 2023. https://www.bis.org/bcbs/history.htm.

Katchova, Ani L., and Peter J. Barry. 2013. "Credit Risk Models: An Application to Agricultural Lending." *Agricultural Finance Markets in Transition.* Kansas City. 7-27.

Quaye, Frederick M., Denis Nadolnyak, and Valentina Hartarska. 2017. "Factors Affecting Farm Loan Delinquency in the Southeast." *Research in Applied Economics* 75-92.

Studenmund, A.H. 2016. *Using Econometrics A Practical Guide.* Boston: Pearson.

Tauer, Loren W., and Jonathan B. Dressler. 2016. "Estimating Expected and Unexpected Losses For Agricultural Mortgage Portfolios." *American Journal of Agricultural Economics* 1470-1485.

Yu, Tom, Tom Garside, and Jim Stoker. 2001. "Effective Credit Risk-Rating Systems." *RMA Journal* 38-43.

Zech, Lyubov, and Glenn Pederson. 2004. "Application of Credit Risk Models to Agricultural Lending." *Agricultural Finance Review* 91-106.

APPENDIX

Table A.1 Commodities by Category

Crops	Livestock	Permanent Planting	Other
Alfalfa	Dairy Cattle and Calves	Almonds	Aquaculture
Barley	Ducks	Apples	Bees
Beans	Farm Chickens	Avocados	Beets
Bluegrass	Feedlot Cattle and Calves	Blackberries	Carrots
Canola Oil	Hogs	Blueberries	CRP
Corn	Horses and Mules	Cherries	Greenhouse/Nursery
Cotton Lint	Milk	Figs	Hops
Cottonseed	Other Livestock	Filberts	Hunting Recreation
Fescue	Other Poultry	Grapes	Lettuce
Hay	Ranch Cattle and Calves	Grapes-Raisins	Misc. Vegetables
Mint	Sheep and Lambs	Grapes-Table	Nut Processing & Shelling
Oats	Turkeys	Grapes-Wine	Other Processing
Other Field Crops		Lemons	Peas, green
Other Seeds		Limes	Peppers, green
Peanuts		Misc. Fruit and Nuts	Timber
Potatoes		Olives	Tomatoes
Rice		Oranges	Watermelon
Ryegrass		Peaches	
Silage		Pears	
Sorghum		Pistachios	
Soybeans		Plums and Prunes	
Sugar Beets		Pomegranates	
Sugarcane		Raspberries	
Tobacco		Tangerines	
Wheat		Walnuts	