**Vehicle Make as a Determinant of
Automobile Loan Default, Exposure at Default, and Duration Until Default**

**Authors**

Colene Trent, Ph.D.**\***

Union University, ctrent@uu.edu

Andrew Tiger, Ph.D.

Angelo State University, andrew.tiger@angelo.edu

Joel Wyatt, Data Analyst II

Leaders Credit Union, jwyatt@leaderscu.com

Kyle Yeoman

Union University Alumnus, kylepyeoman@gmail.com

**Abstract**

In the automobile loan market, lenders face the risk that borrowers will default. In evaluating potential clients, lenders would benefit from the ability to answer three questions: Who will default? What is the potential loss exposure when the borrower defaults? When will the borrower default? This paper suggests that lenders consider the make of vehicle purchased by the borrower when attempting to answer these questions. We analyze applicant-level and loan-level data from an American credit union, employing logistic, tobit, and censored normal regression. Our results suggest that lenders could more efficiently price for risk by including a borrower’s choice of vehicle in predictive models specific to their market.

JEL Classification: G21

Keywords: automobile loan default, exposure at default (EAD), duration

**\* Contact Author**

**Vehicle Make as a Determinant of
Automobile Loan Default, Exposure at Default, and Duration Until Default**

1. **Introduction**

“You are what you drive” is the popular mantra of American auto manufacturers and car dealers alike (Stokes and Hallett, 1992). Soccer moms might find themselves shouting admonitions towards unruly children in the backseat of a Chrysler minivan. Rugged, hard-working types might choose the reliable Dodge Ram. Adventurous? Seeking status? Concerned about the environment? Practical? Whatever your personality there is seemingly an automobile that reflects your ambitions. If cars reflect the drivers’ identity and serve as an extension of who they are, could lenders potentially use this information to reduce their own risk exposure? This paper investigates whether information on vehicle make is useful in predicting auto loan defaults, the lender’s exposure at default, and the duration until default.

Auto loans are vital to the U.S. consumer, with 86.1% of new and 54.7% of used autos financed through loans (Experian, 2018). Although banks, credit unions, and financing companies are eager to provide this financing, lenders may benefit by becoming increasingly creative in screening potential clients to respond to the latest trends in the auto loan industry. Over the past several years, the auto loan market has featured looser credit standards, with larger loan amounts financed over longer terms. In the second quarter of 2018, outstanding automobile loans rose to reach a record high of $1.124 trillion in the United States (Board of Governors of the Federal Reserve System US, 2018). According to the Federal Reserve Bank of New York (2018), auto loan balances are on a six-year upward trend, with 4.3% of auto loan balances 90 or more days delinquent at the end of the second quarter of 2018. The average new and used car loans were just over $30,950 and $19,700 respectively with the average new and used car monthly payment at $525 and $378 respectively (Experian, 2018). Consumers are increasingly taking on longer loans with the average new loan term at just under 5.75 years and the average used loan term at 5.3 years (Experian, 2018). Lower interest rates have made these longer loans attractive to the consumer, but they may signal an increase in delinquencies in the future. Lenders should closely monitor their portfolios, “as more customers roll existing auto loans into new vehicle purchases” (Federal Reserve Bank of Atlanta, 2017). Regardless of whether these trends continue in the future, lenders should remain vigilant in screening potential clients and make better use of available information in order to remain competitive and more efficiently price for risk when offering loans.

Default occurs when the debtor fails to repay the loan. Lenders undertake credit risk management and engage in risk-based pricing, which involves offering different interest rates, loan terms, or fees based upon creditworthiness of the individual borrower, to keep the bank’s risk of default within acceptable parameters. Typically, information about the borrower is used to qualify loans; information on the loan term, year of vehicle, and credit score are typically used to determine the interest rate on the loan. Credit score usually serves as the primary metric to determine the ease with which an institution should extend credit. To minimize defaults on auto loans, lenders should consider adopting additional risk-based pricing methods, which differ from these traditional methods. This paper proposes the use of information on vehicle make to better adjust for default risk. Agarwal, Ambrose, and Chomsisengphet (2008) note that auto insurers “have long recognized that automobile makes and models appeal to different clienteles and that these clienteles have heterogeneous risk profiles and accident rates”. Furthermore, “information about the underlying asset as well as the borrower’s personal characteristics” is vital to mortgage lenders (Agarwal, et al., 2007). If auto and home insurers and mortgage lenders benefit from information on the specific car and home being insured, it is likely that auto lenders could benefit from incorporating information on vehicle make in their lending decisions.

We propose that lenders consider vehicle make when attempting to answer three risk-related questions: Who will default? What is the potential loss exposure when the borrower defaults? When will the borrower default? In other words, lenders should employ information on a borrower’s choice of vehicle when predicting the *probability* that a borrower will default, the bank’s *exposure at default*, and the *duration until default*. Exposure at default (EAD) is the amount of loss that a bank may face due to default, which is a key input into determining their expected credit losses (Bandyopadhyay, 2016). The duration until default is the amount of time that passes before default occurs.

Our results indicate that vehicle make is a significant predictor of default probability, exposure at default, and duration until default. Our notable results for each of three models are as follows: 1) Logistic regression is used to predict the probability of default. Borrower loan characteristics that traditionally predict default continue to perform as expected. Additionally, we find that joint loans were significantly less likely to end in default. Default probabilities can be more accurately predicted by including vehicle make in the analysis. Specifically, Ford/Lincoln/Mercury vehicles were significantly more likely to default while GMC, Dodge/Chrysler, Nissan/Infiniti, and Mazda were significantly less likely to default relative to the control make of Chevrolet. The default probability is also greatly increased for ATVs and motorbikes. 2) Tobit regression analysis is used to predict exposure at default (EAD). As expected, the variables that were significant predictors of probability of default from the logistic regression are significant in predicting EAD. Notable results are that longer loan terms increase the exposure at default, but exposure is lower for joint applicants. Several makes were significant in predicting exposure at default with ATVs and motorbikes and Ford/Lincoln/Mercury vehicles leading to a greater exposure at default relative to the control variable of Chevrolet. 3) Censored normal regression analysis is used to predict the duration to default. GMC, Dodge/Chrysler, and Mazda vehicles had a significantly longer duration while Ford/Lincoln/Mercury and ATV/Motorbike vehicles had a significantly shorter duration, relative to the control make of Chevrolet. Overall, these findings suggest that lenders could more efficiently price for risk by including a borrower’s choice of vehicle in predictive models specific to their market.

1. **Related Literature**

Our paper seeks to add to the existing research on default behavior by suggesting that lenders use information on vehicle make to estimate the probability of default, the exposure at default, and the timing of default. We model how this information could be incorporated to prevent losses. Regarding the progress made in predicting individual credit risk, Carrol and Zeltkevic (2007) note that “the biggest enhancement to this process in recent years has been the use of quantitative models of default risk, which have largely supplanted subjective assessments of “ability and willingness to repay” when making loan-level credit risk decisions”. As access to information and data analytics becomes widespread, even small-scale financial institutions and lenders have begun to seek a “mechanism to identify the potential defaulters, i.e., a predicting model” (Bhadwaj and Bhattacharjee, 2010). Although using risk models based upon industry standards are popular, Carrol and Zeltkevic promote the increased use of models specific to a particular firm that seek to incorporate that firm’s own client experiences (2007). This recommendation is echoed in this paper as we seek to predict default probabilities, exposure at default, and duration until default for a single credit union.

The literature provides many studies that seek to model individual risk behaviors particularly related to auto loans. Heitfeild and Sabarwal (2003) utilize a competing risks model of default and prepayment on pools of subprime automobile loans. They note that increases in unemployment precede increases in default rates and lenders charging higher interest rates experience higher default rates. Agarwal, et al., (2007, 2008) use a competing risks framework to analyze auto loans. Credit score is a significant predictor of default, and an increase in the loan to value ratio and an increase in local area unemployment increases the probability of default. Agarwal, et al. (2008) also provide a helpful analysis of automobile make and its relationship with default rates. They find that used cars have a higher likelihood of default, while economy cars have a lower probability of default, noting that individual automobile-make dummies are significant drivers of default. They also find that significant dispersion exists in the performance of auto loans after controlling for manufacturer country of origin, specifically noting that loans on U.S. cars have significantly higher default rates (Agarwal, et al., 2007). Adams, Einav, and Levin (2009) analyze credit market conditions for subprime borrowers, finding that loan size positively predicts the probability of default. Bhadwaj and Bhattacharjee (2010) use data from India to find that personality traits such as money attitudes, power-prestige, and anxiety are significant predictors of default on auto loans. Einav, Jenkins, and Levin, (2013) details the adoption of credit scoring as a method for auto lenders to increase profitability and to offer better customization of contract terms to lower-risk borrowers. Tanninen (2013) examines Finnish auto loan data, finding that the loan-to-value ratio and the credit score are important indicators of default. Hock-eam and Yeok (2017) examine vehicle loan default in Malaysia and determine that loan-related characteristics, such as areas of residence, vehicle purchase price, length of service, existing relationship with the bank, interest rate, and the presence of a co-signer, are the most important determinants of probability of default, compared to socio-demographic, financial ability, and vehicle related characteristics. Our analysis of duration until default is similar to that of Banasik, Crook, and Thomas (1999) who apply a proportional hazards model to a sample of personal loans to determine when borrowers will default. Dirick, Claeskens, and Baesens (2017) also provide a helpful review of survival and duration analysis techniques applied to time to default studies.

Our paper has three distinct goals. First, by analyzing default probabilities using applicant-level and loan-level data on indirect automobile loans from a credit union in the United States, we seek to determine whether the make of vehicle purchased by a borrower reveals information about the *probability of default*. Logistic regression is employed for this analysis. Second, we seek to add to the literature by determining which factors influence the *amount* of the default or the predicted amount of loss the creditor is exposed to when the borrower defaults on the loan. This is otherwise known to financial institutions as their exposure at default. Tobit regression is used for this analysis. Third, our paper seeks to determine the factors which influence the *time* to default using duration analysis. In other words, if a borrower defaults, when will the default occur? A censored normal regression is employed for this analysis. Our work is the first to employ a dataset of indirect auto loans from a credit union to analyze the impact of vehicle make on default probabilities. This paper is innovative in that it adds to the contributions of Agarwal, et al., (2007, 2008) by being the first to use information on automobile make to predict exposure at default and duration until default.

## Methodology

This paper attempts to predict 1) default probability, 2) exposure at default, and 3) duration until default. First, we investigate whether a borrower’s choice of vehicle is related to the probability of default using logistic regression. The dependent variable, *default*, is a binary variable with 1 indicating that an individual defaulted on the loan and 0 indicating that they did not default. The logistic regression model (Agresti, 1996) takes the form

 (1)

where 𝑝 is the probability of default, is a series of dummy variables representing vehicle makes, and is a series of borrower and loan characteristic variables expected to influence the probability of default.

 Second, this paper seeks to determine whether borrower characteristics and information on the borrower’s choice of vehicle can be used to predict the lender’s exposure at default (EAD). Following Wooldridge (2002), the tobit structural equation is:

, (2)

where includes borrower and loan characteristic variables as well as dummy variables representing various vehicle makes. The latent dependent variable, *exposure at default (),* indicates the dollar value of the loan which is left unpaid after the default. Because the exposure at default is zero for a non-trivial fraction of the population, a corner solution response is modeled. *EAD\** is a latent variable that is observed for values greater than zero and censored otherwise. The observed *EAD* is defined by the following equation:

 (3)

The likelihood and log-likelihood functions are

 (4)

 (5)

where (⋅) is the cumulative distribution function of a standard normal distribution, and (⋅) is the corresponding density function. Since the coefficients from the tobit model indicate the partial effects of the independent variables on *E(\*|x)* where EAD\* is a latent variable, other expected values are of interest: the unconditional expected value, the conditional expected value, and the probability of being greater than the lower bound of zero. Following Wooldridge (2002), these expected values are defined as follows:

Unconditional expected value:

 (6)

Conditional expected value: (7)

Probability greater than lower bound: (8)

where (⋅) is the cumulative distribution function of a standard normal distribution and is the inverse Mills ratio. These expected values provide marginal effects at the means of the independent variables. The corresponding marginal effects are

 (9)

} (10)

 (11)

where (⋅) is the cumulative distribution function of a standard normal distribution, (⋅) is the corresponding density function, and is the inverse Mills ratio. Equation 9 indicates how the observed variable EAD changes with respect to the regressors. Equation 10 indicates how the observed variable EAD changes with respect to the regressors for the subpopulation of uncensored observations (loans which default and thus have a positive exposure at default). Equation 11 indicates how the probability of having a positive exposure at default changes with the regressors.

Third, the relationship between borrower characteristics, vehicle characteristics, and the time to default is investigated using a censored normal regression model. Lenders would benefit from knowing what factors influence the length of time for which borrowers honor their obligations prior to defaulting. The censored normal regression model, introduced by Amemiya (1973), is a generalization of the standard Tobit model. The censored normal regression approach is used when the censoring threshold varies across the observations in the sample. The dependent variable in this analysis is *duration* which is the number of months before default occurs. For some borrowers, default may never happen (because these borrowers pay back their loan during their loan term) or it may happen in the event of a longer loan term than is observed. Therefore, *duration* is right-censored, and the censored normal regression model is appropriate. Following Woodridge (2002), the model is specified as:

 (12)

 (13)

where includes borrower and loan characteristic variables as well as dummy variables representing various vehicle makes and satisfies the classical linear model assumptions. Equation (13) implies that rather than observing , we observe it only if it is less than a censoring value, , which is the loan term for each borrower. Under the assumptions in (12) and (13), the density of given and is:

(14)

where (⋅) is the cumulative distribution function of a standard normal distribution, and (⋅) is the corresponding density function. Define the dummy variable

 (15)

The likelihood and log-likelihood functions are

 (16)

 (17)

## Data

*4.1 Sources of Data*

The data used in this analysis was provided by a credit union which serves a metropolitan statistical area (population of approximately 130,000 people) in the southeastern United States. The National Credit Union Administration categorizes the credit union as belonging to Peer Group 5 which has total assets of $100 million to less than $500 million (2018). The final dataset includes 3,541 observations on loans issued from January 2008 through December 2014 which have either matured or defaulted as of May 2017. When borrowers fall behind on debt payments, the lender must eventually issue a charge-off. The credit union providing this dataset is required to undertake a charge-off when the auto loan reaches a level of delinquency of 180 days past due. However, the lender may initiate a charge-off prior to that time. The credit union uses debt-to-income and payment-to-income ratios as qualifiers for loans; however, these variables are not used in the determination of loan interest rates. To determine the rate, the credit union places potential borrowers into different tiers or groupings based upon credit scores. The vehicle year, loan to value ratio, and vehicle mileage are then used to adjust the interest rate suggested by these tiers.

All the loans in the dataset are indirect loans; thus the loans were ultimately transferred from a dealer who originated the loan to the lender. Auto lenders must offer competitive rates to attract customers, particularly in the market for indirect loans, which serve as the focus of this paper. With indirect loans, dealerships “shop” for the financial institution that will provide the lowest rate to the customer while offering the dealer the best payment when the loan is sold. These financial institutions give the dealership a fixed interest rate to offer to customers. The challenge with an indirect loan is that lenders compete with each other and sometimes with a captive finance company to offer the best rate. Captive finance companies can often offer a lower rate since their parent company is profiting from the transaction. Increased competition may have a negative influence on indirect auto lending programs of some credit unions with heightened competition “prompting credit unions to offer lower interest rates, lengthen amortization periods, and scale down payment requirements” (Credit Union Department – State of Texas, 2014). In 2018, banks enjoyed the largest market share of auto financing, including both new and used vehicles and loans and leases, with 31.6% market share. Captive finance companies had 29% market share while credit unions were close behind at 21.3% market share (Experian, 2018). Credit unions could increase their market share through more widespread use of predictive models tailored to their local market, such as the one suggested in this paper, to compete while appropriately pricing for the risk at hand.

*4.2 Summary Statistics*

Table 1 presents summary statistics for the variables used in the logistic regression. These variables were measured at the time of the loan origination. The first columns consider the summary statistics for the entire sample. The dependent variable in this analysis is *default* with 9% of the loans in the dataset resulting in default. The variables used in the regression equation include applicant-level personal variables such as the *age of the borrower*, *credit score*, *debt to income ratio*, and *payment to income ratio*. Three categories of age are considered in the analysis: 39% of the sample is under 40, 49% of the sample is between the ages of 40 and 65, and 12% of the sample is over the age of 65. The average credit score for the sample was 723. On average, the debt to income ratio for borrowers was 32 while the payment to income ratio averaged to approximately 8. It is generally recommended that borrowers have a debt to income ratio of less than 40 and a payment to income ratio of less than 15. The analysis also considers variables related to the auto loan including the *amount of the loan*, the *interest rate on the loan*, the *term of the loan*, and the *loan to value ratio*. On average, borrowers received loans of approximately $15,700 at a rate of 8% for approximately 4.5 years. A dummy variable is included to indicate whether this loan is a *joint loan*. Joint applicants to a loan each have an ownership interest in the vehicle and are both equally responsible for ensuring that the loan is repaid. For joint loans, the credit union uses the highest credit score of the two applicants for loan approval. For our analysis, we use the higher credit score for the *credit score* variable. We use the age of the applicant with the highest credit score for the *age* variable. 35% of the sample consists of joint applicants. The quarterly *local unemployment rate* at the loan origination date for the labor market serviced by the credit union is also included in the analysis (U.S. Bureau of Labor Statistics, 2018). This variable serves as a proxy for local

economic conditions which change over time and which could impact a borrower’s need for credit as well as a lender’s willingness to provide financing. The average unemployment rate at loan origination date for the entire sample was 9.1% during the time of this analysis.

Columns (2) and (3) of Table 1 also include summary statistics for loans that do not end in default (non-default loans) and loans that end in default (default loans). Non-default loans were compared to default loans using a two-sample t-test for continuous variables and a two-sample z-test of proportions for dummy variables. The t and z statistics and p-values are reported in column (4) of Table 1. A negative t or z value indicates that the sample mean for default loans was greater than the sample mean for non-default loans. There are statistically significant differences at the 1% significance level between non-default and default loans for all the personal and loan-related variables. Results show that loans ending in default had statistically significantly lower credit scores, higher debt-to-income and payment-to-income ratios, higher loan amounts, higher interest rates, longer-term loans, higher loan-to-value ratios, and were issued in times of lower unemployment rates. Joint loans had statistically significantly fewer defaults than single applicant loans. Those under age 40 and over age 65 had statistically significantly more defaults than middle-age individuals (at the 1% and 10% significance levels respectively).

Because the focus of this analysis is determining the impact of vehicle make on default rates, the analysis also includes several variables related to the auto. Table 2 provides summary information on these variables, which include a dummy variable indicating whether the vehicle was an ATV/motorbike, a truck, or was used. Approximately

1% of the loans were for ATVs or motorbikes, while 24% of the loans were for trucks

rather than cars. 80% of the sample was comprised of loans for used autos. Loans on Chevrolet vehicles constitute the largest percentage of the dataset with 21.6% of loans.

**Table 1: Summary Statistics**

Other popular makes for this dataset include Dodge/Chrysler/Jeep (16.9%), Ford/Lincoln/Mercury (16.3%), GMC (12.7%), and Nissan/Infinity (14.6%). Table 2 also provides a distribution of the auto loans by auto make. The ATV/motorbike category experiences the largest percentage of default in the sample with loans for nearly half of these vehicle types ending in default. 11.35% of Acura/Honda loans and 9.77% of Hyundai/Kia loans ended in default in the sample. Toyota/Lexus and Mazda experienced the lowest rates of default in the sample at approximately 6%. When examining the z-test

results for vehicle-related variables in Table 1, loans for trucks had significantly fewer defaults while loans for ATV/motorbikes had significantly more defaults. The results suggest that GMC vehicles had significantly fewer defaults at the 10% significance level.

The dollar value of *exposure at default* is computed by the lender using a factor which annualizes these charge-off amounts. From Table 1, the average dollar value of exposure at default for loans ending in default is $5,043. 25% of the observations ending in default were below $2,158.68, 50% were below $4,407.96, and 75% were below $7,158.66. The average dollar value of exposure at default for all loans is $430.

The variable *duration* indicates the number of months that pass before default occurs. As noted in equation (13) above, the duration variable was created for each observation in the dataset, with . From Table 1, for loans which matured without defaulting, the duration equals the loan term in months. For the sample of loans which did not default, the average loan term is 4.5 years. For loans which ended in default, duration equals the month in which the default occurred. For the sample of loans which ended in default, the average duration was 2 years. The summary variable *percent complete* expresses the average duration in percentage terms. For loans which ended in default the typical borrower had completed approximately 40% of the loan at the time of the default.

**Table 2: Summary Statistics (Auto-Related Information)**



1. **Results**

*5.1 Logistic Regression: Who Will Default?*

The results of the estimation of default probabilities using logistic regression are shown in Table 3. Relative to middle age borrowers, those under age 40 and over age 65 were significantly more likely to default. Young borrowers sometimes lack the experience and income necessary to make payments on time. In our sample, borrowers older than 65 are also more likely to default, perhaps due to declining incomes and deteriorating health conditions. This pattern of greater risk for the lending institution is similar to that experienced in the auto insurance industry when the risks of accidents are high for both younger and much older drivers. In the sample, those under 40 had 1.4 times the odds of default and those over 65 had 2.4 times the odds of default compared to middle age borrowers. As expected, borrowers with higher credit scores were significantly less likely to default, with the odds ratio of 0.996 revealing a small decline in the likelihood of default when the credit score increases by 1.

Loans that were jointly owned were significantly less likely to default, likely due to the increased income available for the loan payments and increased accountability of

having two individuals responsible for the loan. Joint loans have odds of defaulting that

**Table 3: Logistic Regression Results, Dependent Variable: Default (dummy)**



are 0.56 times as great as those of non-joint loans. Loans with higher interest rates and longer terms are significantly more likely to default, although odds ratios for interest rate (eβ=1.238) and loan term (eβ=1.068) indicate little increase in the likelihood of default when the predictors increase by 1% and 1 month respectively. Higher loan to value ratios were statistically significant in predicting default, with a 1% increase in the loan to value ratio resulting in the odds of default that are 1.029 times higher. The unemployment rate at the time of the loan origination was also a statistically significant predictor of default. Specifically, the higher the unemployment rate at the time of the loan origination, the less likely the borrower was to default. The odds of defaulting are 1.6 times higher when the unemployment rate falls by 1%. This result provides evidence that the lender is more selective in a weak economy. Additionally, individuals feeling the strain of weak local economic conditions may either fail to qualify for loans or may opt out of purchasing autos as their budgets tighten.

 Loans for ATVs and motorbikes are 7.5 times more likely to default. This result suggests that individuals reveal their risk preferences through their auto purchases. Individuals purchasing ATVs or motorbikes may be more risk tolerant than individuals purchasing cars; thus, these risk tolerant individuals may be more of a risk for auto lenders. ATVs and motorbikes are usually purchased as an additional vehicle or used for recreational purposes; thus, for individuals facing financial problems, the ATV or motorbike is probably the first payment to be neglected.

Various makes were included as nine dummy variables in the analysis. Chevrolet was chosen as the control category due to its popularity in the local market and because its default rate of approximately 8% is similar to the overall default rate for the sample. Ford/Lincoln/Mercury vehicles were significantly more likely to default. The odds of default are 1.56 higher for Ford/Lincoln/Mercury vehicles compared to Chevrolet vehicles.

At the 10% significance level, GMC, Nissan/Infiniti, and Mazda vehicles are significantly less likely to default than Chevrolet vehicles while Dodge/Chrysler vehicles are significantly less likely to default than Chevrolet vehicles at the 5% significance level. The odds of default for Chevrolet vehicles are 1.54 times higher than GMC and Nissan/Infiniti vehicles, 2.94 times higher than Mazda vehicles, and 1.57 times higher than Dodge/Chrysler vehicles. These findings indicate lenders would be wise to investigate and understand the risk inherent in financing certain vehicle makes in their local market as they determine auto loan rates. The results further indicate that used vehicles are 1.5 times more likely to default relative to new vehicles.

**Table 4: Logistic Regression, Goodness of Fit and Significance Tests**



Table 4 provides information on goodness of fit. Results indicate that the overall model was statistically reliable in distinguishing between loans that would end in default and loans that would reach completion (-2 Log Likelihood = 1465.27, (23) = 599.095, p < 0.0000). The R-square measures indicate that approximately 30% of the variability in loan defaults is accounted for by the predictor variables.

A classification table for *default* is presented in Table 5. The classification table compares the predicted values for default based upon the regression model with the actual observed values from the dataset. If the predicted probability is less than 0.50, the

observation is classified as non-default. The model correctly classified 92% of the cases, indicating that the model is extremely accurate in classifying subjects. The specificity of the model, the ability of the model to predict which loans will not end in default, is 98.64% which indicates extremely accurate prediction. The sensitivity of the model, the ability of the model to predict which loans will end in default, is 18.87%, which indicates that the model is not as accurate in predicting default. This is not particularly troubling for our

research question, because it is acknowledged that predicting loan default is already a complicated task for lenders; if lenders could perfectly predict default, they would work to offset these risks using additional risk-based pricing measures. An improvement of approximately 19% predictive ability is significant for lenders as they seek to price loans and minimize losses.

**Table 5: Logistic Regression, Classification Table**



Figure I shows the receiver operating curve for the fitted model with sensitivity on the y axis and specificity on the x axis. The area below the forty-five degree line represents the classifications occurring purely by chance. The graph indicates that the fitted model is accurate since a substantial portion of the curve lies above the reference line. The area under the curve is 0.88.

***Figure I: Receiver Operating Curve for Estimated Default Model***



*5.2 Tobit Regression: What is the loss exposure when the borrower defaults?*

In addition to predicting the likelihood of default using logistic regression, this paper also seeks to determine which factors influence the *amount* of the default, known to financial institutions as their exposure at default. Predicting this value would be particularly useful for lenders since the expected loss depends upon the size of the balance left on the loan.

A tobit regression was calculated to predict the amount of the default (EAD)

based on personal, loan, and auto characteristics. The results are shown in Table 6. The likelihood ratio test indicates significance at the 1% level with the calculated chi-square equal to 600.35 with 24 degrees of freedom. As expected, significant predictors are similar to that shown in the logistic regression above. Individuals below the age of 40 and over the age of 65 have significantly higher default amounts relative to those between the ages of 40 and 65. Borrowers with higher credit scores and those with joint loans had statistically

**Table 6: Tobit Regression,**

**Dependent Variable: Amount Owed at Time of Default in $ (Exposure at Default)**



significantly lower default amounts. Default amounts were higher for those with higher debt to income ratios, larger loan amounts, higher interest rates, longer terms, and higher loan to value ratios. Loans issued in times of higher unemployment had significantly lower default amounts. Several makes were significant in predicting exposure at default with ATVs and motorbikes and Ford/Lincoln/Mercury vehicles leading to a greater exposure at default relative to the control variable of Chevrolet. Used vehicles had higher default amounts.

To accurately interpret the Tobit estimates, it is helpful to consider the computed marginal effects. The marginal effects are shown in columns 4, 5, and 6 of Table 6. While many of the variables have interesting marginal effects, we focus on those variables which might be of interest for lenders to consider as they attempt to mitigate default risk. Longer terms are becoming increasingly popular, so lenders should be aware of how this impacts their exposure at default. The marginal effects of the loan term variable are as follows: a one month increase in loan term would result in a $6 increase in the EAD for all borrowers in the sample (the unconditional expected value), a $28 increase in the EAD for borrowers with a positive default amount (the conditional expected value), and a 0.17% increase in the probability of a positive default amount. A 1% increase in the interest rate would result in a $20 increase in the EAD for all borrowers, a $94 increase for borrowers with a positive default amount, and a 0.57% increase in the probability of a positive default amount. Another predictor of interest is the *joint* variable. Joint borrowers have a $57 lower

unconditional expected default amount (for all borrowers) and a $285 lower conditional expected default amount (for borrowers with a positive default amount). Joint loans have a 1.63% lower probability of defaulting than non-joint loans. A 1% increase in the unemployment rate results in a $52 decrease in EAD for all borrowers, a $240 decrease for borrowers with a positive default amount, and a 1.45% decrease in the probability of default.

The marginal effect of ATV/motorbike indicates that a borrower would have a $381 increase in the unconditional expected default amount (for all borrowers) and a $905 increase in the conditional expected default amount. Loans for ATVs and motorbikes have a 8.81% higher probability of defaulting than loans for other vehicles. Marginal effects indicate that a borrower taking out a loan for a Ford/Lincoln/Mercury vehicle would have a $57 increase in the unconditional expected default amount (for all borrowers), a $228 increase in the conditional expected default amount, and a 1.54% increase in the probability of default.

Lenders could apply adaptations of the above two models to better protect against losses. Because individual borrower characteristics, borrower preferences, and the availability and popularity of vehicle makes are likely different in different markets, lenders would be best-served to develop risk-based models that are consistent with their specific market. Carrol and Zeltkevic (2007) support this recommendation, stating “a proprietary model built for a specific firm, based on its own client experiences will always outperform a generic model, as long as the underlying sample is large and well chosen.” Lenders should also be careful to follow the law when applying risk-based models. Following Agarwal, et al., (2008) we included *age of the borrower* as a variable in our analysis. Although age is an influential factor in the above models, the Equal Credit Opportunity Act (ECOA) prohibits credit discrimination on the basis of race, color, religion, national origin, sex, marital status, age, or because a borrower receives public assistance (Federal Trade Commission). While creditors may ask borrowers for this information in certain situations, they may not use it when deciding whether to extend credit or in determining the terms of the loan (Consumer Financial Protection Bureau). In some cases, a lender or dealer may relate borrower age to other information about the borrower that the lender or dealer considers in evaluating creditworthiness. For example, a lender or dealer may consider a borrower’s occupation and length of time to retirement to determine whether income, including retirement income, will be adequate for the life of the loan.

We suggest that lenders combine information on vehicle make with information on borrower and loan characteristics in predictive models such as the ones presented in Tables 3 and 6 to ultimately predict their expected loss (EL). While the focus of our paper is on predicting the probability of default and the exposure at default, the lender would need to combine information on the probability of default (PD) and EAD with information on loss given default (LGD) to predict their expected loss. LGD is the amount of borrowed money the lender loses if the borrower defaults. The LGD will depend upon the value of collateral, length of time to recover collateral and at what financial cost, and the amount spent in the recovery process (Stephanou & Mendoza, 2005). LGD is commonly expressed as a fraction of the exposure at default. After estimating LGD, it is possible for the lender to calculate the amount of the expected loss (EL) following Stephanou & Mendoza (2005) and Tanninen (2013) as:

*Expected Loss = PD x EAD x LGD* (18)

As an example, suppose that a loan applicant is determined to have a 10% chance of a default and an expected exposure at default of $4,000. If the lender has an average recovery rate of 50%, then the expected loss is $200. This expected loss could be incorporated into loan pricing in several ways such as an initial fee or by adjusting the interest rate.

*5.3 Censored Normal Regression: When will the borrower default?*

A third and final analysis in this paper is to determine the factors which influence the time to default or duration. Lenders would benefit from knowing what factors influence the length of time for which borrowers will honor their obligations. A censored normal regression model is used to predict the timing of the default for the sample since *duration* is right-censored. The results are presented in Table 7. The likelihood ratio chi-square of 532.73 (df=24) with a p-value of 0.0000 indicates that the model is significant. The natural log of *duration* is used as the dependent variable. In interpreting the results, a longer duration is more desirable. The duration until default is significantly shorter for those under

40 and those over 65 relative to those between the ages of 40 and 65. An individual under the age of 40 has a 23% [-1=-0.23] shorter duration until default and an individual over the age of 65 has a 53% [-1=-0.53] shorter duration until default. Loans to individuals with higher credit scores have a longer duration, but the size of the effect is negligible. An individual with a joint loan has a duration until default that is 66% [-1=0.66] longer, suggesting that joint borrowers are collectively better able to honor their obligations. A 1% increase in the interest rate reduces duration by approximately 17%. An increase in loan term of one year reduces duration by 45% [100x12(-0.038)]. Borrowers with higher loan to value ratios have a significantly shorter duration until default. The higher the unemployment rate at the date of issuance, the longer it takes the borrower to default on the loan. A 1% increase in the unemployment rate at issuance increases duration by 43%. Again, weak economic conditions may result in more selective lenders and more conservative borrowers, and ultimately lower loan amounts and less difficulty in continuing payments.

**Table 7: Censored Normal Regression,**

**Dependent Variable: Log(Duration), where Duration = Months Until Default**



GMC, Dodge/Chrysler, and Mazda vehicles had a significantly longer duration while Ford/Lincoln/Mercury and ATV/Motorbike vehicles had a significantly shorter duration, 29% [-1=-0.29] and 80% [-1=-0.80] shorter, respectively, relative to the control make of Chevrolet. Because ATVs and motorbikes are usually purchased as an additional vehicle or used for recreational purposes, payments for ATVs or motorbikes are likely to be neglected first in times of financial hardship. Based upon these results, the credit union should take additional precaution when lending for ATV/motorbike and Ford/Lincoln/Mercury vehicles. Lenders could compensate for the added exposure on these vehicle types by applying an initial fee or by adjusting the interest rate. These results provide a promising direction for future research and imply that more information would enable lenders to have added precision in predicting how defaults will impact their earnings.

# Summary and Conclusions

Using applicant-level and loan-level data from a credit union, this paper seeks to provide a better understanding of individuals as they make choices regarding auto loans, focusing on the impact of vehicle make. Logistic, tobit, and censored normal regression analysis is employed to analyze the impact of vehicle make on default probabilities, exposure at default, and duration until default.

When predicting the probability of default, our paper offers some of the same conclusions as Agarwal, et al., (2008). Individuals with higher credit scores are significantly less likely to default while those with higher loan to value ratios are significantly more likely to default. Additionally, we find that local labor market conditions are important indicators of default probabilities. Agarwal, et al., did not include information on joint loan performance; our analysis indicates that joint loans were significantly less likely to end in default. Our model also provides evidence that probabilities of auto loan default can be more accurately predicted by including the make of vehicle. Unlike Agarwal, et al., we did not find a significant effect of luxury vehicles in predicting default; however, we did find that loans for used cars were more likely to result in default. Specifically, Ford/Lincoln/Mercury vehicles were significantly more likely to default while GMC, Dodge/Chrysler, Nissan/Infiniti, and Mazda were significantly less likely to default relative to the control make of Chevrolet. Our analysis includes a category for ATVs and motorbikes, which is not studied in the previous literature. The results indicate that lenders should take precaution when issuing loans for ATVs and motorbikes as the default probability is greatly increased for these types of vehicles.

This paper evaluates the factors predicting exposure at default using a tobit regression model. As expected, the variables that were significant predictors of probability of default from the logistic regression are significant in predicting EAD. Notable results are that longer loan terms increase the exposure at default, but exposure is lower for joint applicants. Several makes were significant in predicting exposure at default with ATVs and motorbikes and Ford/Lincoln/Mercury vehicles leading to a greater exposure at default relative to the control variable of Chevrolet. We recommend that this specific lender uses greater caution in loans for ATVs and motorbikes as results indicate that these loans have a practically and statistically significant impact on exposure at default.

A third consideration addressed in this paper is duration until default which was analyzed using a censored normal regression. Like Banasik, et al., (1999) we find that the duration analysis provides a result that is consistent with the logistic regression approach; the factors which significantly predicted the probability of default also significantly predicted duration until default. The results indicate that this lender should take additional precaution when lending for ATV/motorbike and Ford/Lincoln/Mercury vehicles since they will default earlier than other vehicles.

This study has some limitations. The focus of the study was indirect auto loans from a credit union in the southeastern United States. Although our study is focused on a single lender, the model developed in this paper could be applied to other studies with larger sample sizes in different regions. The availability and popularity of vehicle makes and borrower preferences are likely different in different markets; thus, lenders would best be served to develop risk-based models that are consistent with their specific market. Because vehicle make was shown to be a significant predictor in all three models, lenders can be more confident in incorporating information on vehicle make in their risk-adjustment process. After estimating the probability of default and exposure at default using information on vehicle make, lenders could combine this information with estimates of loss given default to determine expected loss. Loans could be adjusted to compensate for the added exposure by applying an initial fee or by adjusting the interest rate. Hopefully, incorporating information about vehicle make in the decision-making process will enable lenders to offer more competitive rates and minimize losses.

**References**

Adams, William., Liran Einav, and Jonathan Levin, 2009. Liquidity Constraints and

Imperfect Information in Subprime Lending. *American Economic Review*, 99(1):

49-84.

Agarwal, Sumit., Brent W. Ambrose, and Souphala Chomsisengphet, 2007. ‘Information

Asymmetry and the Automobile Loan Market’, in Agarwal, S. and Ambrose, B.W. (Eds.), Household Credit Usage: Personal Debt and Mortgages, New York: Palgrave Macmillan, 93-116.

Agarwal, Sumit., Brent W. Ambrose, and Souphala Chomsisengphet, 2008.

Determinants of automobile loan default and prepayment. *Economic Perspectives*,

Federal Reserve Bank of Chicago, 32(3): 17-28.

Agresti, Alan. An Introduction to Categorical Data Analysis. (New York: John-Wiley &

Sons Inc., 1996).

Amemiya, Takeshi., 1973. Regression Analysis when the Dependent Variable is

Truncated Normal. *Econometrica,* 41(6), 997–1016.

Banasik, John., Jonathan Crook, and Lyn Thomas, 1999. Not if but When will Borrowers

Default. *The Journal of the Operational Research* Society 50(12): 1185-1190.

Bandyopadhyay, Arindam. “Exposure at Default (EAD) and Loss Given Default (LGD),”

in *Managing Portfolio Credit Risk in Banks*, (Cambridge: Cambridge University Press, 2016), 137-185.

Board of Governors of the Federal Reserve System (US), ‘Motor Vehicle Loans Owned

and Securitized, Outstanding [MVLOAS]’, retrieved from FRED, Federal Reserve Bank of St. Louis, accessed October 1, 2018. https://fred.stlouisfed.org/series/MVLOAS.

Bhardwaj, Sunil, and Kaushik Bhattacharjee, 2010. Modeling Money Attitudes to Predict

Loan Default. *IUP Journal of Bank Management* 9(1/2), 12-20.

Carrol, Peter, and Michael Zeltkevic, 2007. Safety in Numbers, *The RMA Journal*,

Philadelphia, 90(3), 22-27.

Consumer Financial Protection Bureau, “Can a Lender or Dealer Consider My Age…”,

accessed October 12, 2018. https://www.consumerfinance.gov/ask-cfpb/can-a-lender-or-dealer-consider-my-age-my-receipt-of-part-time-income-or-my-retirement-income-when-deciding-whether-to-give-me-an-auto-loan-en-1181/.

Credit Union Department - State of Texas., 2014, “RB 2014-03 Guidance on Indirect

Automobile Lending.” accessed October 1, 2018. http://cud.texas.gov/reports-

publications/rb-2014-03-guidance-on-indirect-automobile-lending

Dirick, Lore., Gerda Claeskens, and Bart Baesens, 2017, Time to Default in Credit

Scoring Using Survival Analysis: a Benchmark Study. *Journal of the Operational*

*Research Society*, 68(6), 652-665.

Einav, Liran., Mark Jenkins, and Jonathan Levin, 2013, The Impact of Credit Scoring on

Consumer Lending. *The RAND Journal of Economics,* 44(2), 249-274.

Experian, 2018, “State of The Automotive Finance Market: A Look and Loans and

Leases in Q1 2018.” accessed October 12, 2018.

https://www.experian.com/content/dam/marketing/na/automotive/quarterly-webinars/q2-2018-state-auto-finance-markets.pdf

Federal Reserve Bank of Atlanta., 2017, Looking under the Hood: Current Conditions in

Auto Lending [online]. accessed October 2, 2018. https://www.frbatlanta.org/economy-matters/banking-and-finance/viewpoint/2017/05/04/looking-under-the-hood-current-conditions-in-auto-lending

Federal Reserve Bank of New York. Research and Statistics Group. Microeconomics

Studies., 2018, “Quarterly Report on Household Debt and Credit: August 2018”.

accessed September 28, 2018.

https://www.newyorkfed.org/medialibrary/interactives/householdcredit/data/pdf/HHDC\_2018Q2.pdf

Federal Trade Commission, 2018, Your Equal Credit Opportunity Rights. accessed

October 5, 2018.https://www.consumer.ftc.gov/articles/0347-your-equal-credit-opportunity-rights.

Heitfield, Erik., and Tarun Sabarwal, 2004. What Drives Default and Prepayment on

Subprime Auto Loans?, *Journal of Real Estate Finance and Economics*,

29(4):457-477.

Hock-eam, Lim, and Siew Goh Yeok, 2017. Estimating the Determinants of Vehicle Loan

Default in Malaysia: An Exploratory Study. *International Journal of Management*

*Studies*, 24(1): 73-90.

National Credit Union Administration. Research a Credit Union, 2018, accessed October

10, 2018. https://www.ncua.gov/analysis/Pages/research.aspx

Stephanou, Constantinos, and Juan Carlos Mendoza, 2005. Credit Risk Measurement

Under Basel II: An Overview and Implementation Issues for Developing

Countries, Working Paper 3556, The World Bank, accessed October 5, 2018.

https://openknowledge.worldbank.org/bitstream/handle/10986/8557/wps35560cor

rected.pdf?sequence=1&isAllowed=y

Stokes, Gordon, and Sharon Hallett, 1992. The Role of Advertising and the Car.

*Transport Reviews*, 12(2), 171-183.

Tanninen, Matti., 2013. Determinants of Credit Risk in Secured Loans–Evidence from

the Auto Loan Industry. (Thesis, School of Business, Lappeenranta University of

Technology, Lappeenranta, Finland, 2013).

U.S. Bureau of Labor Statistics, 2018, Local Area Unemployment Statistics, accessed

September 3, 2018, https://www.bls.gov/lau/publications.htm

Wooldridge, Jeffrey M., Introductory Econometrics: A Modern Approach. (US: South

Western Educational Publishing, 2002) 575-612.