Is there any dependence between consumer credit line utilization and default probability on a term loan? Evidence from bank-level data*

Anne-Sophie Bergerès
Caisse de dépôt et placement du Québec

Philippe d'Astous
HEC Montréal – Department of Finance

Georges Dionne
HEC Montréal – Department of Finance, CIRRELT, and CIRPÉE

8 February 2013

Abstract

Whereas recent studies on revolving lines of credit suggest a positive relationship between exposure at default and default probability on the line, this paper considers the relationship between two financial instruments through the simultaneous analysis of credit line utilization and default probability on a personal loan. We model both financial instruments endogenously in a simultaneous equation system and find strong evidence of a positive relationship between the two instruments. Individuals in the default state use their credit line 59% more than those in the non-default state, and full utilization of the credit line increases the default probability on the loan by 46% when compared with non-utilization. Our results suggest that banks should manage both financial instruments simultaneously.

Keywords: Consumer finance, consumer risk management, credit line, term loan, default probability, ability to pay, endogeneity, simultaneous equations.

JEL numbers: D12, D14, G01, G21, G33.

* *Financial support by SSHRC and CGI is acknowledged. The authors thank Hind Diboune and Denise Desjardins for their contribution to this research, and Benoit Dostie for his very useful comments.*

Is there any dependence between consumer credit line utilization and default probability on a term loan? Evidence from bank-level data

Consumer finance has been neglected for many years in the economics and finance literature (Campbell, 2006; Campbell et al. 2010; Tufano, 2009), but the recent financial crisis clearly shows that consumer financial services are not well managed, and deserve further study. New revolving credit instruments have been made available to consumers during the past decades; their utilization has grown tremendously in recent years. For example, the Canadian consumer credit line market expanded by 133% from 1999 to 2005. Currently, its value far surpasses that of credit card debt and personal loans combined. In the United States, outstanding total revolving consumer credit, including credit card balances, represents 33% of total consumer debt in 2011. The presence of these relatively new and popular credit instruments should have a significant impact on consumer financial distress, and may affect conventional credit instruments such as term loans.

The dataset used in this study comes from a leading Canadian bank, and is composed of retail borrowers at the institution. We focus on individuals who use both a revolving line of credit (credit line hereafter) and a term loan. The two dependent variables of interest are the percentage of the credit line drawn down relative to the commitment amount of the bank and the default state of the term loan. Early signs of high revolving credit utilization may signal future financial distress, reflected by a higher default probability on the term loan. In this original dataset, most credit lines are granted upon credit evaluation without strict collateral requirement. The revolving line is featured as an add-on to the checking account, allowing the borrower to use the funds up to the authorized limit. The lines are mostly unsecured, contrasting with home equity lines of credit (HELOC), another popular revolving credit instrument, which requires home equity as collateral. The credit lines we study share some similarities with credit cards, but differ with respect to their repayment schedule and their much lower interest rates. The term loans in the dataset are also granted upon credit approval, and are not necessarily subject to collateralization. Both credit instruments should be viewed as consumption credit, in contrast with mortgage-linked or home equity credit.

We analyze both dependent variables in a simultaneous equation model in which credit line utilization is modeled by an instrumented Tobit, and default probability on the loan is modeled by an instrumented Probit. Our assumptions are that both variables are endogenous and that their relationship is positive. Excessive credit line utilization may signal bad liquidity shocks, ultimately affecting the borrower's default probability on their term loan. Conversely, a high default probability on the term loan may lead borrowers to use the liquidity made available by the line excessively, in the hope of reducing the impact of financial distress. Banks interpret such behavior as a signal of credit quality deterioration: this may lead to better risk management of the loan portfolio if corrective measures are taken early.

To deal with the endogeneity problem of the system of equations, we instrument the credit line utilization and the default probability variables. Our results show that exposure at default (EAD), modeled by credit line utilization, is endogenous in the default probability equation of the term loan. Further, the default probability variable is endogenous in the credit line utilization equation. We use valid instruments to overcome the endogeneity problem and to successfully estimate both equations. We find a positive relationship between the two variables, and we verify that the number of active credit lines plays a significant role in the borrower's probability of default on the term loan. We also propose a new measure of debtholder ability to pay back the loan, which complements the credit score variable, to explain the default probability and the credit line utilization. With the proposed specification, the borrower's ability to pay significantly affects credit line utilization--those with the worst ability to pay are more likely to default on the term loan.

To the best of our knowledge, we are the first to highlight the link between revolving credit utilization and the default probability on a different instrument for consumers. Because a default on the term loan may ultimately degenerate into borrower default, banks should carefully manage the interactions between a borrower's credit instruments.

The rest of the article is structured as follows. Section 1 presents a brief definition of consumer lines of credit. Section 2 reviews the existing literature and explains how this study complements it. We depart from what has been done previously by analyzing two different credit instruments simultaneously and controlling for endogeneity. Section 3 presents some general statistics from the Canadian and US markets. The variables used and the dataset are specified in Section 4. The descriptive statistics of the sample are covered in Section 5, while Section 6 provides the hypotheses and methodology. Results are presented in Section 7 and Section 8 concludes the paper.

1. General definition of revolving lines of credit

In its most general definition, a line of credit is the maximum amount a bank commits to grant a borrower for a predetermined period of time. Borrowers who are accepted can use (*i.e.* borrow and repay) funds up to the maximal amount authorized on the credit line. Two types of lines of credit are usually available to consumers: credit cards and revolving lines of credit. One of the main differences between these two instruments is that credit cards usually require no collateral from the client, making them unsecured loans for the institution, whereas credit lines are sometimes secured by collateral. Credit lines are also characterized by a lower interest rate. Some credit lines are secured by the equity owned on a house; they are called Home Equity Lines of Credit (HELOC). However, not every credit line is backed by collateral. For small consumer lines of credit, banks usually require a less valuable asset as collateral or no collateral at all, relying solely on the borrower's credit evaluation for the acceptance decision.

Our dataset differs from what has been widely used in the existing literature because it does not include HELOC or credit cards; upon credit approval applicants are granted a revolving line of credit, attached to their checking accounts. For consumers in our sample, a line of credit usually has no maturity; the bank renews the account periodically as long as it remains active. A minimum monthly payment of 3% of the balance is usually required to keep the line active. The interest charges are added to the amount drawn down by the borrower, and if a client reaches the borrowing limit on the line, the institution can renegotiate the contract into a term loan on which interest and capital must be repaid monthly. However, even upon reaching the borrowing limit, the credit line may remain active if the borrower keeps making the minimum required payments .

2. Literature review

The relevant literature for our study covers both corporate and personal revolving credit analyses. Although some findings are similar in nature, one must be careful when making inferences from the corporate literature to the consumer literature. Recent studies suggest a positive relationship between exposure at default (EAD) and default probability (D) on lines of credit. For example, looking at HELOC, Agarwal et al. (2006a) find that consumers with an *ex ante* higher expectation of credit deterioration use a smaller part of their line of credit at origination. Their findings also suggest that the drawing behavior after origination is inversely related to changes in the borrower's credit quality. This is reflected by higher credit line utilization for borrowers who end up defaulting on the line, and implies that the exposure at default should not be treated independently from the default probability on the credit line. The authors do not calculate the default hazards because not enough borrowers are in default in their sample. We do not have access to such a panel dataset, but we match credit line utilization to default probability on a different credit instrument, and we extend the analysis to address endogeneity.

Studying corporate lines of credit, Jiménez et al. (2009) find similar results and evidence of a positive relationship between the probability of default and the amount drawn by a firm on its credit line. This relationship is strongest near the default event. Such findings show a quality reduction in the credit line market; consumers with deteriorating credit quality are the ones who use their credit lines the most.

Agarwal et al. (2006b) study the differences between the actions individuals take on Home Equity Loans and HELOC. They find that users of these two instruments exhibit differing risk profiles. Their sample shows that borrowers using a term loan have a lower credit rating and contract significantly smaller amounts than individuals using a line of credit. They also assert that the probability of prepayment is higher on lines of credit than on term loans. Their analysis shows that the probability of default on credit lines is lower than on term loans, the latter being more sensitive to changes in the value of the home and twice as sensitive to changes in interest rates. Because borrowers usually have either a HELOC or a Home Equity Loan, they do not highlight the potential link between this revolving credit instrument and a loan, as we do in a different context. Our analysis is possible because our sample consists of consumers' credit activities at the institution; it is not unlikely for borrowers to use both a revolving credit line and a term loan in this context.

Calem et al. (2011) perform panel data estimation and risk hazard analysis on HELOC borrower data to test for an adverse selection effect in the credit line market. They find that contrary to neoclassical theories of consumption, tightening economic conditions are associated with a negative shift in the credit quality of the borrower pool. Further analysis shows that drawdowns when the anticipation of future income is low are associated with a higher probability of borrower delinquency. Such evidence is consistent with an adverse selection problem in the market. We do not consider macroeconomic factors in the analysis because we use a cross-sectional dataset. We model the differences across borrowers instead of the differences across time.

Dey (2005) studied consumers' decision to incur debt through a credit card rather than through a personal line of credit. He finds that theoretically, the collateral required on a line of credit can make consumers reluctant to use the full loan amount authorized by the bank. Depending on their utility function and corresponding risk aversion, agents may prefer to use an unsecured credit card, even if it bears a higher interest rate. This explains why some people carry a positive balance on their credit cards even though their credit line is not maxed out.

Studying the credit card market, Dunn and Kim (1999) find evidence that variables such as the ratio of minimum required payment on the card, the percentage of the line used and the number of cards on which the borrower has reached the limit have a highly significant impact on the default probability. They argue that these variables represent a short-term assessment of the default probability, and may replace the usual debt-to-income ratio used in most research on default probability determinants. Their data come from an original survey but do not directly include bank-level data. Like them, we consider the number of credit lines a borrower has contracted in our default probability equation.

Norden and Weber (2010) investigate checking accounts and the information they supply regarding credit line utilization for both firms and consumers. Using a German dataset, they show that banks tend to monitor clients' credit line utilization through information on the use of their checking accounts. Although there are many similarities between personal and corporate lines of credit, some differences are worth noting. Empirical evidence shows that firms often refinance their lines of credit as term loans before the maximum amount authorized is used, possibly to preserve their short-term borrowing capacity (Agarwal et al., 2004). This happens because firms pay a percentage fee according to the unused portion of their line, whereas consumers do not. Only corporate credit lines are subject to material adverse change (MAC) clauses (Agarwal et al., 2006a). At one specific institution examined, the bank has full commitment on the authorized amount on the credit line during its active period.

Using a Canadian dataset, Mester, Nakamura and Renault (2007) find that monitoring firms' transaction accounts may provide useful information on borrower credit quality. They show that banks intensify the transaction account monitoring activity when loans are perceived as deteriorating. They argue that such monitoring helps the lender gather information about the borrowing firm's accounts receivable and inventories. In our sample, credit lines are attached to the consumer's checking account. Monitoring credit line utilization for consumers can potentially provide information about borrowers' short-term credit quality, as will be shown below.

Strahan (1999), who studied the corporate credit market, argues that liquidity needs and default probability are interrelated, the former being higher when financing from other sources is harder to obtain, usually because the borrower's default risk is above average. He also notes that a typical deal between a firm and its bank should include both a term loan and a credit line. Combining both short-term and long-term borrowing would help an institution divide monitoring costs between these two loans. It can thus gain a better understanding of the borrower's credit quality. Information on both credit instruments should be combined so that credit line drawdown behavior can be connected to the default risk assessment of a term loan and vice versa. We pursue this idea empirically for individual borrowers.

3. Statistics on credit lines

Of all household financial products, lines of credit have grown the most in recent years. The Statistics Canada 2005 Survey of Financial Security reports that credit lines accounted for 9.0% of total Canadian household debt in 2005, versus 5.7% in 1999. This increase represents a change of 133.2% in constant 2005 dollars. At that time, personal credit lines were the most prevalent type of debt after real estate mortgages, which accounted for 75.3% of total Canadian household debt. According to this study, the total amount of debt on credit cards and installment credit[1] has risen by 58.4% (to 3.4% of total family unit debt in 2005) during the same period, the second largest increase after that of credit lines.

These numbers are consistent with an upward trend in consumer lines of credit, which started in the 1990s. Figure 1 presents the assets of Canadian chartered banks used for consumer credit, excluding mortgages for the period of November 1981 to March 2012. [2] During that time, the proportion allocated to personal loans did not increase as much as that allocated to lines of credit and credit cards. Since 1997, the total value of lines of credit has increased dramatically, far exceeding the total value of personal loans. In March 2012, the portion of assets dedicated to personal lines of credit was the largest, almost twice the value of credit card and personal loans combined. This suggests that the proportion of credit line debt used by households is now even higher than what is reported by the Statistics Canada 2005 study.

[Insert Figure 1 about here]

Further, personal revolving credit utilization has risen sharply in the US since the mid-1990s. The latest numbers provided by the Federal Reserve[3] (FBR: G.19) show that in March 2012, outstanding revolving credit for consumers totaled $803.6 billion, representing 31.61% of the total outstanding consumer debt. This number represents all types of revolving consumer credit. The significant increase in revolving credit usage warrants the inclusion of these instruments in economic and financial models to fully understand the evolution in households' consumption decisions (Dey, 2005).

[1] From Statistics Canada: "Instalment debt is the total amount owing on deferred payment or instalment plans where the purchased item is to be paid for over a period of time."
[2] Source: Series v36867, v36868 and v36869 of CANSIM.
[3] Federal Reserve Statistical Release G.19, available online: http://www.federalreserve.gov/releases/g19/current/g19.htm.

4. Variables Used and Dataset

Table 1 presents the variables used in the analysis; they can be divided into three categories:

(a) Variables that affect both the credit line utilization and the default probability on the loan

(b) Variables used as instruments for credit line utilization

(c) Variables used as instruments for the default probability

[Insert Table 1 about here]

4.1 Explanatory variables

To control for the characteristics of borrowers in the sample, we include information about their sex, credit score, seniority at the bank, working situation, number of dependents, age, and ability-to-pay. The credit score is the one calculated by the institution when the borrower applied for credit. It ranges from 1 to 8, with lower values indicating a better credit quality. We develop an ability-to-pay variable to take into account borrowers' income and expenditures, which can have economic significance beyond the information provided by the credit score. Some authors have argued that the information collected by financial institutions on potential individual borrowers is not always accurate (see, for example, Finlay, 2006). Specifically, the information concerning expenditures would be the most problematic because borrowers have incentives to misreport, and such information is hardly verifiable by the bank. We suspect that this is the case in our sample: the mean monthly expenditures reported is $247.39. This does not seem realistic because this amount putatively includes rent and all other personal expenses. Following the methodology used by Finlay (2006) for the United Kingdom, we estimate these expenses based on a larger dataset compiled by Statistics Canada. By taking into account variables such as sex, age, income, type of housing and dependents in an ordinary least squares estimation, we rigorously associate expenditures with each borrower in our sample and construct the ability-to-pay ratio. Results of the estimation used for predicting expenditures in our sample are presented in Table 2.

[Insert Table 2 about here]

We then develop a ratio that measures borrowers' financial constraints. The definition proposed for the ability-to-pay variable is the ratio of the monthly payment on the loan to the borrower's capacity to pay (expressed as income plus tangible assets minus expenditures). Income and expenditures are expressed in monthly amounts, and the tangible assets are divided by the remaining months on the loan contract to fit the time frame. This allows us to capture a short-term assessment of the ability-to-pay. The ratio is expressed as:

$$\text{Ability-to-pay ratio} = \frac{Monthly\ payment\ on\ the\ loan}{Income + tangible\ assets - expenditures} \tag{3}$$

A negative value for this ratio implies that the expenditures are higher than the income and tangible assets added, while a value of more than one implies a monthly payment higher than the customer's ability to pay. More prudent borrowers have a small but positive ratio.

This ratio has been segmented in different categories to better determine the point at which it becomes economically significant. The segmentation is presented in Table 3. Category 5 is taken as the reference category in the next estimations; it corresponds to the category of borrowers with the best ability-to-pay.

[Insert Table 3 about here]

4.2 Instrumental variables

To overcome the endogeneity bias of the model, we use statistically valid instruments for the default probability and the credit line utilization ratio. As in all studies using instrumental variables, the dataset dictates the choice of instruments. The instrument sets have been validated by an Amemiya-Lee-Newey overidentification test.

To instrument the default probability variable, we use the remainder term of the loan expressed in months, the number of active credit lines, and a dichotomous variable indicating whether the loan is secured by collateral. While these variables should be correlated to the default probability conditionally on all other exogenous variables of the model, neither instrumental variable should be correlated to credit line utilization. To instrument the credit line utilization variable, we use the amount drawn on the line and a dichotomous variable indicating the presence or absence of collateral on the credit line. In this case, the instruments should be correlated to the credit line utilization conditionally on all other exogenous variables of the model, and should not be correlated to the default probability.

4.3 Dataset

The dataset we use is original; it has been granted by a leading Canadian financial institution. It consists of a random sample of 37,440 observations of the institution's borrowers with information ranging from January 1, 2005 to December 31, 2007. There is no reason to believe that poorly performing accounts were eliminated from the sample because we observe cases of borrowers with as many as 43 cumulative late payments on the loan. After eliminating duplicates and observations for which information is missing, we obtain a full sample of 34,404 observations on borrowers.

We are interested in individuals using both a credit line and a term loan; they represent 43% of the full sample. We retain only those borrowers using both credit instruments; we thus perform our analysis on a cross-sectional dataset of 14,767 observations at December 31, 2007. At the time of extraction, the bank considered 160 clients to be in default on the term loan (*i.e.,* more than three consecutive monthly nonpayments), a proportion of 1.08% of the database. The extraction date paints a portrait of Canadian households when the financial crisis was already well under way, although the crisis did not affect the Canadian banking system significantly.

Because people who apply for a revolving line of credit face an acceptance or rejection decision made by the bank's credit department, our selection procedure to keep borrowers using both a revolving line of credit and a term loan might entail a bias towards more creditworthy borrowers. To control for such potential bias, we estimate a Heckman selection model (Heckman, 1979) and test for the statistical significance of the inverse Mills ratio. To do so, we first model the variables influencing the likelihood of a borrower's using a line of credit in a Probit regression where the dependent variable is the binary variable equal to 1 if the borrower has a line of credit, and 0 otherwise. As previously shown, consistency of the estimates requires a linear regression model to be used in the second step estimation of the credit line utilization and default probability equations. For this reason, we follow the procedure featured in Wooldridge (2002) for the case of an endogenous independent variable. We use a Two-Stage Least Squares technique to test the statistical significance of the inverse Mills ratio in the second step of the Heckman procedure for both equations. We also carefully use exclusion restrictions to avoid perfect correlation of the inverse Mills ratio in the second step estimation. For example, total debt is used in the selection equation but not in the outcome equation, while the ability-to-pay ratios are used only in the outcome equation. The results for the selection equation are presented in Table 4. From these results, we calculate predicted probabilities and use them in the inverse Mills ratio.

[Insert Table 4 about here]

Table 5 shows that the inverse Mills ratio is not statistically significant at the 10% confidence level in either the credit line utilization or default probability equations. The two sets of estimation results are very similar for each equation, showing that the inverse Mills ratio does not affect significantly the results. This implies that we can pursue our analysis without considering any sample selection bias.

[Insert Table 5 about here]

5. Descriptive statistics

The data we use cover information the bank acquired when it originally accepted the customers. The mean (median) amount authorized on the credit lines is $5,037.19 ($5,000) and the mean (median) value of the term loans is $14,125.86 ($12,878.79). No instruments in the sample are used over the authorized limit. This means the bank most likely applies mechanisms to prevent overdrafts on the credit line. The financial instruments that compose our sample are not tied to the equity on the borrower's home, and should be considered consumption credit. One important contrast with the loans that make up our dataset and the usual HELOC or Home Equity Loans concerns collateralization. For example, only 4.26% of the credit lines that make up our dataset are secured by collateral, compared with about 23.82% of term loans. The bank relies on the applicant's credit history to decide whether or not to grant the line of credit. Although the institution issues only one credit line per customer, we introduce a variable indicating how many more credit lines are active for the customer at other institutions. No other information on the additional lines is available, which makes it hard to assess external limits on the borrower's available credit at other institutions. This may nonetheless serve as a proxy for consumers' number of bank relationships.

Table 6 provides further information on the composition of the sample. Slightly more than one percent of observations are considered in default on the term loan. Because the default state is a binary variable, we use the Probit model to estimate its probability. The table also presents the proportion of individuals with zero and total credit line usage, and the proportion of observations lying between these two extremes. This distribution justifies the Tobit modeling of the credit line utilization variable, although some authors have also used OLS estimation (e.g. Agarwal et al., 2006a; Jiménez et al., 2009b), as we also report.

[Insert Table 6 about here]

Table 7 shows the individuals at corner solutions of credit line utilization as a proportion of the loan status. 38.75% of the observations for which the loan is considered in default use their entire credit line. Only 3.96% of the observations in the non-default group have reached such utilization. The extreme cases in the sample are thus in line with our hypothesis.

[Insert Table 7 about here]

Table 8 presents the mean and median of each variable, depending on whether the loan is in default or not. We test whether these statistics are significantly different across the default and non-default groups. The mean (median) credit line utilization is 78% (99%) for the default group compared with 45% (43%) for the non-default group. These numbers are in line with our hypothesis. In our sample, the higher the credit score, the riskier the borrower is considered by the institution. This is reflected across both groups; the means and medians are statistically different. Regarding the borrower's age and seniority at the institution, the median test shows that defaulters are both significantly younger and have a shorter business relationship with the bank. The control variables of number of dependents, sex and work indicators are not statistically different across both groups. For the non-segmented ability-to-pay ratio, the difference across both groups is statistically significant for the median, but not for the mean. Borrowers with the worst ability-to-pay are more likely to be in the default group. Individuals facing default on the term loan have a shorter median of remaining term on their loans, while the number of active lines of credit is not statistically different between both groups.

[Insert Table 8 about here]

6. Hypotheses and methodology

Research on credit line utilization has mainly focused on the relationship between borrowers' drawdown behavior and the default probability associated with their credit line. The literature generally suggests a positive relationship between drawdown and the default state; lines that end up in default are usually the ones that were used the most.[4] This has significant implications for the calculation of the EAD on the line because the relationship with the default probability must be taken into account.[5]

[4] See Agarwal et al., 2006a, Jiménez et al., 2009b, and Norden and Weber, 2010.
[5] See, for example, Jiménez et al. (2009a).

In contrast, the main hypothesis of our study concerns the relationship between credit line utilization and default probability on a term loan. Due to limited data, we could not perform a dynamic analysis of the two variables. Instead, we analyze the link between the level of credit line utilization and the default probability on the term loan with a cross-sectional dataset of borrowers. Our hypotheses are that both variables are endogenous, and that the relationship between them is positive. Specifically, when borrowers face financial distress and an increase in the default probability on the term loan, they may apply to the liquidity made available by the credit line toward the monthly payments on their term loans. Because the repayment schedule of the loan is tighter than that of the credit line, it may be to the borrowers' advantage to use the maximum amount available on their credit line to avoid default on the term loan. A higher default probability on the loan may thus result in more aggressive revolving credit utilization. Conversely, when borrowers experience bad liquidity shocks, the credit line is the first instrument on which they might be tempted to draw. Such abnormal liquidity shocks are reflected by higher credit line utilization, and may ultimately lead to financial distress, as reflected by a higher default probability on the term loan. For these reasons, we conjecture that both variables are endogenous and that the relationship between them is positive.

To consider the endogeneity of the dependent variables, we build a simultaneous equation model in which one equation represents the default probability on the loan and the other represents the percentage use of the credit line. The bank monitors the monthly non-payments by borrowers on their term loan; after more than 90 consecutive days late they are considered to be in default on the loan. We establish the default state on the loan with a variable provided by the financial institution. A loan classified as "bad" in the database is presumed to be in default. This rating is the worst the institution can attribute to a customer. It reflects more than three consecutive monthly late payments on the loan. Credit line utilization is calculated as the ratio of the amount drawn down on the line on the extraction date divided by the maximum amount allowed.

Neither dependent variable can be assumed to follow the normal distribution, because the default probability is a dichotomous variable and the credit line utilization is bounded by zero and one (respectively usage rates of 0% and 100% of the line). With these specifications, the former variable is modeled by an instrumental Probit equation (1) and the latter variable by an instrumental Tobit equation (2). The model can be represented by the following system:

$$y_1 = \alpha_1 + \beta_1 y_2 + \boldsymbol{z}_{(\mathbf{1})}\, \delta_{(1)} + u_1 \tag{1}$$

$$y_2 = \alpha_2 + \beta_2 y_1 + \boldsymbol{z}_{(\mathbf{2})} \delta_{(2)} + u_2. \tag{2}$$

In such a system, the two endogenous variables are y_1, the default probability on the term loan, and y_2, the percentage of the line drawn down. The vectors $\boldsymbol{z}_{(\boldsymbol{i})}$ of explanatory variables and $\delta_{(i)}$ of parameters are used for the exogenous factors of the model. Variables u_1 and u_2 are the random error terms for equations (1) and (2) respectively.

We estimate three sets of results for this system (see Table 9). The first set of results is estimated with Newey's Two-Step Efficient Estimator (Newey, 1987), a limited information procedure. Standard errors for this estimator are based on Amemiya's (1978, 1979) derivations of the efficient variance-covariance matrices. We thus validate the chosen instruments for each endogenous variable by an Amemiya-Lee-Newey overidentification test (Lee, 1992). The second set is estimated from a full information maximum likelihood procedure and provides a simple Wald test of the exogeneity of an explanatory variable. It allows us to test the endogeneity of the default probability and the credit line utilization variables. For robustness, the last set of results is a joint two-step estimation of OLS and Probit models, known as Two-Step Probit Least Squares (2SPLS). All procedures are instrumented for the endogenous variables of the system.

[Insert Table 9 about here]

7. Estimation results

The second stage results of these three estimations are qualitatively identical, and are presented in Table 9. We derive the marginal effects in Table 10 from the full information maximum likelihood procedure, which is the preferred set of results.

7.1 Endogeneity-Related Tests

Our first hypothesis about the relationship between credit line utilization and default probability on a term loan is that both variables are endogenous. To test this hypothesis, we perform a Wald test after the maximum likelihood estimation. Results, presented in Table 10, show that for both equations, the exogeneity hypothesis is rejected and each variable is considered endogenous in the equation for which it is used. To test the validity of the instruments, we use the Amemiya-Lee-Newey overidentification test available after the minimum chi-square estimation. This test concludes that both instrument sets are valid, and confirms that we have eliminated the endogeneity bias in the system. The result of this test is also presented in Table 10.

[Insert Table 10 about here]

7.2 Marginal Effects

We present the marginal effects at the average of the explanatory variables for the model obtained from the maximum likelihood estimation, to capture the conditional effects. For each of the equations studied, it is possible to derive different marginal effects. We retain the marginal effects of the linear prediction for the probability of default equation, and the marginal effects of the latent variable for the credit line utilization equation; they are presented in Table 10. Table 11 illustrates the Tobit decomposition according to McDonald and Moffitt (1980) for the credit line utilization equation.

7.3 Default Probability Equation

The credit line utilization variable is highly significant in the default probability equation, and is quantitatively the most important factor affecting default probability. The result suggests that higher utilization of the credit line has a statistically significant impact on increasing the likelihood of default on a term loan. This confirms our hypothesis and shows that it is important for financial institutions to assess consumer default probabilities jointly with the use of various financial instruments. This variable is even more economically significant than the credit score assigned by the financial institution. The use of a line of credit is hence a potential signal of the default probability of a consumer term loan. The marginal effect of this variable on the probability of default on the term loan is approximately 46%. The interpretation is that borrowers who go from 0% to 100% utilization of the credit line experience a 46% increase in their default probability on the term loan, everything else held constant. Alternatively, at the sample average, a 1% increase in use of the line causes an increase of 46 percentage points in the default probability.

The credit score is also highly significant and positive. It shows that the bank has successfully managed to classify customers according to the risk they present. Obviously, a riskier consumer has a higher probability of default. However, even after controlling for the credit score, credit line utilization is still an important determinant of default probability.

The first two categories of the ability-to-pay ratio are significant, and show that individuals with a lower ability to pay have a higher probability of default. This significance disappears in categories 3 and 4, comprising individuals with better financial capacity. This leads us to believe that the ability to pay is crucial for the determination of the probability of default on a term loan when the borrower is very financially constrained. The economic effect is important. People with a very poor ability to pay would therefore send a signal of high probability of default to financial institutions, which should use this variable to assess the default risk of their customers as a supplement of the credit score information. It indicates, indirectly, that the credit score variable is not sufficient to assess default probability. The credit score variable is, in fact, more of a proxy for credit delinquency than a measure of the customer's financial constraints.

The number of active credit lines is particularly interesting in this equation; it is highly significant and its economic impact is very important. The results clearly show that borrowers with multiple lines of credit have a higher probability of default on their term loan. The sign of this variable implies that individuals with a high number of active lines of credit do not gain more liquidity. Instead, each additional credit line increases the default probability by 26.39%. We posit that the various lines of credit an individual possesses negatively affect their credit quality because they are already heavily drawn down. It would thus be important for financial institutions to use this variable in their decision models to monitor borrowers' credit activity and to be kept informed of other credit lines contracted.

7.4 Credit line utilization equation

As assumed, the default probability on the term loan is one of the most important determinants of credit line utilization. This variable is significant and is quantitatively the most important one in the model. This confirms the hypothesis that increasing the probability of default on a term loan leads borrowers to use their lines of credit more aggressively, even when the risk rating assigned by the bank is taken into account. The marginal effect of this variable indicates that individuals in default on a term loan use their line of credit about 59% more than individuals who are not in default. The credit score variable suggests that borrowers who present a higher risk for the bank are the ones who use their lines the most.

As for the borrower's ability to pay, results show that in comparison to the omitted fifth category (the best ability to pay), borrowers with the worst ability to pay are the ones who use their credit lines the most. This is very intuitive, because the variable measures the borrower's financial constraints, and thus serves as a measure of short-term liquidity needs that can be met by credit line utilization. Our results confirm that borrowers who have difficulty making loan payments rely more on the use of their credit line. Further, the effect on credit line utilization for category 4 (Abil 4) is more than double that of category 1 (Abil 1), emphasizing the fact that borrowers with a worse ability-to-pay ratio need to rely more on the liquidity provided by the credit line.

McDonald and Moffitt (1980) decompose the marginal effects for a Tobit model with a lower limit into two categories: the intensive margin and the extensive margin. Their analysis can be applied to the Tobit model used in this paper. Such decomposition allows an analysis of individuals' drawdown behavior depending on their usage category. It provides insights about borrowers who shift from zero ($y_2 = 0$) to moderate ($0 < y_2 < 1$) or total use ($y_2 = 1$) of the credit line. The unconditional expectation of the dependent variable can be written as:

$$E[y_2] = P(y_2 = 0) * E[y_2|y_2 = 0] + P(0 < y_2 < 1) * E[y_2|0 < y_2 < 1] + P(y_2 = 1) * E[y_2|y_2 = 1] \quad (4)$$

By taking the derivative with respect to the explanatory variables, we get:

$$\frac{\partial E[y_2]}{\partial X_k} = \frac{\partial P(0<y_2<1)}{\partial X_k} * E[y_2|0 < y_2 < 1] + \frac{\partial E[y_2|0<y_2<1]}{\partial X_k} * P(0 < y_2 < 1) + \frac{\partial P(y_2=1)}{\partial X_k} \tag{5}$$

There are thus three effects to take into account: 1) the extensive margin; 2) the intensive margin; and 3) the change in the probability of using 100% of the credit line. We present the extensive and intensive margins in Table 11. The statistical significance remains very similar; differences come from the estimation of the marginal effects used in the decomposition.

[Insert Table 11 about here]

The important result of the above analysis is that the sign of the default probability variable is not the same for the extensive margin as for the intensive margin. Therefore, the default probability has a positive impact on the marginal use of a credit line for individuals who already have a utilization rate between zero and one. Individuals facing default on the term loan thus decide to increase their credit line utilization to meet their liquidity needs. Even in situations of financial distress, individuals who already use their credit lines would therefore not hesitate to take on more debt to fulfill other financial obligations. However, for individuals who do not use their lines of credit or who already use the maximum amount authorized by the bank, the default situation has a negative impact on the latent utilization rate. Indeed, the effect of the probability of default at the extensive margin is negative. Because the bank has full commitment on the amount authorized even if the borrower defaults, we can rule out a supply effect caused by the bank cutting off the funds available on the credit line. This shows that individuals who are not using their credit lines would be reluctant to start using them in the event of a default on the term loan. Perhaps such borrowers do not wish to aggravate their financial situation. This effect, however, also includes individuals who already use the line fully, although their proportion is very small in the sample. The marginal effects presented thus confirm the assumption of our model and suggest that the situation of default on a term loan affects individuals' decisions to draw on their lines of credit.

8. Conclusion

This research studies two credit instruments simultaneously through joint modeling of credit line utilization and default probability on a term loan. Research on credit line utilization has generally focused on the relationship between the borrower's drawdown behavior and the default probability associated with the line. We innovate by analyzing two financial instruments simultaneously and by quantifying the effect of each instrument on the other. The model thus highlights the need to evaluate credit risk for a portfolio of financial instruments held by borrowers. We estimate a simultaneous equation model in which the default status on the loan is modeled by an instrumental Probit equation, whereas the credit line utilization is modeled by an instrumental Tobit equation.

Our main results are that the two independent variables of the model are endogenous and that their relationship is positive. We use valid instruments to eliminate the econometric bias and affirm that an increased use of the credit line is associated with an increased default probability on a term loan, while an increased default probability is associated with increased use of the line. The estimated marginal effects indicate that a default status on the term loan is associated with an increase of about 59% in credit line utilization. As for the credit line equation, the marginal effects show that for individuals moving from 0% to 100% utilization of their credit line, the probability of default on the loan increases by approximately 46%. Alternatively, at the sample average, a 1% increase in use of the line causes an increase of 46 percentage points in the default probability. We also find that the number of active lines of credit that an individual possesses is an important determinant of the likelihood of default on a term loan. The marginal effect of an additional line of credit for consumers leads to an increase of 26% in the probability of default. We propose a new variable to measure the borrower's ability to pay. This variable complements the credit scoring variable developed by banks and credit agencies, which is more a measure of delinquency. Our results indicate that borrowers with the worst ability-to-pay ratios are the most likely to default on a term loan and to use their credit lines more extensively. These figures are reasonable and reflect the composition of the sample.

8.1 Policy Implications

Basel regulation requires banks to set aside a minimum capital reserve to avoid financial disasters. Such legislation has been adopted by several countries since 1988 and seeks to protect depositors of financial institutions; it is based primarily on portfolio assessment of probabilities of default on bank loans. The Committee allows banks to develop an internal method for computing the capital to keep in reserve. Once their methodology is accepted, banks can use their own estimations of default probabilities, recovery rates, and loss given default. Given the results of our analysis, the amount borrowers draw on the credit line (EAD) is likely to be correlated with the default probability on the term loan. The inclusion of such a correlation should allow banks to manage risk diversification more effectively. Banks could therefore manage the borrower's risks as a portfolio by taking into account the significant dependence across the borrower's various financial obligations. By creating portfolios of loans that allow greater diversification, financial institutions could reduce the minimum capital reserve required by regulators.

8.2 Limits and possible extensions

Our analysis is limited to the data available. Studies on the use of a line of credit suggest that this instrument is positively correlated with changes in the borrower's creditworthiness over time (Agarwal et al., 2006a; Jiménez et al., 2009b; Norden and Weber, 2010). To our knowledge, no study on consumer credit has analyzed the temporal relationship between a revolving credit instrument and a term loan, as we have done statically. It would thus be interesting to test our model in the context of panel data that could address the effects associated with different economic cycles. Such an analysis would provide more information on the dynamics of borrower behavior and could increase the potential correlation between the two financial instruments. Credit card balances could also be incorporated in our framework, if such data were available.

References

Agarwal, Sumit, Brent W. Ambrose, and Chunlin Liu (2006a). "Credit Lines and Credit Utilization." *Journal of Money, Credit, and Banking* 38(1), 1-22.

Agarwal, Sumit, Brent W. Ambrose, Souphala Chomsisengphet, and Chunlin Liu (2006b). "An Empirical Analysis of Home Equity Loan and Line Performance." *Journal of Financial Intermediation* 15, 444-469.

Agarwal, Sumit, Souphala Chomsisengphet, and John C. Driscoll (2004). "Loan Commitments and Private Firms." Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System (U.S.) 2004-27.

Amemiya, Takeshi (1978). "The Estimation of a Simultaneous Equation Generalized Probit Model." *Econometrica* 46(5), 1193-1205.

Amemiya, Takeshi (1979). "The Estimation of a Simultaneous Equation Tobit Model." *International Economic Review* 20(1), 169-181.

Calem, Paul, Matthew Cannon, and Leonard Nakamura (2011). "Credit Cycle and Adverse Selection Effects in Consumer Credit Markets – Evidence from the HELOC Market." Federal Reserve Bank of Philadelphia Working Paper 11-13.

Campbell, John Y. (2006). "Household Finance." *Journal of Finance* 61(4), 1553-1604.

Campbell, John Y., Howell E. Jackson, Brigitte C. Madrian, and Peter Tufano (2010). “The Regulation of Consumer Financial Products: An Introductory Essay with Four Case Studies.” HKS Working Paper RWP10-40.

Dey, Shubhasis (2005). “Lines of Credit and Consumption Smoothing: The Choice between Credit Cards and Home Equity Lines of Credit.” Bank of Canada Working Paper 2005-18.

Dubin, Jeffrey A. and Douglas Rivers (1989). “Selection Bias in Linear Regression, Logit and Probit Models”. *Sociological Methos and Research* 18(2&3), 360-390.

Dunn, Lucia F. and TaeHyung Kim (1999). “An empirical investigation of credit card default.” The Ohio State University Working Paper 99-13.

Finlay, Steven M. (2006). “Predictive models of expenditure and over-indebtedness for assessing the affordability of new consumer credit applications.” *Journal of the Operational Research Society* 57, 655–669.

Heckman, James J. (1979). “Sample Selection Bias as a Specification Error.” *Econometrica* 47(1), 153-162.

Jiménez, Gabriel, Jose A. Lopez, and Jesús Saurina (2009a). “EAD Calibration for Corporate Credit Lines.” Federal Reserve Band of San Francisco Working Paper 2009-02.

Jiménez, Gabriel, Jose A. Lopez, and Jesús Saurina (2009b). “Empirical Analysis of Corporate Credit Lines.” *The Review of Financial Studies* 22(12), 5069-5098.

Lee, Lung-Fei (1992). "Amemiya's Generalized Least Squares and Tests of Overidentification in Simultaneous Equation Models with Qualitative or Limited Dependent Variables." *Econometric Reviews* 11(3), 319-328.

McDonald, John F., and Robert A. Moffitt (1980). "The Uses of Tobit Analysis." *The Review of Economics and Statistics* 62(2), 318-321.

Mester, Loretta J., Leonard I. Nakamura, and Micheline Renault (2007). "Transactions accounts and Loan Monitoring." *The Review of Financial Studies* 20(3), 529-556.

Nelson, Forrest and Lawrence Olson (1978). "Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables." *International Economic Review* 19(3), 695-709.

Newey, Whitney, K. (1987). "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics* 36, 231-250.

Norden, Lars, and Martin Weber (2010). "Credit Line Usage, Checking Account Activity, and Default Risk of Bank Borrowers." *The Review of Financial Studies* 23(10), 3665-3699.

Statistics Canada (2006). "Assets and Debts Held by Family Units, Total Amounts." Available online: http://www40.statcan.gc.ca/l01/cst01/famil109-eng.htm.

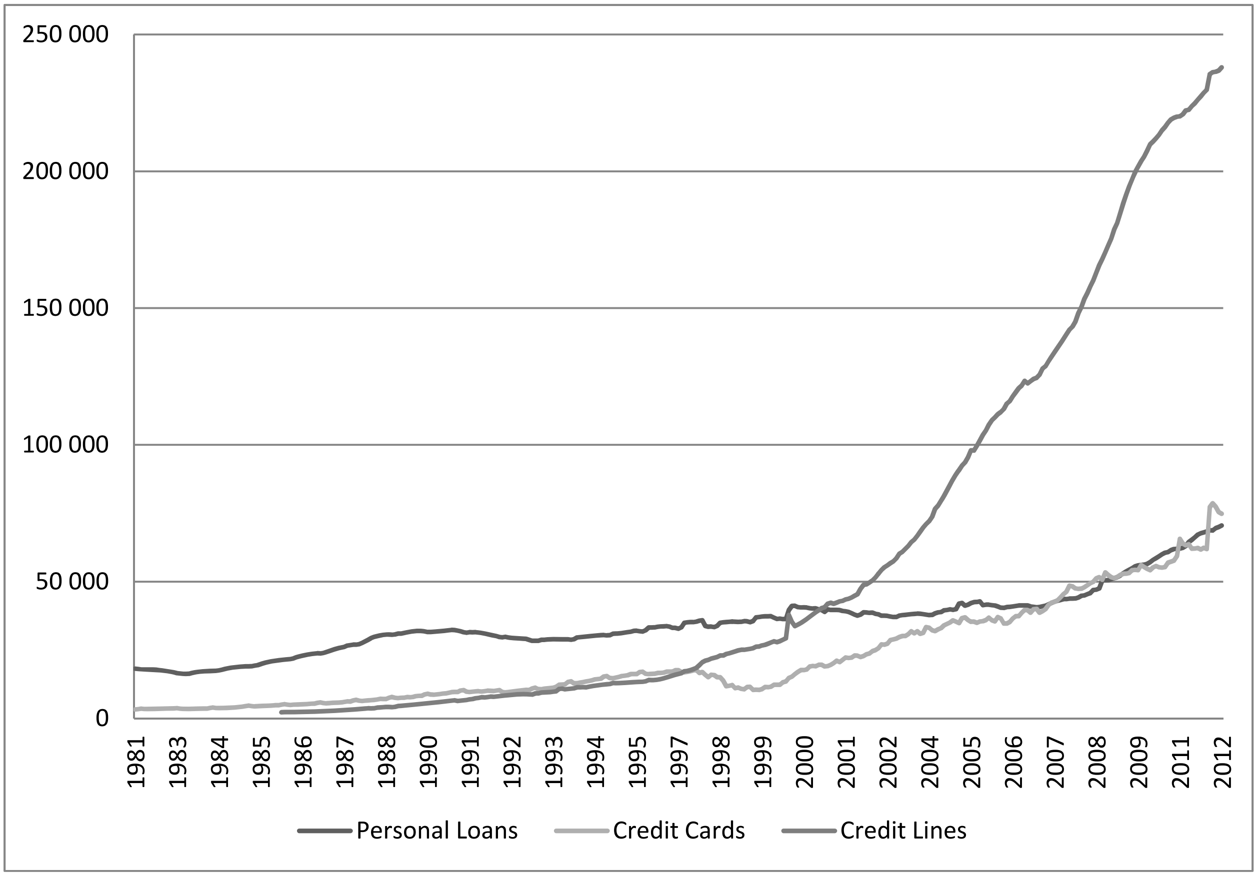
Strahan, Philip E. (1999). "Borrower risk and the price and nonprice terms of bank loans." Federal Reserve Bank of New York Staff Report 90.

Sufi, Amir (2009). “Bank lines of credit in corporate finance: An empirical analysis.” *The Review of Financial Studies* 22(3), 1057-1088.

Tufano, Peter (2009). “Consumer Finance”. *Annual Review of Financial Economics* 1, 227-247.

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Figure 1
Assets of Canadian Chartered Banks ($ million), 1981-2012



The figure plots the monthly average value (in $ million) of personal loans, credit cards, and credit lines for Canadian Chartered Banks from 1981 to 2012. The data come from the series v36867, v36868 and v36869 of CANSIM and do not include information concerning mortgages.

Table 1
List of variables used in the econometric analysis

Variable	Code		Description
Default probability	Def	=	1 if the term loan is in default, 0 otherwise.
Credit line utilization	Util	=	$\frac{\text{Value of the credit line drawn on December 31, 2007}}{\text{Total amount authorized by the bank}}$.
Borrower sex	Sex	=	1 if the borrower is male, 0 otherwise.
Credit score	Score	=	Internal credit rating given to the client by the bank (1 to 8).
Seniority	Sen	=	Seniority of the client at the institution (months).
Employment	Work	=	1 if the borrower is employed, 0 otherwise.
Dependents	Dep	=	Number of borrower's dependents
Borrower age	Age (categories 1 to 6)	=	1 if consistent with the category, 0 otherwise.
Borrower ability to pay	Abil (categories 1 to 5)	=	1 if consistent with the category, 0 otherwise.
Remainder on the loan	Rem	=	Remaining term on the loan, in months.
Loan collateral	Loan_coll	=	1 if the loan is secured by collateral, 0 otherwise.
Additional credit lines	Lines	=	Number of additional active credit lines
Credit line used	Used	=	Amount used on the credit line
Credit line collateral	Line_coll	=	1 if the credit line is secured by collateral, 0 otherwise.
Total debt	Debt	=	Borrower's total debt at the institution or elsewhere

The table presents the variables, their code, and the definition used in the econometric analysis. The first two variables are used as dependent variables in the analysis; the other ones are used as independent variables.

Table 2
Estimates used for predicting expenditures

Variables	Male Sample	Female Sample
Intercept	601.44 (0.0001) ***	208.54 (0.0007) ***
Income	0.6126 (0.0001) **	0.7257 (0.0001) **
Dependents	469.97 (0.0001) ***	274.58 (0.0001) ***
Owner	147.77 (0.0028) **	201.17 (0.0002) ***
Age 1	33.08 (0.7117)	183.08 (0.0717) *
Age 2	64.57 (0.3269)	62.57 (0.3269)
Age 4	-129.91 (0.0323) **	143.76 (0.0213) **
Age 5	-151.74 (0.0421) **	286.32 (0.0007) ***
Age 6	-174.68 (0.2690)	-37.77 (0.7890)

The table presents the estimates used for predicting the expenditures in the ability-to-pay ratio. The first column reports the OLS results for the male sample, while the second column reports the OLS results for the female sample. Values in parentheses represent the p-value of the test statistic for the null hypothesis that the values are zero. *, **, and *** represent significance at the 10%, 5%, and 1% levels respectively.

Table 3
Ability-to-pay ratio

Category	Ability to Pay
Abil 1	Ratio ≤ 0 or Ratio ≥ 1
Abil 2	0.498 ≤ Ratio < 1
Abil 3	0.249 ≤ Ratio < 0.498
Abil 4	0.127 ≤ Ratio < 0.249
Abil 5	0 < Ratio < 0,127

The table presents the segmentation used for the ability-to-pay ratio. This ratio is expressed as the monthly payment on the loan divided by the sum of income and tangible assets minus the predicted expenditures. A negative value for this ratio implies that the expenditures are higher than the income and tangible assets added, while a value of more than one implies a monthly payment higher than the client's ability to pay. Safer borrowers have a small but positive ratio (corresponding to the Abil 5 category).

Table 4
Credit Line Probit
(Heckman selection regression)

	Credit line indicator variable (Probit Regression)
Score	-0. 1459 (0.000) ***
Sen	0.0004 (0.000) ***
Work	0.1640 (0.000) ***
Dep	0.0252 (0.053) **
Age 1	0. 0341 (0.401)
Age 2	0. 2073 (0.000) ***
Age 3	0.2191 (0.000) ***
Age 4	0.1942 (0.000) ***
Age 5	0.1515 (0.000) ***
Sex	0.0845 (0.000) ***
Debt	0.0028 (0.000) ***
Intercept	-0.2271 (0.000) ***
Observations	34,404
LR chi2(14)	2375.98
Prob > chi2	0.000
Pseudo R^2	0.0506

The table reports a Probit selection model for borrowers using a line of credit. The dependent variable is a binary indicator of borrowers using a revolving line of credit. The results are used to construct the inverse Mill ratio, testing for sample bias in the outcome regression (Table 5). Values in parentheses represent the p-value of the test statistic for the null hypothesis that the values are zero. *, **, and *** represent significance at the 10%, 5%, and 1% levels respectively.

Table 5
Heckman outcome regressions

	Credit Line Utilization (2SLS - Heckman)	Credit Line Utilization (2SLS)	Default Probability (2SLS - Heckman)	Default Probability (2SLS)
Util			0.0048 (0.201)	0.0045 (0.229)
Def	5.8350 (0.000) ***	5.9139 (0.000) ***		
Score	0.0341 (0.002) ***	0.0198 (0.019) **	0.0054 (0.000) ***	0.0065 (0.000) ***
Sen	0. 002 (0.001) ***	0. 002 (0.000) ***	0.0000 (0.662)	0.0000 (0.819)
Work	0.0132 (0.553)	0.0308 (0.120)	-0.0003 (0.924)	-0.0017 (0.576)
Dep	0.0129 (0.205)	0.0163 (0.108)	-0.0016 (0.325)	-0.0018 (0.242)
Age 1	-0.0219 (0.528)	-0.0232 (0.507)	0.0001 (0.979)	0.0002 (0.967)
Age 2	-0.0533 (0.095) *	-0.0328 (0.260) *	0.0084 (0.078) *	0.0068 (0.119)
Age 3	-0.0152 (0.638)	-0.0069 (0.813)	0.0092 (0.054) *	0.0075 (0.081) *
Age 4	-0.0465 (0.179)	-0.0294 (0.368)	0.0133 (0.006) ***	0.0119 (0.009) ***
Age 5	-0.0679 (0.043) **	-0.0554 (0.089) *	0.0109 (0.025) **	0.0098 (0.037) **
Sex	-0.0640 (0.000) ***	-0.0560 (0.000) ***	0.0059 (0.004) ***	0.0052 (0.005) ***
Abil 1	0.1117 (0.000) ***	0.1040 (0.000) ***	0.0092 (0.007) ***	0.0098 (0.003) ***
Abil 2	0.0979 (0.000) ***	0.0909 (0.000) ***	0.0057 (0.055) *	0.0062 (0.031) **
Abil 3	0.0938 (0.000) ***	0.0903 (0.000) ***	-0.0019 (0.488)	-0.0016 (0.557)
Abil 4	0.0388 (0.023) **	0.0371 (0.031) **	0.0024 (0.358)	0.0026 (0.332)
Rem			-0.0001 (0.006) ***	-0.0001 (0.004) ***
Loan_coll			-0.0001 (0.941)	-0.0000 (0.992)
Lines			0.0082 (0.000) ***	0.0081 (0.000) ***
Line_coll	0.0736 (0.007) ***	0.0746 (0.007) ***		
Used	-0.0652 (0.000) ***	-0.0655 (0.000) ***		
Intercept	0.2267 (0.371)	0.2267 (0.022) **	-0.0374 (0.006) ***	-0.0273 (0.000) ***
IMR	-0.1423 (0.101)		0.0114 (0.409)	
Observations	14,767	14,767	14,767	14,767
Wald chi2(19)	2354.31	2301.61	222.37	221.70
Prob > chi2	0.000	0.000	0.000	0.0000

The table reports the coefficients of the second-stages of the 2SLS model when including the inverse Mills ratio (IMR). Because the coefficient on the IMR is not statistically significant in either equation, we conclude that there is no selection bias in our sample. Values in parentheses represent the p-value of the test statistic for the null hypothesis that the values are zero. *, **, and *** represent significance at the 10%, 5%, and 1% levels respectively.

Table 6
Dependent variables

Panel A: Loan Status	Observations	Proportion of data
Default	160	1.08%
Non-default	14,607	98.92%
Total	14,767	100%

Panel B: Credit Line Utilization	Observations	Proportion of data
0%	4,433	30.02%
0 % < Credit Line Utilization < 100%	9,693	65.64%
100%	641	04.34%
Total	14,767	100%

Panel A reports the number of observations in the default and non-default states, and their proportion in the data. Panel B reports the number of observations for credit line utilization of 0%, between 0% and 100%, and 100%, and their proportions in the data.

Table 7
Credit line utilization by loan status

	Number of observations (proportion in terms of loan status)	
Loan Status	0% utilization	100% utilization
Default	12 (7.50 %)	62 (38.75 %)
Non-default	4421 (30.27 %)	579 (3.96 %)

The table presents the proportion of observations in the default and non-default states for borrowers using 0% and 100% of their credit line.

Table 8
Explanatory variables

	Default Group			Non-Default Group			Comparison Tests	
	Mean	Median	Standard deviation	Mean	Median	Standard deviation	Mean Comparison (T-test)	Median Comparison (Chi(2))
Util	0.78	0.99	0.39	0.45	0.43	0.33	0.000 ***	0.000 ***
Score	4.55	4	1.82	2.88	3	1.68	0.000 ***	0.000 ***
Age	38.94	37	10.84	41.70	42	11.76	0.003 ***	0.023 **
Sen	160.45	125	126.58	200.12	180	131.44	0.000 ***	0.000 ***
Dep	0.14	0	0.54	0.17	0	0.56	0.578	0.721
Sex	0.72	1	0.45	0.67	1	0.47	0.161	ψ
Work	0.93	1	0.26	0.89	1	0.31	0.201	ψ
Abil	0 .97	0.38	7.15	0.78	0.25	43. 68	0.952	0.001 ***
Rem	17.96	19.13	19.132	23.67	19.23	22.62	0.002 ***	0.097 *
Loan_coll	0.22	0	0.41	0.24	0	0.43	0.5635	0.628
Lines	0.84	1	0.44	0.82	1	0.48	0.4664	0.885

The table presents the mean and median of each explanatory variable by the borrower's loan status, along with a mean and median comparison test. *, **, and *** indicate significance at the 10%, 5%, and 1% levels respectively. ψ means that the test statistic is not available.

Table 9
Estimation results

	(1) Newey's Two-Step Estimation		(2) Maximum Likelihood Estimation		(3) Two-Stage Probit Least Squares	
	Instrumented Tobit	Instrumented Probit	Instrumented Tobit	Instrumented Probit	Instrumented OLS	Instrumented Probit
Util		0.4661 (0.005) ***		0.4631 (0.005) ***		0.2802 (0.039) **
Def	8.5962 (0.000) ***		8.6808 (0.000) ***		0.1382 (0.000) ***	
Score	0.0267 (0.029) **	0.1827 (0.000) ***	0.0261 (0.035) **	0.1817 (0.000) ***	0.0318 (0.000) **	0.1804 (0.000) ***
Sen	0.0003 (0.000) ***	-0.0001 (0.784)	0.0003 (0.000) ***	-0.0001 (0.783)	0.0002 (0.000) ***	-0.0001 (0.839)
Work	0.0429 (0.136)	0.0293 (0.824)	0.0430 (0.138)	0.0296 (0.820)	0.0180 (0.349)	0.0202 (0.873)
Dep	0.0221 (0.130)	-0.0863 (0.188)	0.0223 (0.131)	-0.0858 (0.187)	0.0174 (0.080) *	-0.0866 (0.177)
Age 1	-0.0207 (0.683)	0.2347 (0.256)	-0.0207 (0.686)	0.2329 (0.256)	-0.0527 (0.093) *	0.2257 (0.256)
Age 2	-0.0391 (0.353)	0.3846 (0.018) **	-0.0398 (0.350)	0.3826 (0.018) **	-0.0428 (0.112)	0.3722 (0.017) **
Age 3	0.0157 (0.710)	0.3996 (0.013) **	0.0150 (0.725)	0.3977 (0.013) **	-0.0011 (0.967)	0.3844 (0.013) **
Age 4	-0.0370 (0.435)	0.6228 (0.001) ***	-0.0380 (0.427)	0.6197 (0.001) ***	-0.0408 (0.219)	0.5952 (0.001) ***
Age 5	-0.0851 (0.073) *	0.5740 (0.005) ***	-0.0859 (0.072) *	0.5711 (0.005) ***	-0.0688 (0.044) **	0.5233 (0.009) ***
Sex	-0.0831 (0.000) ***	0.2231 (0.003) ***	-0.0835 (0.000) ***	0.2221 (0.003) ***	-0.0549 (0.000) ***	0.2167 (0.004) ***
Abil1	0.1470 (0.000) ***	0.3829 (0.003) ***	0.1461 (0.000) ***	0.3809 (0.003) ***	0.1033 (0.000) ***	0.4042 (0.001) ***
Abil 2	0.1341 (0.000) ***	0.2918 (0.016) **	0.1335 (0.000) ***	0.2904 (0.016) **	0.0825 (0.000) ***	0.3109 (0.009) ***
Abil 3	0.1373 (0.000) ***	-0.0593 (0.657)	0.1374 (0.000) ***	-0.0587 (0.659)	0.0831 (0.000) ***	-0.0246 (0.850)
Abil 4	0.0592 (0.018) **	0.1612 (0.191)	0.0589 (0.019) **	0.1605 (0.190)	0.0268 (0.153)	0.1788 (0.136)
Rem		-0.0058 (0.004) ***		-0.0058 (0.004) ***		-0.0060 (0.002) ***
Loan_coll		-0.0376 (0.646)		-0.0381 (0.640)		-0.0243 (0.759)
Lines		0.2652 (0.000) ***		0.2639 (0.000) ***		0.2639 (0.000) ***
Used	0.0861 (0.000) ***		0.0860 (0.000) ***		0.0648 (0.000) ***	
Line_coll	-0.1024 (0.011) **		-0.1028 (0.011) **		-0.0579 (0.011) ***	
Intercept	-0.1071 (0.106)	-4.1698 (0.000) ***	-0.1050 (0.116)	-4.1468 (0.000) ***	0.5046 (0.000) ***	-3.9745 (0.000) ***
Observations	14,767	14,767	14,767	14,767	14,767	14,767
Pseudo R^2	-	-	-	-	0.4364	0.1135
Wald chi2(df)	1946.21	176.16	1914.82	173.45	-	-
P-value	0.000	0.000	0.000	0.000	0.000	0.000

The table presents the three sets of instrumented results estimated for the model. The first set of results is estimated with Newey's Two-Step Efficient Estimator (Newey, 1987), a limited information procedure. Standard errors for this estimator are based on Amemiya's (1978, 1979) derivations of the efficient variance-covariance matrices. The second set of results is estimated from a full information maximum likelihood procedure. For robustness, the last set of results is a joint two-step estimation of OLS and Probit models, known as Two-Step Probit Least Squares (2SPLS). All procedures are instrumented for the endogenous variables of the system and all results reported are for the second stage estimation. Values in parentheses represent the p-value of the test statistic for the null hypothesis that the values are zero. *, **, and *** represent significance at the 10%, 5%, and 1% levels respectively.

Table 10
Marginal effect coefficients derived from the MLE

	Credit line utilization (Instrumented Tobit)	Default Probability (Instrumented Probit)
Util		0.4631 (0.005) ***
Def	0.5887 (0.000) ***	
Score	0.0104 (0.035) **	0.1817 (0.000) ***
Sen	0.0001 (0.000) ***	-0.0001 (0.783)
Work	0.0171 (0.137)	0.0296 (0.820)
Dep	0.0089 (0.131)	-0.0858 (0.187)
Age 1	-0.0082 (0.686)	0.2329 (0.256)
Age 2	-0.0158 (0.349)	0.3826 (0.018) **
Age 3	0.0060 (0.725)	0.3977 (0.013) **
Age 4	-0.0151 (0.426)	0.6197 (0.001) ***
Age 5	-0.0340 (0.071) *	0.5711 (0.005) ***
Sex	-0.0333 (0.000) ***	0.2221 (0.003) ***
Abil 1	0.0583 (0.000) ***	0.3809 (0.003) ***
Abil 2	0.0533 (0.010) ***	0.2904 (0.016) **
Abil 3	0.0548 (0.000) ***	-0.0587 (0.659)
Abil 4	0.0235 (0.002) **	0.1605 (0.190)
Rem		-0.0058 (0.004) ***
Loan_coll		-0.0381 (0.640)
Lines		0.2639 (0.000) ***
Used	0.0342 (0.000) ***	
Line_coll	-0.0406 (0.010) ***	
Amemiya-Lee-Newey overid. test	0.289 (0.8654)	1.165 (0.2804)
Wald test of exogeneity	108.35 (0.000) ***	4.45 (0.0349) **
Observations	14,767	14,767

The table reports the marginal effects of the independent variables derived from the Maximum Likelihood Estimation. This estimation allows a test of the exogeneity of the default probability and the credit line utilization variables by a Wald test of exogeneity. The null hypothesis of the test is the exogeneity of the variable. Newey's Two-Step Efficient Estimation allows a validation of the chosen instruments for each endogenous variable by an Amemiya-Lee-Newey overidentification test (Lee, 1992). The null hypothesis of the test is the validity of the instruments. Marginal effects reported for the Probit model are based on linear prediction. Marginal effects reported for the Tobit model are presented for the latent variable. Values in parentheses represent the p-value of the test statistic for the null hypothesis that the values are zero. *, **, and *** represent significance at the 10%, 5%, and 1% levels respectively.

Table 11
McDonald and Moffitt (1980) decomposition

	Credit line utilization (Tobit)	
	Intensive Margin	Extensive Margin
Def	0.1627 (0.000) ***	-0.1910 (0.000) ***
Score	0.0009 (0.035) **	0.0007 (0.036) **
Sen	0.0000 (0.000) ***	0.0000 (0.000) ***
Work	0.0015 (0.138)	0.0012 (0.184)
Dep	0.0008 (0.131)	0.0006 (0.133)
Age 1	-0.0007 (0.686)	-0.0006 (0.704)
Age 2	-0.0014 (0.350)	-0.0011 (0.383)
Age 3	0.0005 (0.725)	0.0004 (0.718)
Age 4	-0.0013 (0.427)	-0.0010 (0.454)
Age 5	-0.0030 (0.072) *	-0.0027 (0.134)
Sex	-0.0029 (0.000) ***	-0.0019 (0.000) ***
Abil 1	0.0052 (0.000) ***	0.0023 (0.000) ***
Abil 2	0.0047 (0.000) ***	0.0024 (0.000) ***
Abil 3	0.0048 (0.000) ***	0.0024 (0.000) ***
Abil 4	0.0021 (0.019) **	0.0013 (0.007) ***
Used	0.0030 (0.000) ***	0.0022 (0.000) ***
Line_coll	-0.0036 (0.011) **	-0.0035 (0.044) **
Observations	14,767	14,767

The table reports the McDonald and Moffitt (1980) decomposition for the Tobit equation of credit line utilization based on Equation (5). The intensive margin reports the marginal probability of utilization of the credit line for an individual who already has a utilization rate between 0% and 100%. The extensive margin reports the marginal utilization of the credit line for an individual who has a 0% or 100% utilization of the credit line. Values in parentheses represent the p-value of the test statistic for the null hypothesis that the values are zero. *, **, and *** represent significance at the 10%, 5%, and 1% levels respectively.