Consumer Credit: Learning Your Customer's Default Risk from What (S)he Buys

Annette Vissing-Jorgensen
Kellogg School of Management, Northwestern University, NBER and CEPR

November 16, 2009

Abstract

Using a novel panel data set for half a million customers of a large Mexican retail chain I study determinants of consumer credit default. I document that information about which products a customer buys provides substantial information about potential default losses on a given loan. Differences in default losses across product categories are robust to controlling for characteristics of the loan contract, demographics and more standard measures of credit risk and do not diminish substantially with how long the borrower has been a customer. The differential loss rates across product categories are driven mainly by which types of individuals buy particular products, as opposed to being product-specific features. High loss products tend to be luxuries and tend to be purchased by individuals who consume abnormally large fractions of luxuries given their income. I discuss how differences across consumers in their desire for indulgence or their degree of self-control may explain why loans to people who consume more luxuries incur higher loss rates. I propose that providers of consumer credit could benefit from adjusting credit terms (down-payment requirements, interest rates, or credit limits) as a function of product mix purchased to date, and thus that product mix should be an important component of credit scoring.

I. Introduction

U.S. households have increasingly large amounts of debt. From 1945 to the second quarter of 2009, household liabilities grew from 20% of disposable personal income to 129%. The increase in liabilities has been driven by both increasing mortgage debt (as a percent of real estate assets and as a percent of disposable personal income) and an increase in consumer credit. At the end of the second quarter of 2009, home mortgages stood at $10.4 Trillion, with consumer credit adding another $2.5 Trillion to household liabilities.

A central issue in the context of household debt is default. Compared to the large literature on determinants of household portfolio choice, relatively little is known about what drives households' choices of debt levels and their decision of whether and when to default. Industry models for predicting default emphasizes past borrowing and repayment behavior. For example, in the calculation of an individual's FICO score (a commonly used measure of credit risk in the U.S.), a weight of 35% is given to on time payment of past debt, 30% weight is given to the current amount of debt of various types, how many accounts the individual has, and how large the debt is relative to the total available credit, 15% weight is given to the length of time of credit history, 10% to the number of new accounts and recent requests for credit, while 10% is given to the mix of credit (credit cards, installment loans, finance company loans, and mortgages) used in the past (Fair Isaac Corporation (2005)). FICO scores do have predictive power for default (e.g. Keys, Mukherjee, Seru and Vig (2010) in the context of mortgage delinquencies). However, from the perspective of understanding the underlying economic drivers of debt and default, predicting default based on past repayment behavior is not informative. Furthermore, models predicting default tend to have a modest statistical fit. For example, Gross and Souleles (2002) find a pseudo-R^2 of about 0.14 in a probit model predicting credit card delinquency using time effects, account age, measures of account risk (including credit scores), and local economic conditions as explanatory variables.

This paper seeks to improve our understanding of household default from both an economic and a statistical perspective. Using a new proprietory data set from one of the largest retail chains in Mexico I document that information about which products a customer buys provides substantial information about potential default losses on a given loan. The data set is large both in terms of the number of borrowers covered -- about half a million -- and in its panel dimension, with monthly data available from January 2005 to August 2009. The unique feature of the data which enables me to study the link between which goods are purchased and subsequent loan losses for the lender is that each good purchased at this retailer has its own loan associated with it. For example, a customer may first take out a loan to buy a washing machine and then later take out a separate loan to buy new tires for her car.

I find that lender loss rates (measured as the ratio of the amount not repaid to the size of the loan), are dramatically higher for certain types of products than others. Lender losses are low (below 12%) on loans for appliances, kitchen equipment and furniture, while they are about 21% on loans for electronics (cell phones, stereos, TVs etc.), and almost 40% on loans for jewelry purchases. These differences in default losses across product categories are robust to controlling for characteristics of the loan contract (e.g. the size of the loan and interest rate on the loan), demographics, and more standard measures of credit risk based on past repayment behavior, and the differences do not diminish substantially with how long the borrower has been a customer. From a statistical perspective this implies that which products people borrow money to buy is a useful additional predictor of subsequent default, above and beyond known predictors documented in past work.

To begin assessing the economic forces underlying differential lender losses across product categories, I estimate models of loss rates that include customer fixed effects. With these fixed effects included differences in loss rates across product categories are economically small. This indicates that differential loss rates across product categories are driven mainly by which types of individuals buy particular products, as opposed to being product-specific features. In other words, customers who tend to buy electronics generate high lender loss rates both when they buy electronics and when they buy other products.

This raises the question of why people with preferences for particular goods on average are worse risks. I find that high loss products tend to be luxuries and that they tend to be purchased by individuals who (in their spending at the retail chain analyzed) consume abnormally large fractions of luxuries given their income. A one standard deviation increase in the fraction spent on luxuries (controlling for income), increases the predicted lender loss rate by about 3.4 percentage points. I confirm that the link between preferences for luxuries and getting into repayment difficulties is not specific to the particular retail chain studied or to Mexico. Using U.S. data from the 2003-2007 Consumer Expenditure Survey I find that households who consume abnormally large fractions of luxuries given their overall consumption have larger amounts of consumer credit (an extra $300 for a one standard deviation increase in the fraction spent on luxuries) and incur larger amounts of finance charges on their consumer credit.

While many aspects of budget constraints or preferences (income risk, discount rates, the elasticity of intertemporal substitution, or risk aversion) could happen to be correlated with people's preferences for luxuries, I discuss whether the existing literature on consumer choice includes theories that have predictions about both which products particular people tend to buy and why particular people would have higher default rates on consumer credit. The marketing literature (e.g. Dhar and Wertenbroch (2000)) distinguishes between hedonic goods which provide more experiential consumption, fun, pleasure, and excitement, and utilitarian goods which are primarily functional. Utilitarian goods tend to be necessities while hedonic goods tend to be luxuries. Furthermore, Hoch and Loewenstein (1991) model consumption outcomes as the result of an inner struggle between a desire for immediate gratification (i.e. hedonic pleasure, or indulgence) and self-control. I argue that high spending relative to available resources (and thus high default risk) and spending an abnormally high fraction on luxuries may be two dimensions of having a high preference for indulgence or low self-control to overcome impulses to indulge. I discuss possible field surveys and experiments at the Mexican retailer that could further shed light on this interpretation of the evidence.

The paper's findings have immediate implications for lenders. The fact that people who tend to buy certain products generate larger loss rates implies that providers of consumer credit could benefit from adjusting credit terms (down-payment requirements, interest rates, or credit limits) as a function of product mix purchased to date, and thus that product mix should be an important component in credit scoring. This requires lenders to estimate the borrower fixed effects in ``real time", i.e. using only information available up to the data of purchase. I propose that lenders use the average (product level) default rate of products purchased by a particular customer to date as an additional indicator of the customer's creditworthiness, above and beyond standard predictors of default such as credit scores and demographics.

Since an increasing fraction of consumer debt is securitized, my findings also suggest that information about which products (and in turn which types of buyers) a given security was issued to finance could help price the security more accurately. More broadly, understanding what drives heterogeneity in borrowing and default behavior across households in the market for consumer credit is likely to be informative for other loan markets and for understanding consumption and savings behavior more generally. For example, the same factors that make some individuals buy more luxurious consumer goods than would be predicted by their income may be relevant for understanding the behavior of home buyers some of whom may buy more luxuriously houses than suggested by their income, with corresponding implications for mortgage default. Furthermore, if substantial preference heterogeneity can be documented within a sample of households who all have consumer debt, one would expect an even larger degree of preference heterogeneity in the full set of households, with correspondingly broader implications for explaining heterogeneity in household net worth.

The paper is organized as follows. Section II describes the data set and how lending works at this retailer. Section III documents differences in loss rates across product categories. Section IV documents that these differences are driven mainly by high-risk borrowers disproportionately buying certain products as opposed to being product-specific characteristics. It also discusses how lenders can construct an estimate of a given borrower's type using real-time data. Following this, section V turns to the link between borrowers' risk and their preference for luxury goods as well as to what more fundamental heterogeneity may drive this link. Section VI concludes.

II. The Data Set and the Mechanics of the Loan Process

A. The Basics of the Data Set

The data sets consists of information about 499,906 new customers who purchased one or more products on credit at one of the largest Mexican retail chains between January 2005 and December 2006. During this time period, this set of customers made a total of 1,364,864 credit-financed purchases. The payment history of these purchases is followed up to August 2009.

The retail chain which made the data available makes about 90% of its sales on store credit, with the remaining sales paid in cash or using credit or debit cards. The chain was founded several decades ago is now represented in all 32 Mexican states and has millions of customers. During the 2005-2006 period the chain was not represented in a few states. The purchases in my sample are made across 220 different stores.

The company's target customers are middle and lower income households. 88% of customers in my sample have monthly household incomes below 16,800 pesos ($1,268). 52% have monthly household incomes below 4,200 pesos ($317). For comparison, in the Mexican population as a whole, 85% have monthly household incomes below 16,800 pesos, while 26% have monthly household incomes below 4,200 pesos (calculated using ENIGH 2005, a national household survey conducted by the Mexican government). A large part of the company's success is attributable to its ability to sell products on credit to this segment of households, many of whom have no other sources of credit. One of every five employees work in credit supervision.

The data set provided by the company contains monthly information of four types. First, the company collects and updates customer demographics, specifically age, gender, marital status, household income, education, home ownership, years at current address, and household size. Second, information is provided about any movements in the customer's accounts. Movements include new purchases made, payments on past purchases, assignment of additional interest (due to late payments), merchandise returns, as well as records of any payment renegotiation plans entered into between a customer and the firm. The data set covers purchases made on credit only. For each new purchase, the data set contains information about the store at which the purchase took place, the amount of the purchase, the size of the down-payment, the interest rate on the loan, the term of the loan, and the type of product purchased. Third, monthly data are available for the customer's account balances, the customer's track record of repaying loans, and the customer's credit limit. Fourth, the data contains information about ``lost loans'', meaning loans on which the company has given up collecting any further payments. For such accounts, records are kept on the date of purchase, the date the account was declared lost, and the amount of the loss to the company. The average time between date of purchase and date the account was declared lost is just over two years (with a 1st percentile of 730 days and a 99th percentile at 761 days). This reflects the fact that loan terms are 12 or 18 months and it takes the company a while to determine whether any further payments can be collected on a given loan. The two year lag between a purchase and the typical date of a loan being declared lost motivates my focus on purchases made in 2005 and 2006. Since the sample runs to August 2009, this provides sufficient panel dimension to follow the outcome of each loan.

Of particular importance for my analysis is the information about the type of good purchased and the way loans are made. For the purchases made in 2005 and 2006 I have for each purchase a basic product description such as ``DVD player'', ``lamp'', or ``washing machine''. This product description refers to the largest item purchased on a given visit to the store. For a separate sample (not overlapping with the main sample described above) covering purchases made between December 2008 and August 2009, both the basic product description and a product category assigned by the company is also available. I create a mapping between basic product descriptions and product categories in this sample and use it to assign a product category to each purchase in the 2005-2006 sample.[1] The product category is based on the company categorizing products into 9 different departments, with each of the department further sub-divided into classes. Purchases in the 2005-2006 sample fall into 124 product categories. Some of the product categories account for very small fractions of overall purchases. Within each of the nine departments, I therefore group some of the classes together and work with a total of 32 product categories.

What enables me to study the relation between what a customer purchases and default is the following unique feature of the lending process. Rather than having one revolving credit account at the company to which various purchases could be charged, the company issues a separate loan for each purchase. For example, suppose a customer buys a refrigerator and then comes back a few weeks later and buys an armchair. The company will make one loan for the refrigerator and another for the armchair and I am able to follow the repayment (or lack of repayment) of each of these two loans. Clothing and cell phone minutes are an exception to this principle since these are charged to a revolving credit account much like a U.S. credit card. This makes it difficult to compare losses on clothing and cell phone minutes to losses on other product and I therefore leave out clothing and cell phone minutes from the majority of the analysis. The counts of individuals and purchases stated above exclude purchases of clothing and cell phone minutes.

The first column of Table 1 shows the distribution of sales (by peso value) across the nine departments, along with the fraction of sales constituted by clothes and cell phone minutes.

---

[1] To create this mapping, I calculate (in the December 2008-August 2009 sample) the most common product category for each basic product description. For most basic product descriptions (73%) only one product category is used and for the remaining product descriptions the same product category is assigned by the company in the vast majority of the occurrences of a given basic product description. This implies that one with a high degree of accuracy can use this mapping to define product categories in the 2005-2006 sample by assigning the most common product category for that basic product description to all sales with a given basic product description.

Appendix Table 1 provides a more detailed breakdown across the 32 product categories, clothes and cell phone minutes. About 41% of sales are electronics, 31% are clothes, 9% are appliances, 8% are various types of furniture, 6% are kids gear and toys or auto parts (of which kids items constitute more than half), 2% are kitchen equipment, with the remaining 3% constituted by watches, jewelry, eye glasses, and cell phone minutes. The range of products sold is thus very diverse, though one should keep in mind that consumer goods only represent a fraction of overall spending, with food and housing likely constituting a larger fraction for most households.

B. The Mechanics of the Loan Process

The lending process for a particular purchase starts with the customer deciding on which item(s) he or she would like to buy. A sales person then accompanies the client to the credit desk. For new clients a host of information is then collected, including the clients name, identity documentation, address, demographics, employer and income. If the client does not work, the spouse/partner's employment information is collected instead. The credit desk then verifies the clients identity, home address and work information by phone via a call center. This takes only about 5 minutes during which the client watches an informational video.

The credit desk then proposes a minimum down-payment. The rules for down-payments have changed over time, but the latest rules are as follows.

% of customer's authorized credit	Type of client				
	A	B	N	C	D
From 0 to 100	0	20	10	30	30
101 to 150	10	20	20	30	40
151 to 200	20	30	30	40	50
201 to 300	20	40	40	50	60
301 to 400	30	50	50	60	70

Each cell states the required minimum down-payment as a percent of the cost of the item as a function of the cost of the item relative to the customer's authorized credit (credit limit) and the company's internal credit score for the client (the down-payment numbers are not binding in that the client, or the sales person on the client's behalf, can bargain with the credit desk to reduce the down-payment). A new customer's authorized credit (credit limit) is 25% of the customer's annual income. Subsequent limits are updated based on the client's payment history. A customer can borrow more than the limit but will then be required to pay a larger down-payment as laid out in the table above. The customer's internal credit score is calculated based on the customer's repayment efficiency to date. Repayment efficiency is calculated as the sum of actual payments divided by the sum of payments due since the customer first started borrowing at the company. New customers are assigned a credit score of ``N" meaning that they have no repayment history.

The monthly payment on a loan is calculated as:

(1) $$\text{Monthly payment} = \text{Loan amount} * (1+r) / \text{Loan term}$$

where r is the interest rate on the loan. The implied annual percentage rate on the loan is higher than r. For example, an interest rate of 24% on a 12-month loan leads to the same monthly payment using the above formula as an annual percentage rate of 41.6% with monthly compounding would. Interest rates are surprisingly homogeneous across borrowers. Notably, they do not depend on the borrower's credit score, the down-payment, or the size of the purchase. The only variation in interest rates (at a given point in time) is that they are higher for cell phones than for other product categories, higher for 18-month loans than 12-month loans, and higher for cities considered high risk. The latest schedule of interest rates is:

City type:	Zone 1 (low risk)	Zone 2 (high risk)
Furniture/household item (12 month loan)	24%	30%
Furniture/household item (18 month loan)	36%	45%
Cell phone (12 month loan)	32%	38%
Cell phone (18 month loan)	44%	38%

Once the loan is granted, monthly bills are delivered by hand and explained in person. Additional visits are paid if the customer is overdue on his/her payments. If a customer misses payments, there are three possible outcomes: (1) The customer renegotiates to pay over a longer period, (2) the customer agrees to return the product to the store, or (3) the customer never pays what is owed and the firm declares the amount owed on the particular loan at this point a loss.

III. Differences in Loan Loss Rates Across Product Categories

A. Defining the Loss on a Loan

When a customer do not make the full set of payments on a particular loan, the amount declared lost by the company is given by:

(2) Loss=Loan*(1+r)-Payments

which can be decomposed into how much of the principal is not repaid and how much of the interest is not repaid:

(3) Loss=[Loan-min(Loan,Payments)]+[Loan*r-(Payments-min(Loan,Payments)]

=Principal loss+Interest loss.

I define the loss rate as the loss divided by the size of the loan. One can approximate the company's realized return over the term of the loan as:

(4) 1+realized return=Payments/Loan=(Loan*(1+r)-Loss)/Loan

=(1+r)-[Principal loss/Loan]-[Interest loss/Loan].

If all payments were due at the end of the payment term this would be the exact realized return. Since in practice payments are due monthly, the realized return for the company accounting for the fact that it can re-lend payments received before the end of the term, will be higher. On the other hand, if some of the payments received are made after the term of the loan these should be discounted back to the loan maturity date to calculate the exact realized return over the term of the loan. Ignoring this issue in the above approximation will tend to overstate the realized return.[2]

B. Main Result

Table 1 documents the main descriptive result of the paper: Dramatic differences in the company's loss rates and realized return on loans across products. Column (4) shows the lender's loss rate for each of the nine departments of products. The rates are calculated at the department level (as opposed to being an average across purchases in the category) in order to account for any potential correlation between losses and purchase size. Four of the departments -- kitchen equipment, the two types of furniture (mattresses, dining set and other furniture; living room and

[2] A related issue is that the return given in equation 4 is for the period of the term, as opposed to being on an annual basis. For the nine department that I focus my analysis on the loan term is either 12 or 18 months, but more than 99% of loans are 12 month loans. Dropping the 18 month loans from the analysis has very limited impact on any of the results.

bedroom furniture), and appliances -- have substantially lower default rates than the others. For, these four departments, loss rates are between 11% and 12%. In contrast, loss rates for the category kids gear and toys, auto parts and bikes, and for watches, and eye glasses, range between 15% and 17%. Electronics, which constitute a large fraction of both sales and loans, have default rates above 20% and loans given to finance jewelry purchases have default rates of almost 40%. Appendix Table 1 shows the loss rates for the 32 more detailed product categories. Products with a given department tend to have similar loss rates. Notice, however, that cell phones have a loss rate substantially above that of other electronics.

From the description of the loan process in section II.B it is clear that the differences in loss rates do not fully translate into corresponding differences in interest rates across products. The company does charge higher interest rates for cell phone loans, but otherwise charges the same interest rate for all products. Column (7) of Table 1 and Appendix Table 1 show the average interest rate charged for each department and for each of the 32 product categories. Rates are around 25% for each product category for the 2005-2006 sample analyzed here, with the exception of an average rate around 30% for cell phones. For the categories other than cell phones the small differences across categories are driven by slight differences in the timing of purchases (since interest rates change over time) and the location of purchases (across high and low risk cities). Since a higher interest rate mechanically will lead to a higher interest loss rate for identical payments by the customer, column (5) and (6) of Table 1 and Appendix Table 1 decompose the loss rates into principal loss rates and interest loss rates. It is clear that even the principal loss rate (which is not mechanically affected by the loan interest rate) is higher for cell phones and thus for electronics than for all the other products aside from jewelry.

Column (8) of Table 1 and Appendix Table 1 summarizes the impact of loss rates and interest rates on lender profits by showing the realized return earned by the lender for each type of product. The lender return is negative for loans given to finance jewelry purchases, due to the large loss rates for this category. For the other categories, the lender return is substantially positive since interest rates are far above loss rates, with patterns across departments and across product categories driven by the patterns in loss rates. It is important to emphasize that the lender return on loans calculated here do not account for the large expenses the company incurs as a result of employing thousands of staff to manage the loan process. Accounting for differences in these expenses across product categories would likely increase differences in lender returns across product categories since additional costs from extra home visits and loan renegotiation are incurred when a customer starts missing payments on the loan.

These findings indicate that the company could likely benefit from conditioning loan terms -- interest rates, down-payments, or credit limits -- much more on product type than is currently done. Preliminary discussions with company management reveal that the company to some extent is aware of loss differences across product categories and is open to conducting field experiments of the impact of further conditioning of loan terms on product type. In fact, in early 2009, the firm increased the down-payment requirements for new clients from 10% to 20% for the following products: Cell phones, stereos, video games, iPods, computers, laptops, and jewelry. The profit impact of this is not yet known.

A central issue in designing the optimal loan terms is whether the differences in loss rates and lender returns across products are truly product specific effects or whether they instead are driven by a tendency for high risk borrowers to buy certain products.[3] In the former case, setting different loan terms for each product category is likely to be optimal while that approach is unlikely to be optimal in the latter case. For example, if borrowers who buy a lot of electronics tend to default on both electronics and other purchases, the company should try to elicit a given customer's risk type from his/her purchase patterns and then condition loan terms for all products purchased by the consumer on this information. I return to these issues in section IV.

An additional issue in designing the optimal loan terms is whether the patterns of differential loss rates and lender returns across product categories is specific to new customers or whether they diminish with time as customer. I document next that while loss rates decline with time as a customer (due to the company learning about each customer's risk type and adjusting down-payment terms and credit limits accordingly), large differences in loss rates across product categories remain even for more seasoned borrowers. Figure 1 sorts purchases by how long the customer has been a customer at the time of purchase, measured in months. The average loss rate across purchases made by customers who had been customers a given number of months is plotted against the time as customer. A standard identification issue arises in interpreting this relation. Customers who have been with the company for more months will tend to have become customers earlier (in calendar time) and will tend to be making purchases in later months (also in

[3] Economically, truly product-specific effects could arise from some products being easier for the borrower to resell in the used goods market, or from some product depreciating faster in value over time. Both of these effects would make strategic default more attractive for the borrower.

calendar time), since the sample of purchases studied are for a fixed time sample, 2005-2006. Panel A therefore shows separate lines for groups of purchases made by customers who all became customers in a given narrower time period, while Panel B shows separate lines for groups of purchases made in a given narrower time period. Panel A thus show the relation between loss rates and time as a customer controlling for cohort effects, while Panel B shows the relation controlling for time (date of purchase) effects. In both approaches, there is a strong negative relation between loss rates and time of purchase, with loss rates of 20-25% on loans made for purchases by new customers and loss rates of 10-15% on loans made by customers who have been with the company more than a year. Cohort and time effects are modest in comparison to the effect of time as customer.

Importantly, though loss rates are lower for more seasoned borrowers, Table 2 shows that the differences across product categories remain about as large in relative terms for seasoned as for new borrowers. For example, the default rate on electronics for loans to customers in their first month with the company is about 1.8 times the default rate on kitchen equipment, while the ratio of the default rates for these two categories is about 2.0 for loans to customers who have been with the company between 18 and 24 months at the time of purchase. Similarly, the default rate on jewelry is 3.9 times that for appliances for loans to customers in their first month as borrowers, and this ratio remains as high as 2.8 even for loans to customers who have been with the company between 18 and 24 months. The fact that differences in loss rates across products are large for both new and more seasoned borrowers suggests that conditioning lending terms on purchase patterns is relevant across the customer population.

While the relation between purchase patterns and lender losses has not (to my knowledge) been emphasized in past work, several papers have documented a relation between default rates and past default rates (as captured by credit scores), loan terms, and demographics. To ensure that the differences in lender losses across product categories remain once these known predictors of default are controlled for, and to investigate how much additional predictive power is gained by considering product categories, I turn next to statistical models for predicting loss rates.

C. The Predictive Power of Product Mix for Losses, Controlling for Standard Default Predictors

For a given loan, the loss rate (denote it by y) is either zero or positive. I am interested in how E(y|X) depends on a set of predictors X. One possible approach to modeling this relation would be to estimate a Tobit model. In that setup, E(y|X) would be non-linear in X (see Wooldridge (2002), equation (16.14)). I instead take a simpler approach and assume that E(y|X) is approximately linear in X over the relevant range of variation in X and proceed to estimate linear regression models by OLS. The reason for this simplification is that I later turn to estimating models with customer fixed effects. While one can estimate the regression coefficients (β's) in a Tobit model with fixed effects by transforming the model in a way that eliminates the fixed effects (see Honore (1992)), dE(y|X)/dX (with fixed effects included in the set of X-variables) remains a function of the fixed effects and no unbiased estimator of the fixed effects exists. For comparability of results I therefore proceed to estimate linear regression models both for the cases without individual fixed effects and for the cases with individual fixed effects.

Table 3 predicts loss rates using time as customer dummies (to account for the strong negative relation between loss rates and time as a customer documented in Figure 1), transaction characteristics (including loan terms), measures of borrower credit risk, demographics, and store fixed effects. Table 4 adds product dummies. It is known that transactions characteristics are related to lender losses. For example, in a study focusing on auto loans to subprime borrowers, Adams, Einav and Levin (2009) find that default rates are higher on larger loans. This is consistent with both models of adverse selection (where high-risk individuals self-select into larger loans) and models of moral hazard (where a larger loan increases the likelihood of default either via strategic default or simply lack of affordability of the payments). Adams et al. (2009) argue, however, that one can include the excess of down-payments above a statistically predicted value as a proxy for an individual's risk type (since low risk borrowers use this to signal their type to affect the interest rate), in which case the effect of loan size measures only the moral hazard effect. This effect is fund to be positive. At the firm analyzed here, the interest rate on a loan does not depend on the down payment made thus preventing signaling. Therefore, while I include both loan amount and down payment/purchase price as regressors, the effect of loan amount will pick up both adverse selection and moral hazard and any impact of the down payment variable does not pick up signaling but could instead pick up differences across consumers in how long they have been planning for the purchase as well as a causal effect of making certain customers save up more before purchase in order to both ensure lower losses if the customer never makes any further payments and ensure lower monthly payments that are more affordable for the borrower. I also include the interest rate on the loan as a regressor, with an expected positive effect to the extent that the lender has information about likely losses and this is reflected in the interest rate. Such interest variation is used by this particular company in the interest variation across calendar time and term of loan, and across cities perceived by the company as low risk or high risk. Finally, I include the term of the loan (in months) as a transactions characteristic. The predicted sign for this variable is unclear. If borrowers who can afford lower monthly payments take out longer term loans, they may not generate systematically higher or lower loss rates than borrowers who take out a shorter term loan.

As a measure of credit risk I include the credit scores used by the company. The company assigns each borrower a credit score based on past repayment efficiency. As noted above the company calculates repayment efficiency as the sum of actual payments divided by the sum of payments due since the customer first started borrowing at the company. The repayment efficiency calculation is based on both loans made for products in the nine departments as well as for clothing and cell phone minutes. Customers with repayment efficiencies above 75% are assigned a credit score of A, repayment efficiencies between 50% and 75% imply a credit score of B, repayment efficiencies between 25% and 50% imply a credit score of C, while repayment efficiencies below 25% imply a credit score of D. New customers are assigned a credit score of N. In addition to the credit score, I include the underlying repayment efficiencies for both the main account (I use main account to refer to loans for products in any of the nine departments I study, but remember that loans are made at the purchase level not the account level) and the clothing account. I include the number of purchases made to date as an additional risk control. In my set of measures of credit risk I furthermore include variables that would matter for credit scores in the U.S. FICO score system and for which I have data from the company studied here but not from any other credit the borrower may have obtained elsewhere. These variables are as of the end of the month prior to the month of the loan analyzed, or as of the date of the first loan for customers getting their first loan, to make sure they are observable at the time of the loan. The variables included (with separate variables available for the main account and the clothing account when available) are: Credit limit, current amount of account balance, current amount of late balances, amount of moratory interest accumulated (and not paid) to date due to late or missing payments, and maximum credit level obtained in the past.

The set of demographics collected by the firm includes age, gender, marital status, a categorical variable for income, education, living situation (home owner, renter, living with family, or living as guest with someone else), years at current home address, and three measures of household size (number of people living in the customer's house, number of people living in the customer's house who work, and number of people who are economically dependent on the client). For a given loan I use the demographics as of the end of the prior month, or as of the date of the first loan for customers getting their first loan. The only exception is that the household size variables are only available as of December 2008. Results are largely unaffected by excluding the household size measures. The company restricts credit for minors, so I include a dummy for being a minor (age<21 for men, age<18 for women) in addition to age in the regressions.

The results in Table 3 show that both transactions characteristics, measures of credit risk, and demographics have explanatory power for predicting the loss rate on a given loan. The statistical significance of each variable is of little interest in the current setting, since the large sample (over 1.3 million loans) implies that most variables are significant at the 1% level (Table 3 uses 3, 2 and 1 asterisks to indicate significance at the 1, 5, and 10% levels). More interesting are the signs and magnitudes of the effects and the R^2 of the regressions. For reference, Appendix Table 2 shows the summary statistics for the variables included in the regression.

Loan amount and loan interest rate both enter with the expected positive sign. Based on column (5), a one standard deviation (1309 peso) increase in loan amount increases the predicted loss rate by 1.1 percentage point, while a one standard deviation (3.9 percentage point) increase in the interest rate increases the predicted loss rate by 5.4 percentage points. Higher down payments are associated with a 0.9 percentage point decrease in the predicted loss rate for a one standard deviation (0.083) down payment/price increase.

The measures of borrower credit risk also generally enter with the expected signs. Customers with an A credit score have loss rates 6 percentage points lower than customers with an N credit score. Large numbers of purchases, large balances, late balances, or moratory interest are associated with higher predicted loss rates (likely due to indicating a larger financial strain imposed on the borrower relative to available resources), while large credit levels in the past are associated with lower predicted loss rates, possibly by indicating that the borrower has had the ability to repay large balances in the past. Of the demographics, age and years living at home address have the strongest relation to loss rates in economic terms (for a one standard deviation change). A one standard deviation (10.8 year) increase in age lowers the predicted loss rate by 2.3 percentage points, while a one standard deviation (11.3 year) increase in years living at home address lowers the predicted loss rate by 2.0 percentage points.

Table 4, column (2)-(6) repeats regressions (1)-(5) from Table 3, but now adding dummies for the 32 product categories. The objective is first to determine whether the large differences in loss rates across product categories remain once transactions characteristics, credit risk measures, and demographics are controlled for, and second to determine how much incremental explanatory power the product category dummies add. For reference, column (1) of Table 4 shows a regression of loss rates on only the product category dummies themselves. In order to focus on differences in loss rates across product categories, I pick the category with the lowest default rate in Appendix Table 1, sewing machines, as a reference category (omitted dummy) and show the dummies on the other product categories which then measure how much higher the average loss rate is for a given product category relative to the average loss rate for loans for sewing machines. The table indicates significance levels for the product dummies by using a smaller and italic font for coefficients that are not significant at the 5% level.

Consistent with Table 1 and Appendix Table 1, column (1) of Table 4 confirms that average loss rates are substantially higher for electronics (especially cell phones and car audio), kids gear and toys, auto parts and bikes, watches, jewelry and glasses, relative to the benchmark sewing machine category. The relative magnitudes differ a bit from those of the loss rates presented in Appendix Table 1 because those loss rates were at the product category level, while the product dummy coefficients in the regressions estimate the average loss rates across loans in a given category (relative to the benchmark) and thus implicitly weights each loan equally regardless of its size.

Moving from column (2) to column (6) in Table 4 I add still more regressors as indicated in the top part of the table. Controlling for time as customer fixed effects, transactions characteristics, measures of borrower credit risk, demographics, and store fixed effects has very little effect on the relative differences in loss rates across product categories. For example, entertainment electronics have an average loss rate that is 8.9 percentage points higher than the average loss rate for sewing machines when no controls are included, and the difference is still as high as 6.8 percentage points when including the controls listed above. The large differences in loss rates across product categories are thus robust to controlling for standard predictors of default.

In terms of explanatory power, the product category dummies on their own generate an R^2 of 0.021. Comparing column (2)-(6) of Table 4 to column (1)-(5) of Table 3 allows for an evaluation of the incremental R^2 from adding the product dummies. In each case, the v in Table 4 is between 0.01 and 0.02 higher with the product dummies. While this is small in absolute terms, it is economically meaningful given the fact that R^2-values in regressions that predict loss rates tend to be very small both for the company analyzed here and in prior work (Gross and Souleles (2002), discussed in the introduction).

IV. What Drives Differences in Loan Loss Rates Across Product Categories?

To start assessing the economic forces underlying differential lender losses across product categories, I next estimate models of loss rates that include customer fixed effects to determine whether the differential loss rates across product categories are driven mainly by which types of individuals buy particular products, or whether loss rates then to be high in certain products regardless of who buys them. I find that cross-product loss differences are driven mainly by which types of individuals buy particular products and therefore consider how lenders best estimate individual effects using real time data.

A. Product effects versus individual effects

Consider a decomposition in which the loss rate on a loan made to individual i for buying product p has a product-specific component and an individual-specific component as well as a component driven by observables

(5) $$Loss\ rate_{i,p} = f_p + f_i + x_{i,p}{}'\beta.$$

The average loss rate across I individuals borrowing for purchasing product p is then

(6) $$Average\ Loss\ Rate_p = \frac{1}{I}\sum_{i=1}^{I} Loss\ rate_{i,p} = f_p + \left(\frac{1}{I}\sum_{i=1}^{I} f_i\right)_p + \frac{1}{I}\sum_{i=1}^{I} x_{i,p}{}'\beta.$$

In this setting, if one estimates a loan level regression for loss rates, including product dummies and observables, the regression coefficient on the product dummy for product category p will estimate $f_p + \left(\frac{1}{I} \sum_{i=1}^{I} f_i \right)_p$. It will thus capture both the product-specific effect for product category p and the average individual-specific component for individuals taking out loans to purchase products in product category p. If one instead estimates the same regression, but now include both product dummies, individual dummies (individual fixed effects), and observables, then the regression coefficient on the product dummy for product category p will estimate only the product-specific effect f_p.

In order to statistically be able to identify the product-specific effect f_p by including individual fixed effects in the regression, it is necessary that a lot of individual make purchases across several categories of products. Of 499,906 customers represented in the regressions, 179,311 purchased goods in both one or more of the four departments in Table 1 with lowest loss rates and in one or more of the five departments in Table 1 with the highest loss rates. Furthermore, focusing on the 32 more detailed product categories, the difference between the lowest and highest (of 32) default categories purchased by a given customer is 5.5% on average across customers. This suggests that there should be sufficiently many individuals with purchases across both high and low loss categories to separately identify the impact of product-specific effects and individual-specific effects.

Column (7) of Table 4 adds individual fixed effects to the product-level loss rate regression. The impact on the coefficients for the product category dummies is dramatic. The majority of them are now economically small and 12 of them are not significant. As an example, while the average loss rate on jewelry is 17.8 percentage points higher than that for sewing machines in column (6) which does not include individual fixed effects, this difference drops to only 3.1 percentage points once the individual fixed effects are added, suggesting that the high loss rate on jewelry purchases does not have much to do with jewelry per se (e.g. that people may buy gifts for their wives that they really cannot afford) but instead is due to the type of people who buy jewelry being higher risk regardless of what product they buy.

Figure 2 illustrates the impact of including individual fixed effects on the regression coefficients for the product category dummies. The figure sorts the 32 product categories based on their average loss rate from column (1) of Table 4. These average loss rates are illustrated by the upward sloping line in the figure, with each point labeled with the number of the product category used in Appendix Table 1. The flatter line in the figure illustrates the coefficients on the product category dummies from column (7) of Table 4, i.e. the ``true'' product-specific effects once the impact of which customers tend to buy particular products is taken out. Most of the product-specific effects are economically small. The vertical difference between the two lines shows the average individual-specific component for individuals taking out loans to purchase products in the category (plus the small effect of the observables). The vertical differences are large implying that there are large risk differences across products in the risk of the customer pool they attract.

B. Advice for Lenders -- Estimating Individual Effects Based on Purchase Patterns to Date

When including individual fixed effects in the regression in Table 4 one is using information about losses on past, current and future loans for a particular customer to estimate that customer's fixed effect. This information is not all available at the time a given loan is made. Suppose the lender had just read the results of Table 4. How does the lender best exploit this knowledge to estimate a given borrower's fixed effect in real time? I propose the following.

Estimating borrower fixed effects in real time: In a given month (from the start of 2007 onward), the lender can estimate the regression in column (7) of Table 4 using data for completed loans available up to that month.

(a) For a new customer borrowing to buy product p on some day during the following month, estimate that customer's individual fixed effect by $\left(\frac{1}{I}\sum_{i=1}^{I} f_i\right)_p$, i.e. the average individual effect of other customers buying product p in the past as estimated using available data.

(b) For a customer borrowing to buy a product on some day during the following month, and who has made other purchases in the past, estimate that customer's individual fixed effect by the average of $\left(\frac{1}{I}\sum_{i=1}^{I} f_i\right)_p$ across product p_1, p_2, ...,p_n purchased in the past or on this day.

Intuitively, using available data, the company can estimate what types of people (in terms of loss rates) tend to buy a particular product. It can then using these estimates to construct an estimate of what the individual fixed effect is for a particular customer based on what that customer is buying today and what that customer has purchased previously. Note that this real-time individual fixed effects estimator can be calculated for a given customer without any loss rate data being available for that customer. If repayment information is available for the particular customer, this information can be used to further improve the predicted loss rate using the regression coefficients on the repayment efficiency variables included in the regressions in Table 4.

In column (8) of Table 4 I investigate whether the real-time estimate of the individual fixed effect for a given customer is successful in capturing all the loss information contained in the product category dummies. This should be the case if the ``true" product-specific effects are economically small. The relevant comparison for column (8) is column (6). Comparing these two columns, it is clear that once the real-time individual fixed effect estimate is included, the coefficients on the product category dummies are small, indicating that the real-time individual fixed effect estimate comes close to capturing all the information about loss rates contained the information about what a customer is buying.[4] In the regression in column (9) I drop the product category dummies, resulting in only a tiny reduction in R^2 relative to column (8). Finally, comparing the R^2's for columns (8) and (9) to the R^2 from column (6), the R^2's are a bit larger for column (8) and (9). This is due to the fact that these columns use information about both what the customer is currently purchasing and what the customer has purchased in the past while column (6) only uses information about what the customer is currently purchasing.

How can the company best use the information about which borrowers have high real-time estimated fixed effects? The finding that the relation between loss rates and purchase patterns is driven mainly by customer fixed effects (as opposed to product-specific effects) implies that a policy of having different loan terms (down payments, interest rates) on loans for particular products is unlikely to be optimal. For customers assessed as being high risk based on their purchase patterns, the company should change loan terms even when these customers are purchasing products in categories that usually attract lower risk customers. For example, a person who has purchased only a cell phone in the past should be considered a high risk borrower and should not get a low down payment or interest rate even when buying an appliance or a sewing machine.

[4] The real-time individual fixed effect estimates are generated regressors. In principle this should be accounted for in the calculation of significance levels, but the t-statistic on the coefficient on the real-time individual fixed effect estimates are above 100, implying that any such correction will not change the conclusion that this variable is highly significant.

V. Luxuries Versus Necessities and its Relation to Individual Effects

While the determination that the relation between loss rates and purchase patterns is driven mainly by customer fixed effects is useful for lenders, it does not shed light on the underlying economics of why customers who tend to buy certain products are higher risks than others. I document next that high loss products tend to be luxuries and that they tend to be purchased by individuals who (in their spending at the retail chain analyzed) consume abnormally large fractions of luxuries given their income. Using less detailed U.S. data, I confirm that the link between preferences for luxuries and getting into repayment difficulties is not specific to the particular retail chain studied or to Mexico. I then discuss whether the existing literature on consumer choice includes theories that could explain why customers who buy a lot of luxury goods would tend to be the same people who default on consumer credit.

A. Products With High Loss Rates Tend to be Luxuries

A luxury good is defined as a good for which the budget share spent is increasing in the consumer's total spending while budget shares for necessities are decreasing in the consumer's total spending. For the individuals studied here I know only what is purchased at the particular retail chain that provided the data. I therefore calculate the budget share for a particular product category as spending on products in that category divided by all spending at the retail chain. For each of the 32 product categories I then run a pooled customer-level regression of the budget share on the customer's log annual income and use the regression coefficient (denoted by beta) as an indicator of the product's luxuriousness. I use (log) annual income as the regressor, as opposed to (log) total consumer spending at the retailer, because total consumer spending at the retailer is a very inaccurate measure of total consumption (with the implication that if this variable was used in both the budget share denominator and as a regressor, mechanical biases due to measurement error could arise). Since income is available at a categorical level, I set a given customer's income equal to the mid-point of the income range the customer's income belongs to. For the highest income category I set the customer's income equal to 1.5 times the income cutoff for belonging to this category (the exact multiple used is not crucial for the results that follow). I estimate that product categories 5, 6, 8, 10, 11, and 12 (all in the electronics department), 15 (office furniture), 20 (baby items), 30 (watches), and 31 (jewelry) are luxuries, i.e. have positive betas, where the product category numbers refer to the numbers given in Appendix Table 1.

Figure 3, Panel A graphs the average loss rate for each of the 32 product categories against the measure of luxuriousness, beta. The size of each point is proportional to the fraction of overall sales accounted for by this product category. The red line in the figure is the predicted value from a product level regression of average loss rates on the betas, where each product category is weighted by the fraction of overall sales it accounts for. The regression is shown in Table 5, Panel B, column (2).[5] There is a significant positive relation between product category loss rates and betas. The extreme example of this relation is the point to the top right in the figure which is cell phones -- these are the most luxurious items sold at the store and tend to have very high loss rates. To ensure that the relation is not simply driven by cell phones, Panel B of Figure 3, and column (4) of Table 5, Panel B, drops the cell phone category, and still finds a positive relation between average loss rates and betas, of stronger economic magnitude (based on the regression slope coefficient). A one standard deviation (0.00367) increase in beta is associated with a 9.3% percentage point increase in the loss rate (Panel A provides summary statistics for beta and related variables).

Table 5, Panel C, column (1) re-estimates the relation between loss rates and beta, this time at the purchase level, and with controls for product cost and log annual income of the customer. The relation between loss rates and beta is about as strong with the controls added as in the product level regression.

B. People Who Spend a Lot on Luxuries Tend to Have High Default Rates

Since the prior analysis in section IV found that the higher loss rates for certain product categories were driven by individual effects, it is relevant to determine whether the relation between loss rates and luxuriousness is also driven by individual effects. This would mean that the beta for a given product category should have little relation to loss rates after controlling for the type of individuals who tend to buy luxuries. I therefore calculate, for each customer, the fraction of spending at the retail chain which is spent on luxuries. Summary statistics are given in

[5] Table 5, Panel B uses heteroscedasticity robust standard errors since the average default rates for the different categories are based on different numbers of purchases and thus estimated with different amounts of precision. Furthermore, beta is a generated regressor and this should be accounted for in the calculation of standard errors in Table 5. The next draft will incorporate this adjustment.

Table 5, Panel A. For the average customer, luxuries account for 56% of spending, with a large (36%) standard deviation across customers.

Table 5, Panel C, column (3) adds the fraction spent on luxuries to the purchase-level regression, finding that it is quite strongly related to the loss rate with a one standard deviation increase in the fraction spent on luxuries being associated with a 3.4 percentage point increase in the predicted loss rate. The coefficient on beta drops from 13.3 to 9.0 when including the fraction spent on luxuries. Since the fraction spent on luxuries is likely to be an imperfect measure of an individual's loss type, column (4) includes individual fixed effects to determine if this further reduces the coefficient on beta. The beta coefficient is now 6.1, less than half than the value from column (1) which did not include any measure of individuals' loss types. More dramatically, column (5)-(8) repeats the analysis of column (1)-(4) leaving out cell phone purchases (cell phones are still included in the definition of the fraction spent on luxuries). Without cell phones, the coefficient on beta drops from 33.0 without any measure of individuals' types, to only 9.3 with individual fixed effects included. This suggests that the majority of the reason for higher loss rates for luxuries is due to the types of people who tend to buy luxuries being of high risk.

C. U.S. Evidence on Loss Rates for Individuals with Large Luxury Consumption

To ensure that the link between preferences for luxuries and getting into repayment difficulties is not specific to the particular retail chain studied or to Mexico, I use U.S. data for 19,815 households from the 2003-2007 Consumer Expenditure Survey (CES). The CES contains detailed information about a household's spending patterns, as well as information about the amount of consumer credit owed and the amount of interest charges, finance charges and late fees paid on consumer credit during the past year. The survey does not contain information about default on consumer credit, or about bankruptcies. I study the amount of consumer credit as one proxy for being more likely to incur repayment difficulties and supplement that evidence with analysis of finance charges which more directly measure whether the borrower has difficulties repaying the debt.

Each household is in the survey for four quarters (aside from attrition) and reports consumption quarterly. I exclude two CES consumption categories from total consumption: The category miscellaneous (miscpq) includes finance charges and if included could generate a mechanical relation between consumption patterns and finance charges. The category personal insurance and retirement (perinspq) includes life insurance and retirement contributions and thus should more appropriately be categorized as savings.[6]

By consumer credit I refer to borrowing on major credit cards, store credit cards, gas credit cards, store installment credit accounts, and consumer credit from financial institutions, with borrowing on major credit cards accounting for about 64% of such debt, and borrowing on store credit cards accounting for another 27%. Consumer credit information is collected in an annual supplement conducted in the household's last survey quarter.

I use the consumption categories defined by the CES and listed in Table 6, Panel B. For each category I estimate the beta (luxuriousness) by regressing the annual budget share for the category on the log of total annual consumption (which unlike in the Mexican data is available in the CES). Since total consumption thus enters both in the denominator of the budget shares and as the explanatory variable, I instrument log total consumption by log household annual income after tax to avoid any potential biases due to measurement error in total consumption. Table 6, Panel B shows the beta estimates. Food, housing, health care, and tobacco are necessities, while reading personal care, alcohol, apparel, education, cash contributions, entertainment, and transportation are luxuries.

Table 6, Panel C investigates the relation between a household's fraction of consumption spent on luxuries and the amount of consumer credit and annual finance charges (I use finance charges to refer to all types of charges, including interest, late fees, and any other finance charges).[7] Column (1) and (2) show that households who spend a larger fraction of their total consumption on luxuries (controlling for the log of annual consumption) have more consumer debt both in dollar terms and relative to their annual consumption. A one standard deviation (0.162) increase in the fraction spent on luxuries increases the dollar amount of consumer debt by $311 dollars -- a substantial effect given mean consumer debt of $3,068. Consistent with this, column (3) and (4) document that households who spend a larger fraction of their total consumption on luxuries (controlling for the log of annual consumption) incur more finance charges. The average household incurs finance charges of $169 per year and a one standard

[6] Following standard practice of papers using the CES (e.g. Vissing-Jorgensen (2002)), I drop households with incomplete income reports whose data are thought to be of lower quality. I also drop households who have less than the full four interviews, who reside in student housing, or who report a change in age of the respondent in between quarterly interviews which is negative or greater than a year.

[7] For consistency with the earlier regressions, the estimations are done by OLS. The next version will use Tobit models to avoid the assumption of a linear relation between E(y|X) and X in this part of the analysis.

deviation increase in the fraction spent on luxuries increases annual finance charges by about $12 (i.e. by about 7 percent of the mean). The less detailed U.S. data thus confirms the link between preferences for luxuries and getting into repayment difficulties documented in the Mexican data.

D. Why do People who Spend a Lot on Luxuries Tend to Generate High Loss Rates?

The evidence linking preferences for luxury consumption to loan loss rates is a first step towards addressing the economics underlying the main finding of the paper, the differences in lender losses across loans made to finance purchases in different product categories. In this section I speculate on what deeper economic heterogeneity across consumers may be at play. I argue that high spending (relative to available resources) and spending an abnormally high fraction on luxuries may be two dimensions of having a high preference for indulgence or of having low self-control to overcome impulses to indulge.

While economists define luxuries and necessities based on whether a good's budget share is increasing or decreasing in total spending, the marketing literature uses the labels in a broader sense, linking luxury consumption to hedonic pleasure and indulgence. Khan, Dhar and Wertenbroch (2004) note that ``Some researchers have used the terms luxury and necessity more broadly, in a less technical sense, to imply that luxuries and consumed primarily for hedonic pleasure while necessities are required to meet more utilitarian goals". The definition of hedonic and utilitarian goods in Dhar and Wertenbroch (2000) helps clarify the meaning of hedonic pleasure: ``Broadly speaking, hedonic goods provide more experiential consumption, fun, pleasure, and excitement (designer clothes, sports cars, luxury watches, etc.), whereas utilitarian goods are primarily instrumental and functional (microwaves, minivans, personal computers etc.)....". Kivetz and Simonson (2002) relate these concepts to indulgence, stating that ``Indulgence is closely related to both luxury and hedonics, often involving spending on items perceived as luxuries relative to one's means; these items are typically hedonic rather than utilitarian." Several papers, e.g. Hoch and Loewenstein (1991), model consumption outcomes as the result of an inner struggle between a desire for immediate gratification (indulgence) and will-power (self-control). Hoch and Loewenstein (1991) use a setting of time-inconsistent preferences in which self-control is a mechanism for enforcing the preferences of the long-term self over the preferences of what one could label the "short-sighted" self. Within the formalization of time-inconsistent preferences in Laibson (1997), in which life-time utility is given by $U_t =$

$E_t[u(c_t) + \beta \sum_{\tau}^{T-t} \delta^{\tau}\, u(c_{t+\tau})]$, the long-term self refers to a decision maker with β=1 while the short-sighted self refers to a decision maker with β<1. Within this formalization, one could model β as an increasing function of self-control. A possible interpretation of my empirical findings is then that some households have either stronger preference for immediate consumption (i.e. a β further below 1 for given self-control), or have less self-control (and thus lower β for given desire for immediate consumption). These consumers have higher current spending (relative to available resources) and could also have a stronger preference for indulgence within current consumption, in the sense of spending an abnormally high fraction on luxuries relative to their total consumption.

One approach to test whether this is the correct interpretation of the empirical findings would be to investigate the link between self-control and purchase patterns. Do consumers who spend a large fraction on the product categories that generate high lender losses have less self-control in the sense that more of them succumbed to a desire for immediate gratification during their visit to the store? One could test this by surveying customer to ask whether the decision to buy was made in the store today or was the result of careful prior consideration weighing the costs and benefits of the purchase. Or one could use a more elaborate scale to measure impulsive shopping based a several questions (see Rook and Fisher (1995) or Puri (1996)). I am aware of only one study which has empirically investigated the link between impulsive buying and which good was purchased. Bellenger, Robertson and Hirschman (1978) conducted a survey of 1,600 consumers at a U.S. department store, asking consumers what they bought today and when they decided they wanted to purchase each item. The consumption categories for which the largest fraction of purchase decisions were made in the store today were jewelry, food, and women's shoes. These findings are consistent with the idea that consumers with a stronger desire for immediate gratification or lower self-control tend to spend more on certain products, and while Bellenger et al. (1978) did not relate purchase patterns to consumer default risk this would be possible in the Mexican data.

An alternative would be to conduct a field experiment in which down-payment requirements were increased for consumers with high estimated default risk based on their purchase patterns to date. If a large fraction of these consumers are impulse buyers this would have a large and permanent impact on sales to and loan losses from this segment of customers (since these customers by their impulsive nature do not plan ahead for down-payments). If instead default risk is not related to planning and self-control, one would expect such an intervention to have only a temporary and smaller impact on sales and losses since consumers would learn to plan ahead for the larger down-payment.

VI. Conclusion

The paper seeks to add to our understanding of default on consumer credit using a large new data set from Mexico. The main finding is that lender losses on consumer loans differ dramatically by the type of product purchased. This is shown to be driven primarily by high-risk borrowers having a tendency to buy some product more than others, as opposed to being a product-specific effect. In terms of the economics driving the relation between product type and lender losses, I showed that high loss products tend to be luxuries and that consumers who spend a lot on luxuries given their income on average are higher risk. Heterogeneity across consumers in desire for immediate gratification or in self-control provides on possible explanation for why some consumers both have high current consumption relative to their resources (and thus generate higher lender losses) and have a preference for luxury goods (hedonic goods) within their current consumption.

References

Adams, William, Liran Einav and Jonathan Levin, 2009, ``Liquidity Constraints and Imperfect Information in Subprime Lending", *American Economic Review*, 99, 1, 49-84.

Bellenger, Danny N., Dan H. Robertson, and Elizabeth C. Hirschman, 1978, ``Impulse Buying Varies by Product", *Journal of Advertising Research*, 18, 6, 15-18.

Dhar, Ravi and Klaus WertenBroch, ``Consumer Choice between Hedonic and Utilitarian Goods", *Journal of Marketing Research*, 37, 1, 60-71.

Fair Isaac Corporation, 2005, ``Understanding Your FICO Score".

Gross, David B and Nicholas S. Souleles, 2002, ``An Empirical Analysis of Personal Bankruptcy and Delinquency", *Review of Financial Studies*, 15, 1, 319-347.

Hoch, Stephen J. and George F. Loewenstein, 1991, ``Time-inconsistent Preferences and Consumer Self-Control", *Journal of Consumer Research*, 17, 492-507.

Honore, Bo. E, 1992, ``Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects", *Econometrica*, 60, 533-567.

Keys, Benjamin J., Tanmoy K. Mukherjee, Amit Seru and Vikrant Vig, 2010, ``Did Securitization Lead to Lax Screening", forthcoming, *Quarterly Journal of Economics*.

Khan, Uzma, Ravi Dhar, and Klaus Wertenbroch, 2004, ``A Behavioral Decision Theoretic Perspective on Hedonic and Utilitarian Choice", INSEAD working paper, 2004.

Laibson, David, 1997, ``Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics*, May, 443-477.

Puri, Radhika, 1996, ``Measuring and Modifying Consumer Impulsiveness: A Cost-Benefit Accessibility Framework", 5, 2, 87-113.

Rook, Dennis W. and Robert J. Fisher, 1995, ``Normative Influences on Impulsive Buying Behavior", *Journal of Consumer Research*, 22, 3, 305-313.

Vissing-Jorgensen, Annette, 2002, ``Limited Asset Market Participation and the Elasticity of Intertemporal Substitution", *Journal of Political Economy*, 110, 4, 825-853.

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