

Predicting Credit Card Delinquency:
A Fundamental Model of Cardholder Financial
Behavior

Håvard Huse
NHH Norwegian School of Economics

May 2019

Abstract

This thesis proposes a model of credit card customer delinquency based on theoretical advancements in financial decision-making. As follows, this thesis has two main research purposes.

First, credit card delinquency is modeled explicitly, incorporating mechanisms from mental accounting and financial decision-making. This allows for more realistic modeling of cardholder behavior, while simultaneously inspecting the validity of these theoretical concepts.

Second, the modeling specification advances previous research in the behavior scoring literature. Accounting for individual-level heterogeneity, dynamic effects are assigned as individual lag weights using a segmented approach. Hence, potentially different behavioral patterns between non-delinquent and eventual delinquent cardholders are modeled directly.

Using a comprehensive dataset combining credit and debit transactions of cardholders between June 2008 and June 2011 from a Norwegian bank, support is found for the following three hypotheses related to mental accounting and present bias. First, increased payment decoupling leads to a higher likelihood of delinquency, when continued borrowing is promoted by reduced salience of past expenses. Second, the results show that behavior consistent with persistence of decision-making ineptitude also increases the likelihood of delinquency. Some cardholders habitually spend excessively, refusing to accommodate consumption to a financially reasonable level. Third, a lower concern for future consequences also increases the likelihood of delinquency. Present-biased individuals tend to discount future credit card repayments at a higher rate and consistently spend at perilously high rates.

Further, the results reveal how the structure of dynamic effects improves prediction of delinquency. Capturing the heterogeneous effects of previous financial status leads to a more precise understanding of cardholder behavior. The proposed model has greater predictive performance than machine learning algorithms that are frequently applied to credit scoring data.

Acknowledgments

There are several people to whom I am grateful for helping me complete this thesis. First and foremost, I would like to thank my advisor, Professor Sven Arne Haugland. I am confident that he went above and beyond what is expected of him, displaying patience and faith when progress stalled. The existence of this thesis is also due to the help and insight provided by Professor Auke Hunneman. His assistance was crucial to ensuring the quality of my work and the completion of the project. This thesis also benefited greatly from feedback on later versions by Professors Mike Palazzolo, Koen Pauwels, and Rutger Daniel van Oest.

I would also like to thank my former and current employers: the Norwegian University of Science and Technology and BI Norwegian Business School. This thesis is only possible because of their faith in me and their willingness help in any way possible. Several former and current colleagues deserve thanks, including, but not limited to: Johannes Mauritzen, Ove Gustafsson, Espen Gressetvold, Arve Pettersen, Arne Morten Ulvnes, Tor Georg Jacobsen, Tor-Eirik Olsen, Magnus Ressem, Steinar Wold, and Morten Erichsen.

My research was largely facilitated by two stays at the University of Michigan. I would like to thank Professor Richard Bagozzi for welcoming me to the Ross School of Business and the marketing faculty there. Allowing me to attend the Ph.D. classes in marketing, econometrics, and statistics at the University of Michigan provided me with knowledge that has been indispensable to my work.

Last, but not least, I would like to thank my family and loved ones. My sister, Irene, my brother, Bjørnar, my mother, Edel, and my father, Odd-Bjørn. They, especially my father, consistently urged me to move forward providing unfaltering support.

Trondheim, May 2019

Håvard Huse

Contents

1	Introduction	1
2	Theory	11
2.1	The Mental Accounting Framework	13
2.1.1	Core Concepts	13
2.1.2	Relevant Empirical Applications of Mental Accounting	15
2.1.3	Payment Decoupling	19
2.1.4	Decision-making Ineptitude	25
2.1.5	Present Bias	28
2.1.6	Financial Risk Aversion	31
2.2	Classification Methods in Credit Scoring	32
2.2.1	Single Classifiers	33
2.2.2	Classifier Ensembles	35
3	Methodology	39
3.1	Data and Selection of Variables	39
3.2	Sample Selection	40
3.2.1	Descriptive Analysis	45
3.2.2	Selection and Construction of Variables	49
3.3	Model Specification and Estimation	56
3.3.1	Estimation of Proposed Model	58
3.3.2	Specification of Classification Algorithms for Prediction	61

3.3.3	Classifier Performance Indicators	62
4	Results and Analysis	65
4.1	Model Estimates	65
4.2	Model Fit and Hypothesis Tests	75
4.3	Predictive Performance	76
5	Discussion and Conclusion	81
5.1	Discussion of the Results	81
5.1.1	Decoupling	82
5.1.2	Decision-making Ineptitude	84
5.1.3	Present Bias	85
5.1.4	Implications for Theory	87
5.1.5	Implications for Classification Methods	91
5.2	Managerial Implications	92
5.3	Limitations and Directions for Further Research	94
5.4	Conclusion	96
6	Bibliography	97
	Appendices	116
A	Predictive Performance – Resampled Training Data	117
B	Simulation Descriptive Statistics	119
C	Convergence Diagnostics and Figures	122

1 — Introduction

Credit card spending accounts for a large amount of consumer debt, as well as numerous daily transactions. According to the Federal Reserve Bank of New York, outstanding U.S. credit card debt totaled \$870 billion in the fourth quarter of 2018.¹ As a means of consumer transactions, credit cards were employed in 22 billion transactions totaling an estimated \$2.1 trillion in the United States in 2011 (Canner and Elliehausen, 2013). Notwithstanding the Credit Card Act of 2009², which sought to improve consumer protection, the credit card industry still sees delinquency rates of around 3%.³ Though appearing minuscule, the rate of delinquency misrepresents the actual number of cardholders that encounter repayment difficulty. Following a panel of cardholders from 2007 to 2011, Canner and Elliehausen (2013) found that 30% fell behind on payments at least once, while 20% fell behind at least 30 days. While this might indicate that credit cards are detrimental to consumer finances, the same report cites survey data in which 85% of consumers who “hardly ever” pay their balance in full also attribute credit cards to making financial management “less difficult.”

¹The Federal Reserve Bank of New York Center for Microeconomic Data provides current data at <https://www.newyorkfed.org/microeconomics/data.html>.

²The Credit Card Accountability Responsibility and Disclosure Act of 2009.

³Delinquency rates have steadily declined since the end of the Great Recession, when delinquency rates reached a peak of 6.8%. See the Statistical Release “Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks” by the Board of Governors of the Federal Reserve System at <http://www.federalreserve.gov/releases/chargeoff/> for current delinquency rates.

In 2013, 38.1% of U.S. families carried a credit card balance, owing an average of \$5,700⁴ (Bricker et al., 2014). Given that roughly two-thirds of adults hold one or more credit cards (Canner and Elliehausen, 2013), the market is not only sizeable but also unexpectedly profitable given the number of competitors in the industry (Ausubel, 1991). Early research on credit card behavior focused primarily on two inconsistencies: 1) Why are credit card interest rates persistently high (e.g., Calem and Mester, 1995)? and 2) Why do consumers continue to borrow using credit cards, when other sources of funds are available (e.g., Brito and Hartley, 1995)? Pursuing the latter area of investigation, consumer behavior as it relates to credit card usage and subsequent customer profitability is a natural extension of this line of research.

Credit card lenders meticulously combat the perpetual risk inherent in the industry; some customers are likely to borrow more than they can afford. Assessing the risk of lending to a particular customer has long been synonymous with credit screening techniques, spearheaded by numerous credit scoring schemes. Avoiding the adverse selection problem inherent in financial lending corresponds to targeting and selective customer acquisition in marketing theory, which has been essential to marketing strategy since the inception of enterprise and business.

In marketing, assessing the efficiency of marketing strategies, such as targeting, building brand awareness, or product design to glean a return on investment from marketing activities, has been at the forefront of the discipline this millennium (e.g., Rust et al., 2004). While customer asset and portfolio assessment has been restricted to contexts that include rich purchase histories and service encounters (often found in business-to-business scenarios), the development of customer value techniques adapted to conditions often found in consumer markets has been more arduous and less

⁴The median amount for a family carrying a credit card balance was \$2,300, indicating a significantly skewed distribution.

straightforward.⁵ A similar evolution is mirrored in the credit card industry, where models that identify risky customers *ex ante* are well developed and ruthlessly implemented to avoid bad customers. While these methods have been commonplace in the finance industry for decades, the management of existing customers based on observed behavior is not equally well developed in theory or practice.

Background for the Study

Assessing the financial accountability of marketing decisions has previously been difficult, because the field has lacked accurate data-driven methods to measure and track the effects of these decisions (Rust et al., 2004). However, the inception of Bayesian statistical methods has encouraged flexible approximations of individual preferences and subsequent behavior. These developments are also natural extensions of the data that have been available; as scanner panel data became attainable, they were a logical point of departure for many of the modeling dilemmas. From the precision of segments (Kamakura and Russell, 1989) to individual-level heterogeneity (Allenby and Rossi, 1999), from brand choice (Guadagni and Little, 1983) to the shopping basket (Manchanda et al., 1999), and from consumer packaged goods (Gupta and Chintagunta, 1994) to related technology products (Sriram et al., 2010), the usefulness and flexibility of these models have flourished along with their complexity and diversity.

Especially pertinent to marketing activities is the fact that consumer heterogeneity plays a prominent role in how product attributes are specified and in the eventual success of the products. Pricing and design decisions are based on this axiom: valuation, preferences, and price sensitivities differ between consumers. Rather than a bothersome nuisance that causes noise in model estimation, as it might be regarded in the econometrics literature, this unobserved heterogeneity is the innate lifeline the market extends to

⁵For a review of customer lifetime value research, see Berger et al. (2002).

marketers and researchers of consumer behavior. Accurately estimating the costs of acquiring and keeping different customers, along with customer profitability, have become central tenets in the marketing literature. Concepts like customer lifetime value, customer equity, and customer profitability seek to unmask desirable customers while balancing the marketing costs required. Maximizing the profitability of the customer portfolio not only entails assessment of which marketing actions are most effective but also relinquishing customers who do not maximize customer portfolio profitability.

Seeking to maximize the customer portfolio profitability of credit card customers, “behavior scoring” attempts to model cardholder behavior based on revealed preferences. In contrast to the selection-based doctrine of application scoring, behavior scoring tries to measure risk and customer value given the actual spending and repayment behavior of the customer in question. In marketing terms, this translates to measuring customer profitability, whereby cardholders are judged not on assets or demographics but on individual behavior as credit card consumers. Balancing acquisition and retention in most industries not only involves screening of potential customers but also tailoring marketing actions towards existing customers. For credit card applicants and cardholders, maximizing customer portfolio value primarily entails evaluating risk of default, supplemented by considerations of customer churn and loyalty.

Credit cards represent a substantial amount of payment transactions, as well as an integral part of mundane consumption decisions. Like most other financial products, credit cards have a lack of available data coupled with an intricate role in consumption that has led to relative neglect of cardholder behavior in marketing research. Granted, there have been a few studies of credit card preference using survey data (e.g., Yang and Allenby, 2000; Yang et al., 2007), but research using actual credit card spending has only recently appeared. Most research regarding credit card customer profitability has centered on improving application scoring techniques with modern

statistical advancements (see Lessmann et al., 2015). Recently, avant-garde research on cardholder value has incorporated aggregate monthly data to investigate cardholder value, showcasing how individual-level heterogeneity is paramount in explaining variations in behavior (Zhao et al., 2009; Khandani et al., 2010). These contributions attempt to capture delinquency by applying existing model structures, thereby circumventing the essential exercise of specifying appropriate theoretical model structures. In addition, credit data alone will likely be inadequate when trying to understand individual financial situations. Predicting the likelihood of delinquency requires a comprehensive approach to financial decision-making and financial status. This necessitates a realistic model structure that includes not only credit transaction data but also debit transactions.

Understanding credit card delinquency incites an examination of how consumers appraise decisions to borrow and repay credit. Modeling revealed preferences of credit card borrowing and repayment in a meaningful way requires a theoretical foundation that realistically reflects decision-making. Behavioral economics (Kahneman, 2003*a*) represents a maturing field that attempts to understand authentic human behavior, adding nuance to the hyperrational assumptions found in the utility maximizing framework. The mental accounting branch of behavioral economics (Thaler, 1980, 1985, 1999) invokes several concepts that are assumed to be essential to consumption and financial decisions, providing a logical scheme to understand cardholder behavior.

Positioning of the Study

Facing a decision to purchase, a consumer encounters a potentially complex problem involving several tradeoffs: paying with credit or debit, assessing if the eventual debt can be paid off before the interest-free period expires, and comparing the cost of the item including eventual interest and fees with the perceived utility received. A model of rational choice as expressed by the

permanent income hypothesis (e.g., Hall, 1978) asserts that consumers will spend to maximize their permanent utility. This includes rational decisions to either save or accelerate spending, such that consumers will attempt to balance a current budget while anticipating future expenditures (including instances of bad luck or fluctuations of liquid assets). Accordingly, consumers will choose to borrow when confronted with incidental outlays, adapting to sudden shifts of income or cost, or accelerating purchases when appropriate.⁶ Such circumstances explain why cardholders, on average, choose to utilize their credit card. However, understanding why some cardholders are more prone to borrowing and delinquency requires an examination of the psychology of spending decisions. While the apparent heterogeneity in risk of delinquency is apparent, a vital query should be asked: What are the behavioral antecedents that promote differing risks in cardholders?

Various departures from the assumption of human rationality inherent in the traditional utility maximizing framework have been developed in the still-growing field of behavioral economics (DellaVigna, 2009). The field of behavioral economics examines how *homo sapiens* departs from *homo economicus* when making decisions. These theories of human behavior have uncovered a range of biases across several research veins. Instead of careful deliberation, utility maximizing, and stable preferences, behavioral economics introduces social preferences, time inconsistency, heuristics, and bounded rationality. A considerable constituent of behavioral economics is founded on concepts from mental accounting, which are frequently applied to various decision-making “anomalies” where the strict assumption of rational choice fails. Mental accounting itself is an amalgamate of microeconomic theory and cognitive psychology (Thaler, 1985, 1999) developed to realistically describe observed behavior where assumptions of rational and steadfast decision makers are often inaccurate. The framework integrates concepts relevant to financial

⁶Exemplifying these behaviors with a domestic flavor could include: buying a new refrigerator when the old one breaks down, paying the heating bill when facing reduced employment, and buying a new TV in time for a major sporting event.

decisions and consumption, portraying choices as subjective and often incomplete judgments of utilities.

Research examining the concepts in mental accounting usually considers individual concepts using settings with fictional experiments (DellaVigna, 2009). Though some recent contributions have examined mental accounting effects empirically (e.g., Hastings and Shapiro, 2013; Seiler et al., 2012; Hilovich and Gilovich, 2014), studies juxtaposing numerous effects are scarce. Variance in cardholder behavior, whether they are spending patterns or repayment decisions, are likely borne out of multiple psychological processes, not only economic situations. In addition, experimental studies have shown that judgment of financial decisions, along with discounting rates, differ significantly between consumers (e.g., Thaler, 1981). Regarding credit card usage, this implies that cardholders have differing preferences regarding credit card spending and repayment. Some cardholders choose to borrow and spend more in proportion to their income, while fiscally conservative cardholders choose to neglect their desires and wants. The present study incorporates several concepts from mental accounting using a model that captures heterogeneity in decision-making.

Consumer rationality regarding credit card borrowing is clearly challenged by theories like present-biased preferences, whereby consumers prefer different payoffs in the long term and short term. This suggests that intra-month spending behavior will be a useful indicator of an individual's concern for their future financial situation, when present-biased individuals will exhaust their monthly disposable income quicker. Although this is not a dilemma *per se*, those who indulge undaunted likely have a lackadaisical approach to their future financial situation. In addition, credit cards inherently promote decoupling, the tendency to separate costs and benefits of expenses (Prelec and Loewenstein, 1998). Some cardholders are more affected by these tendencies, suggesting why some use credit cards haphazardly with no heed of past expenses. The deftness and capability of understanding financial de-

cisions is also considered heterogeneous (Thaler, 1999) and durable. Thus, choosing to borrow at recklessly high levels is, at least partly, presumably due to enduring personal characteristics. Individuals with a history of poor decision-making regarding credit uptake will likely repeat that behavior.

Modeling these concepts necessitates comprehensive data: individual-level spending and borrowing data for both credit and debit accounts. The model structure and variables also need to account for heterogeneity in behavior, and more importantly, the dynamic effects proposed by the mental accounting concepts. The research questions in this thesis contemplate the relevancy of mental accounting in understanding delinquency and predicting credit card customer delinquency:

Do mental accounting concepts significantly explain credit card delinquency in a behavior scoring model? If so, does the mental accounting model outperform current machine learning algorithms when predicting delinquency?

Accordingly, three hypotheses are constructed based on concepts from the mental accounting literature in addition to a hypothesis consistent with a rational choice model. The first hypothesis suggests that some cardholders have a higher tendency to decouple payment from utility. As a consequence, their valuation of future repayment is biased, leading to a higher likelihood of delinquency. The second hypothesis suggests that personality traits consistent with poor financial decision-making are persistent. Accordingly, the likelihood of delinquency is inherently greater for particular cardholders. The third hypothesis relates willpower and present bias to a higher likelihood of delinquency, demonstrating how recurrent shortsightedness eventually catches up to cardholders. Finally, the fourth hypothesis suggests that risk aversion lowers the likelihood of delinquency, in accordance with the permanent income hypothesis.

Contribution of This Thesis

In this thesis, a fundamental model of credit card delinquency based on debit and credit spending behavior is constructed.⁷ The model specification is designed to reflect the financial situation that cardholders face, mirroring recent advancements to the mental accounting literature (Thaler, 1985). Using proprietary data from a large credit card provider in Norway, individual credit card transactional data is matched with individual debit transactional data. This unique dataset allows for a more precise and nuanced understanding of the financial situation cardholders face when they are confronted with a decision to repay credit card debt.

The full model with credit and debit data features variables that attempt to capture variation in individual financial status and subsequent effect on delinquency. More importantly, a segmented lag weight configuration of financial status allows for differing effects of previous financial conditions. The segmented approach to random individual lag weights in addition to the nuanced data allows for an extensive investigation of mental accounting effects in credit card delinquency. Specifically, delinquent cardholders exhibit behaviors consistent with the effects of decoupling (Prelec and Loewenstein, 1998), decision-making ineptitude (Ameriks et al., 2003), and present-biased preferences (Meier and Sprenger, 2010) that are believed to promote financial peril. Compared to modern machine learning algorithms, the modeling procedure that incorporates behavioral dynamics and variables based on mental accounting predicts delinquency at a superior rate. This thesis not only contributes to the behavior scoring literature in providing a superior modeling procedure but also forwards the mental accounting literature by applying the framework to an empirical setting. To conclude, this thesis provides a more

⁷A fundamental model attempts to explain a concept using elements that are perceived to be influential (Bettman et al., 2009). For example, fundamental analysis in finance refers to the process of explaining share prices with factors such as book value per share and earnings per share. This is in contrast to technical analysis, which looks to factors, such as lagged price and momentum, to predict share price.

precise understanding of the risk taken by cardholders using nuanced data that attempt to reflect some of the underlying behavioral processes inherent in spending and lending.

2 — Theory

Central to the assessment of profitability for credit card customers is establishing a likelihood of delinquency. This likelihood is fastidiously examined *ex ante* using credit scoring techniques, similar to the selection process of customer acquisition in traditional marketing nomenclature. Estimating the customer equity for a portfolio of cardholders is inherently tied to evaluating the likelihood of delinquency in the portfolio. Examining this likelihood of delinquency, behavioral economics provides a plethora of effects that potentially yield explanatory power superior to a model featuring rational choice. Economic choices can be portrayed as rational analysis leading to optimal behavior, as in the neoclassical expected utility model, or understood as fallible and limited judgments. Thaler (2016) argues that while the traditional model of utility maximization provides a prescription of what characterizes optimal behavior, describing revealed (actual) behavior requires a framework to explain deviations from rational choices.¹ The concepts introduced in behavioral economics mostly stem from the idea of bounded rationality (Kahneman, 2003*a*), though the linkages between the concepts are loosely defined. Thus, in spite of the additional explanatory power provided, insights often observed in laboratory experiments are difficult to apply in empirical

¹Thaler (2018) understands the utility maximizing model as normative (and he implores his students to make decisions according to it), while behavioral economics provides a descriptive framework of behavior. Further, he argues that behavioral economics should not supplant the standard model, though we should be aware of the predictive shortcomings of the latter framework.

settings of economic choice (Barberis, 2013).

Reviewing the empirical evidence of the concepts, DellaVigna (2009) organizes the field of behavioral economics into three domains: nonstandard preferences, nonstandard beliefs, and nonstandard decision-making. Relevant to credit card payment and repayment decisions is how preferences deviate from a rational choice model and lead to suboptimal evaluations of utility. As Prelec and Loewenstein (1998) note, evaluating the utility of payment decisions involves several psychological aspects in addition to evaluating liquidity and future outlays. Moreover, choosing to finance consumption with a credit card involves evaluating the utility of the purchase itself, along with assessing the present value of future payments and repayment of current debt. Mental accounting (Thaler, 1980, 1985, 1999, 2018) provides a realistic framework of how utility is perceived by decision makers, representing a sensible depiction of how consumption and borrowing decisions are framed and coupled. Here, the likelihood of delinquency is modeled using the mental accounting framework in addition to concepts from behavioral economics.

Credit scoring in general identifies credit card customers in risk of delinquency either *ex ante*, referred to as application scoring, or in the form of existing cardholders, referred to as behavior scoring.² Although application scoring and behavior scoring are related, research has disproportionately favored application scoring. The proliferation of machine learning algorithms for classification has, along with publicly available datasets, created a large influx of studies of application scoring (Lessmann et al., 2015). As such, examining research on credit scoring necessitates an assessment of both behavior and application scoring, while being mindful of the similarities and discrepancies in data and methods. Here, the general field of credit scoring will be probed for effective ways of classifying risky cardholders to provide

²Colloquially, credit scoring and application scoring are often treated as synonyms, hence the ubiquitous concept credit score, which is related to creditworthiness. Here, credit scoring is treated as an umbrella term encompassing both application and behavior scoring.

baseline comparisons to the model proposed in this thesis.

The following section details the central axioms of mental accounting, relevant applications of the mental accounting framework, and hypotheses related to specific effects. Section 2.2 examines extant classification methods in behavior scoring in addition to relevant methods employed in application scoring.

2.1 The Mental Accounting Framework

The fundamental assertion that mental accounting makes is that decision makers are simultaneously affected by bounded rationality, deficient willpower, and pro-social behavior (Jolls et al., 1998). As such, decision outcomes are influenced by various innate compensation strategies and are often suboptimal. Perhaps the most prominent and widely studied compensation strategy is assigning experiences and expenditures to different mental categories. Constructing a model of delinquency based on prescriptions from mental accounting requires a delineation of the central doctrines of the theory. The use of mental categories, in particular, is important to understand how processing financial decisions will contradict models assuming rational individuals that maximize expected utility. Accordingly, the next section presents the core concepts of mental accounting along with the factors employed in the research model.

2.1.1 Core Concepts

Assigning outcomes to mental categories has a potentially large impact on how they are perceived, as different individuals could experience the same outcome as either a gain or a loss (Thaler, 1985). Consider a couple who recently paid for an unanticipated car repair preceding a shopping trip. A rational contemplative individual would recognize that a reduction of discretionary spending would be favorable in light of the sudden car expense.

However, using mental accounts, another individual would designate the car repair as belonging to a separate budget. Accordingly, this individual might haphazardly prefer to keep discretionary spending at previous levels, perhaps requiring additional financing by borrowing.³ Tendencies to partition expenses illustrates how outcomes are framed, often in a piecemeal fashion. While a utility-maximizing model of choice generally assumes perfect information and rationality in decision-making, actual human behavior often entails splitting decisions into several parts or assigning outcomes to different categories (Kahneman and Tversky, 1984). Using appropriate terminology, mental accounting establishes the consequences and effects of eschewing a comprehensive account in favor of a topical or minimal account (Thaler, 1999).

Central to mental accounting is how outcomes are framed as either gains or losses using the value function. Rather than objective utility, the value function is a representation of how individuals subjectively perceive transactions (Kahneman and Tversky, 1979). The suggested shape of the function implies three essential concepts: transactions are evaluated as relative change in utility, decisions are evaluated sequentially relative to a reference point, and losses provide more disutility than the utility of a proportional gain (Thaler, 1999). Regarding economic decisions, the value function in mental accounting has simultaneously explained and provided a point of departure for several observed concepts. Examining empirical studies of mental accounting reveals more precise indications of the possible effects which are relevant to include in a value function representation of credit card usage.

³Thaler (1985) illustrates the same effect using a couple who is compensated by an airline after losing their catch from a fishing trip. The couple could have assigned the compensation to something more general, such as “vacation,” though in the example they designate it as “food,” and enjoy an unusually pricey meal. Also, the fact that the compensation is coded as a gain and not simply a null-sum gain integrated with the adjoining loss of the fish is not self-evident.

2.1.2 Relevant Empirical Applications of Mental Accounting

The act of assigning outcomes to different mental accounts has been examined in a variety of settings: coupon and gift card redemption (Milkman and Beshears, 2009; Cheng and Cryder, 2018; Hilovich and Gilovich, 2014), tracking and allotting time to mental accounts (Rajagopal and Rha, 2009), pricing trade-ins for durable goods (Okada, 2001), income sources and spending (Feldman, 2010; Davies et al., 2009), disposition effects in stocks and real estate (Lim, 2006; Seiler et al., 2012), and how children evaluate outcomes (Webley and Plaisier, 1998). Recently, investigations have examined the boundaries of versatility and construction of mental accounts. To wit, Cheema and Soman (2006) pry into the malleability of mental accounts, conducting experiments that examine the assignment of costs dependent on the existence of preconceived mental accounts. Subjects were shown to assign different costs to mental accounts according to the ambiguity of the costs, which in turn decided how likely they were to continue spending on similar expenses when left with a surplus. Subjects were also more likely to spend when given the flexibility of constructing mental accounts on their own, suggesting that this let them justify prior expenses. In addition, mental accounting effects have also been shown to be sensitive to modality, such as when comparing time to money (Duxbury et al., 2005; DeVoe and Pfeffer, 2007), decision simultaneity (Chatterjee et al., 2009), and comparing relative changes to absolute changes in outcome (Heath et al., 1995).⁴

⁴Interestingly, mental accounts have also been suggested as a useful heuristic when precise calculations are impractical (Antonides et al., 2011). Although prevailing research in mental accounting generally examines how suboptimal solutions are achieved (DellaVigna, 2009), Antonides et al. (2011) investigate if mental accounting can be advantageous. Interestingly, long-term orientation is positively correlated with mental accounting, rather than short-term orientation. In addition, mental accounting seems to improve financial overview and financial management, suggesting that individuals who have a proactive approach to managing their finances use mental accounts. However, it should be noted that the retrospective approach used in the scale likely measures actual successful accounting,

In general, applying mental accounts and subsequent heuristics is expected to lead to sound but also foolish decisions, where the latter is essential to understanding how decision makers reach unfavorable situations (for a review, see DellaVigna, 2009). An especially illustrative example of how mental accounts shape decision-making is cited by Thaler (2015) who refers to the findings portrayed in Hastings and Shapiro (2013). The results alluded to show that the share of premium gasoline increases beyond income effects when gasoline prices fall. Thus, the behavior displayed is consistent with assigning expenditures to different mental categories, where individuals justify splurging on premium gasoline when their allotted budget allows it. Examining consumer behavior regarding credit cards also reveals behaviors at odds with assumptions of rational decision makers. Agarwal et al. (2007) use a natural field experiment to investigate consumer spending following the Bush 2001 tax relief. Spending was initially lower after the tax relief, while consumers reduced their credit card borrowing. However, spending increased shortly thereafter, especially by low limit cardholders and cardholders who utilized most of their available credit, contrary to the permanent income hypothesis. In a mental accounting framework it could be argued that these cardholders are accustomed to keeping credit borrowing at a certain level, and will gradually revert to the level of borrowing they are suited to. As such, these differing tendencies to employ mental accounts should influence decision-making, eliciting actions inconsistent with rational behavior.

Extending the implications of keeping separate mental accounts, Thaler (1999) suggests a framework prescribing how individuals separate (segregate) or combine (integrate) outcomes. When evaluating several outcomes, that are framed as gains or losses, individuals will segregate or integrate the outcomes to justify past or current behavior. Literature examining segregation

rather than planned mental delineation of spending or the extent of mental accounting entering into spending decisions. The survey employed also seems to operationalize mental accounting as comprehensive evaluations, in which mental accounts are used as a shorthand in place of more involved calculations.

and integration has probed the range of applicability of the purported effects (Thaler and Johnson, 1990; Lim, 2006), where temporal contiguity has demonstrated a strong effect (Cowley, 2008). Considering purchases and financial decision-making, the exchange of money for a good is not evaluated as a loss, as long as the cost is budgeted for (Novemsky and Kahneman, 2005). Research focusing on credit decisions, such as McHugh et al. (2011) and Ranyard et al. (2006), suggest that consumers may segregate or integrate the cost of debt in a heterogeneous manner. As explained in Ranyard et al. (2006), credit decisions can be framed by two different representations: a comprehensive account (assessing the total cost of the loan) and a topical account (choosing a loan based on heuristics, such as interest rate).⁵ Using a comprehensive account, consumers will choose a loan based on total cost and not interest rate. Consumers framing the decision using topical accounts will select the loan that gives the lowest monthly payment, preferring payment plans with a lower cost per period. Similar findings were also reported in McHugh et al. (2011), demonstrating how heuristics influence the choice of credit terms. This indicates how individuals will frame the costs of using credit differently, in addition to either segregating or integrating the gain of increased consumption with the loss of increasing debt.

Designation of funds in separate mental accounts leads to funds potentially being labelled with varying fungibility or liquidity. Accordingly, consumers assign funds to mental categories that either encourage them to spend routinely or discourage spending (Shafir and Thaler, 2006). Differing tendencies to place funds in restrictive mental accounts is observed in self-control problems (Thaler, 1999), leading to the planner-doer concept of mental accounting (Thaler and Shefrin, 1981; Shefrin and Thaler, 1988). In the broader field of behavioral economics, lack of self-control has received considerable attention (DellaVigna, 2009). These studies of nonstandard preferences frame

⁵Referred to as a *total account* and a *recurrent budget period account* by Ranyard et al. (2006).

the phenomenon through the lens of present bias (O'Donoghue and Rabin, 1999) using various models of time-inconsistent preferences (Frederick et al., 2002). Naturally, present-biased preferences have been shown to correlate with credit card borrowing (Meier and Sprenger, 2010) and cardholder spending behavior (Agarwal et al., 2013). Studying the distribution of gains or losses across time has also been proven to influence evaluations for various experiences. When combining events that are unpleasant, the intensity at the most extreme moment of pleasure or pain (the peak) and the final moment (the end) contribute most to our evaluations of the experience (Ariely, 1998; Fredrickson and Kahneman, 1993; Redelmeier and Kahneman, 1996; Carmon and Kahneman, 1996). The peak, duration, and end effects have also been shown to be significant in monetary decisions (Langer et al., 2005), supporting the notion of time influencing the evaluation of payment and utility (Prelec and Loewenstein, 1998).

In conclusion, examining mental accounting studies of monetary or financial decisions related to credit borrowing indicates that nonstandard preferences and nonstandard decision-making influence decisions in three ways. First, consumers tend to segregate or integrate utility and payment at different rates. Thaler (1999) refers to this tendency as payment decoupling, where payment salience is suggested to be a significant driver of credit uptake (Prelec and Loewenstein, 1998). Second, tendencies to employ mental accounts, instead of comprehensive evaluations of economic decisions, will likely differ and influence credit uptake. Mental accounts are heuristics, which, on average, could lead to suboptimal choices. The use of these heuristics is likely stable, leading to persistently inept decision-making, which could influence cardholder behavior. Third, consumers employ mental accounts which restrict or encourage spending, leading to observed differences of present bias. These self-control problems are related to credit uptake (Meier and Sprenger, 2010), and could influence cardholder behavior.⁶ These concepts, payment

⁶Using a broad definition of the mental accounting framework (Thaler, 1999), other

decoupling, decision-making ineptitude, and present bias, are examined in the following sections and hypotheses related to cardholder behavior are developed. Lastly, section 2.1.6 presents a hypothesis controlling for behavior consistent with a rational model of consumer decision-making, suggesting that financial risk aversion will influence cardholder behavior.

2.1.3 Payment Decoupling

Pertinent to mental accounting is the decoupling of payment and consumption. As Thaler (1999) notes, credit cards allow customers to postpone payment and thereby to separate payment from consumption. Customers prefer to consume without invoking thoughts of the cost (Prelec and Loewenstein, 1998), *ceteris paribus*, so, besides allowing for liquidity flexibility, credit card payment is attractive because of decoupling. When payments are decoupled from consumption, the cost of the item becomes much less salient. As Soman (2001a) shows, customers have a harder time remembering the cost of an item paid with credit than with cash. Extending this work, Soman and Gourville (2001) investigate the decoupling of consumption through price bundling. Perhaps the most relevant finding concerns the extent of coupling; Soman and Gourville (2001) demonstrate that the likelihood of decoupling is higher when the act of doing so is cognitively simpler or when the alternative is more attractive. Having a prepaid bundle of theater tickets, subjects were less likely to attend the final play when faced with a more attractive alternative (a party versus helping someone move) or when inferring the cost of individual plays was difficult. Thus, using decoupling as a crutch to justify haphazard consumption is likely more prevalent when doing

factors such as bracketing of several decisions or reference dependence could be justifiably included in a model of credit card borrowing. However, these concepts and several others require knowledge of individual value functions or decision structures, making their inclusion in a model of observed choice prohibitive. In addition, several mental accounting concepts share theoretical foundations, such that operationalization of the concepts with sufficient discriminant validity would be challenging. The concepts examined in this thesis have shown some influence on monetary decisions.

so entails less effort or is more attractive. Soman (2001*a*) speculates that the delayed payment of a credit card bill will decouple payment and utility, which, along with the bundling of several purchases, will lead to more consumption.⁷ Equivalently, the pain of paying that is experienced when paying at point-of-purchase likely has a significant effect on consumption and tendencies of habitual overspending. Kivetz (1999) illustrates how the pain of paying is reduced when consumers are allowed to explain their purchases. In addition, purchases that are labeled as frivolous, unnecessary, or discretionary spending will be systematically underconsumed when they are difficult to rationalize or for self-control reasons (Thaler, 1985). However, the effects of mental accounting often mitigate the effects of the actual cost, as Shafir and Thaler (2006) illustrate with wine collectors who tend to dismiss the investment cost of wine as time passes. Applying different mental accounts to expenses (such as classifying seemingly unnecessary purchases of expensive wine as an investment) will permit the decision maker to decouple payment from utility.

A different aspect of decoupling payment and utility is examined in Hirst et al. (1994). Specifically, they discuss temporal contiguity of borrowing and consumption, the extent of which the terms of a loan correspond to the life of a good. As expected, and predicted in Thaler (1985), consumers are more likely to prefer loan terms that correspond to the asset's useful life. Consumers choose a car loan that coincides with the expected utility the car provides, how long they expect to keep the car, and the eventual resale value. Hirst et al. (1994) explain this phenomenon with the concept of integrating gains and losses (Thaler, 1985) based on the value function in prospect theory

⁷The same effect of decoupling can be found with pre-paid pricing policies of resort vacations like Club Med (Thaler, 1980). Here, consumption while on vacation will be driven by the sunk cost effect, defined as a "greater tendency to continue an endeavor once an investment of money, time or effort has been made" (Arkes and Blumer, 1985, p. 124), which should reinforce and beget extravagant behavior. According to traditional economic theory, we should ignore sunk costs. Only incremental and future costs should be considered when considering options.

(Kahneman and Tversky, 1979). As noted in Hirst et al. (1994), “Because of the value function assumption, jointly valuing, or integrating, gains and losses enhances the utility of credit purchases.” A durable good purchased using credit will continue to provide utility, offsetting the loan payments. Payments coinciding with utility are even shown to be preferred such that consumers are willing to pay for it. Decoupled debt, specifically consumer loans, can be shown to have a greater negative effect on psychological well-being. Brown et al. (2005) use survey data to show that non-mortgage debt increases psychological distress, while mortgage loans do not have a significant effect on well-being. Persistent payments will be perceived as losses when they are not justified by a concurrent utility experience.⁸ Considering that the value function is convex for losses (Thaler, 1985; Kahneman and Tversky, 1984), the relative pain of paying with credit (and increasing the credit balance) will eventually be lower than the relative disutility of the transaction. Accordingly, it can be argued that the loss frame activated by consecutive months of credit payment might be conducive to continued borrowing. Carrying a substantial credit card debt could motivate certain consumers to offset the disutility of credit payments with unrelenting (credit financed) consumption, spiraling deeper into a futile process of increasing debt.

Examining the intricacies of mental accounting of loans, Kamleitner and Kirchler (2006) conduct a qualitative study of loan perceptions. Consumer loans are indeed shown to be coupled with the good at the time of the purchase, but this deteriorates over time.⁹ Kamleitner et al. (2010) use samples of mortgage loans and consumer loans to show that associations between the loan and utility provided clearly dominate perceptions of loans, as suggested

⁸See also Krichler et al. (2008) for a short review on decision-making related to borrowing in households. Noted is the possibility of disjunct coupling of payments and utility between partners.

⁹Kamleitner and Kirchler (2006) separate the concepts of hedonic editing, integration and double-entry mental accounting. Though they approach mental accounting slightly differently, the similar concepts will be treated as equal in this thesis.

in previous research (Kamleitner and Kirchler, 2006). When consumers elicit thoughts of their car, the car loan is not necessarily salient. However, when consumers elicit thoughts of their car loan, they often associate this with the car. Also, when associations between the utility provided and the loan were strong, the perceived loan burden (both financially and psychologically) increased (Kamleitner et al., 2010). The structure of loan repayment is also a significant contributor to the utility, or disutility, of a loan. Hoelzl et al. (2011) investigate loan repayment plans, showing how psychological effects outweigh rational financial considerations. Subjects were shown to prefer a constant or decreasing payment profile, even when total cost favored a rising payment plan. This result was consistent across goods purchased, duration, and subject focus (emotional or financial).

The coupling of payments and utility has also been investigated from a dynamic perspective. Hoelzl et al. (2009) reveal that the association between the utility provided and the loan (integration in their terms) seem to change over time, in addition to being heterogeneously distributed among individuals. Credit card debt is usually a mix of past transactions, encouraging higher tendencies of decoupling than for loans financing singular goods. In addition, the eventual utility provided by the goods financed by credit depreciates over time. While credit cards (compared to cash) decouple payment intrinsically and promote more spending (Raghubir and Srivastava, 2008), the effect is likely further instigated by decoupling the borrowed amount from the utility provided.

Expressing decoupling tendencies by applying mental accounting notation demonstrates how the likelihood of delinquency increases. The decision of credit uptake will be formulated using the value function (Thaler, 1985), as introduced by Tversky and Kahneman (1981). As opposed to a utility function, the value function expresses outcomes as perceived gains or losses. Individuals are assumed to use subjective weights as functions of probabilities, and the function assumes a S-shape (Tversky and Kahneman, 1992).

The chosen outcome in a value function will depend on the individual perceptions of the inputs and evaluations of the possible outcomes. A decision to borrow d_t to approximate consumption that is affordable C_t to desired consumption level $C_t^* = C_t + d_t$ is made if the outcome is valued higher than reduced consumption, $V(C_t + d_t) > V(C_t)$. Or, equivalently, borrowing to achieve the desired consumption level is valued higher than reducing consumption: $V(C_t^*) > V(C_t^* - d_t)$. Adding future debt d_t to the equation, using δ as a discount factor, λ as the cost to borrow, and r as a factor of intended repayment yields the following equation:

$$V\left(C_t + d_t - \lambda r \sum_{t=1}^T \delta^t d_t\right) > V\left(C_t\right). \quad (2.1)$$

The choice to borrow is made if the value of the cost of debt is lower than the value of reduced consumption. In the following month, assuming the individual chooses to borrow, the value function expands to include the accumulated debt

$$V\left(C_t + d_t - \lambda r \sum_{t=1}^T (\delta^t d_t) - r_l \alpha^1 d_{t-1}\right) > V\left(C_t - r_h \alpha^1 d_{t-1}\right), \quad (2.2)$$

where the repayment rate r_l of current debt d_{t-1} is assumed to be low when continuing to borrow or higher r_h if intended debt repayment is initiated. The decoupling factor, designated α , is a factor proportional to how efficiently the debt is linked to utility and subsequent salience of the credit debt and payments. While Prelec and Loewenstein (1998) partition this into a buffering and attenuation effect when modeling payment decisions, the model presented here of borrowing and repayment deals with aggregates of payments. The decoupling effect is likely heterogeneously distributed among individuals and between payments (e.g., paying for durables versus haphazard hedonic consumption). Debt accrued, which is loosely coupled to current utility, gives lower pleasure of payment compared to paying for items that

offer current utility (Prelec and Loewenstein, 1998). Thus, the value of repaying is reduced and the subsequent amount of debt repaid is reduced. Extending equation 2.2 to a general decision of continued current borrowing or reducing consumption for M preceding months of borrowing exposes the pivotal role of decoupling:

$$V\left(C_t + d_t - \lambda r \sum_{t=1}^T (\delta^t d_t) - r_l \sum_{m=1}^M (\alpha^m d_{t-m})\right) > V\left(C_t - r_h \sum_{m=1}^M (\alpha^m d_{t-m})\right). \quad (2.3)$$

Individuals who have a moderate decoupling rate (along with a strong desire to uphold desired consumption) will choose to continually borrow without approaching delinquency. These are the debt revolvers, who sufficiently repay their debt when reaching a certain threshold:

$$V\left(-\lambda r \sum_{t=1}^T \delta^t - r_l \sum_{m=1}^M \alpha^m d_{t-m}\right) = V\left(-d_t - r_h \sum_{m=1}^M (\alpha^m d_{t-m})\right). \quad (2.4)$$

This equilibrium is reached when the utility of the sum of discounted future payments and minimal repayments is equal to the utility of reduced consumption and higher repayments. In the words of Prelec and Loewenstein (1998), debt is reduced when the utility of exorbitant consumption is less than the pain of repaying in higher amounts. An individual who borrows indiscriminately to approach C_t^* will repay at a minimal rate r_l and has a decoupling rate a approaching zero. Imposing a budget limit W reveals the corner solution, whereby the individual is forced to initiate repayment, by virtue of being unable to honor the minimal payment:

$$W_t = V\left(C_t + d_t - \lambda r \sum_{t=1}^T \delta^t d_t - r_l \sum_{m=1}^M (\alpha^m d_{t-m})\right). \quad (2.5)$$

This equation constrains the individual decoupling rate to $\alpha = 1$, reflecting actual repayment capability (which, incidentally, corresponds to the level

expected to be observed in a rational cardholder), and future borrowing d_t is irrelevant. As a corollary, this illustrates that past decisions to borrow are prudent when an individual's valuation of the decoupling parameter and discounting factor approaches unity. By virtue of a high decoupling rate (illustrated here), the cardholder is rendered illiquid and the account is marked as delinquent.

Kivetz (1999) notes that credit cards are especially conducive to payment decoupling, and they promote overspending (Prelec and Simester, 2001). To further beget decoupling, the challenge of retrospective evaluation of credit card transactions is more difficult than with cash transactions (Soman, 2001*a*). Following the predicted effects of decoupling, a weaker association between payments and utility will reduce motivation to repay. Individuals who have a high decoupling rate, α , are more prone to borrow indiscriminately, eventually facing repayment problems and delinquency (as shown in equation 2.5):

H1: The likelihood of delinquency increases as the rate of decoupling increases.

2.1.4 Decision-making Ineptitude

Using application scores as a heuristic for underlying behavior, personality traits have been linked to financial decision-making aptitude. Individuals with higher credit scores have been shown to exhibit higher levels of financial knowledge and an internal locus of control (Perry, 2008; Lusardi and Mitchell, 2014). Similarly, our financial behavior may be closely linked with our biopsychological profile, suggesting that traits such as sensation seeking, risk judgment, and risk appraisal induce proportional credit scores (Brockett and Golden, 2007). Rick et al. (2008) show significant differences in credit card debt and savings among “tightwads” and “spendthrifts,” in which the latter have a lower anticipatory pain of paying. Indeed, “spendthrifts” are more insensitive to future expenses, neglecting them when anticipating future

financial slack (Berman et al., 2016).

Moreover, financial prudence related to credit card usage has been connected to such elemental characteristics as cognitive abilities (Agarwal and Mazumder, 2013),¹⁰ cognitive reflection (Frederick, 2005), and financial knowledge (Hilgert et al., 2003). Cardholders who rarely repay their balance in full report lower confidence in managing finances and do not think setting a budget is important (Shefrin and Nicols, 2014). Revolving the credit balance has also been linked to general traits, such as self-control, self-esteem, self-efficacy, and internal locus of control (Wang et al., 2011). Holding debt in and of itself has been suggested to be part of a pattern of dysfunctional economic behavior, such as weak money management (Lea et al., 1995). In addition, attitudes about debt shift toward higher debt tolerance when individuals borrow (Davies and Lea, 1995), reinforcing beliefs about carrying debt.

It is expected that a lower financial understanding and ability to process financial decisions is partially enduring, giving rise to permanent differences in decision-making aptitude. As a consequence, cardholders resort to mental accounting using topical accounts when they are unable to administer a comprehensive accounting decision process to consumption dilemmas. Decisions will, to a greater degree, involve sequential evaluation of decisions as well as framing consumption decisions into incidental categories. The distribution of suboptimal outcomes generated by a mental accounting process is therefore expected to be heterogeneous among individuals.

Tendencies to overspend (or sufficiently adapt consumption) is consequently state dependent, whereby current desired consumption C_t^* is dependent on past desired consumption C_{t-1}^* . Representing the degree of persistence in spending with h and discretionary current spending with a_t , desired current consumption is a function of habit and monthly variation:

¹⁰The authors used previous cognitive test scores, where math proficiency led to greater financial understanding.

$C_t^* = h(C_{t-1} + d_{t-1}) + a_t$. Note, current desired consumption is still assumed to require credit financing $C_t^* = C_t + d_t$. Extending equation 2.3 where past consumption enters the model using a distributed lag approach, the following equation illustrates the effect of the preceding month on current consumption:

$$V\left(h(C_{t-1} + d_{t-1}) + a_t - \lambda r \sum_{t=1}^T (\delta^t d_t) - r_l \sum_{m=1}^M (\alpha^m d_{t-m})\right) > \quad (2.6)$$

$$V\left(hC_{t-1} + a_t - r_h \sum_{m=1}^M (\alpha^m d_{t-m})\right).$$

Persistent desires of indiscreet spending influence delinquency likelihood proportional to the persistence parameter (and level of borrowing). Expanding the equation to accommodate a (more realistic) prolonged history of purchases, including M months of borrowing, yields:

$$V\left(\sum_{m=1}^M (h^m(C_{t-m} + d_{t-m})) + a_t - \lambda r \sum_{t=1}^T \delta^t d_t - r_l \sum_{m=1}^M \alpha^m d_{t-m}\right) > \quad (2.7)$$

$$V\left(h^m \sum_{m=1}^M (C_{t-m}) + a_t - r_h \sum_{t=m}^M \alpha^m d_{t-m}\right).$$

The decision to continue borrowing or desist and repay at accordingly higher levels is dependent on the sum of past consumption and magnitude of habitual persistence, that is, the sum of the $h^m(C_{t-m} + d_{t-m})$ components. Again, imposing a budget constraint W_t reveals the corner solution:

$$W_t = V\left(\sum_{m=1}^M (h^m(C_{t-m} + d_{t-m})) + a_t - \lambda r \sum_{t=1}^T \delta^t d_t - r_l \sum_{m=1}^M \alpha^m d_{t-m}\right), \quad (2.8)$$

where this case illustrates the effect of habit persistence on delinquency likelihood. As in equation 2.5, the decoupling parameter is constrained to $\alpha = 1$

and future payments are disregarded to reflect actual repayment capability; delinquency is achieved here by virtue of persistently careless borrowing. As intermittent financial carelessness (a temporary high d_t value borrowed) is tolerated by most individuals, repayment difficulties will usually require habitual carelessness in the form of the h parameter approaching 1.

It is to be expected that losses in the form of the sum of expenses charged, interest, and fees are only fully realized when the value exceeds some arbitrary threshold set by the cardholder. Put differently, debt accretion will continue until the cardholder reaches a certain self-imposed limit. Using the logic put forth in Kahneman and Tversky (1979), this limit is probably where the loss function flattens relative to the gains function; continued borrowing for discretionary consumption (gains) will give less utility than refraining from the inverse loss. This explains why individuals eventually decide to stop borrowing, although some stop too late.

Individuals who borrow indiscriminately display behavior diverging from what is suggested by liquidity constraint hypotheses or consumption smoothing, indicating that their behavior as cardholders will eventually lead to delinquency. Delinquent cardholders are likely to have low financial understanding, routinely substituting patient deliberation by applying mental accounts. It is expected that their inferior internal locus of control will not hinder their continued overspending. Financial decision-making ineptitude induces persistently careless spending, as a result of lower financial understanding and poor financial decision-making skills in general. Accordingly, the following effect is expected:

H2: The likelihood of delinquency increases as the rate of decision-making ineptitude increases.

2.1.5 Present Bias

In addition to personality characteristics, individual self-control, in the form of acting as a planner or a doer, is a fundamental part of the mental account-

ing framework (Thaler and Shefrin, 1981; Shefrin and Thaler, 1988). Accumulating credit card debt and the decision to spend or refrain from spending can be looked at as individual differences in the propensity to plan (Ameriks et al., 2003), the usage of personal rules and willpower (Bénabou and Tirole, 2004), or as an internal duality between the myopic doer and patient planner (Ali, 2011). Neurological studies have confirmed the heterogeneity in planning; Raab et al. (2011) show how compulsive shoppers associate purchases with affect, and McClure et al. (2004) illustrate how immediate and delayed monetary gratification activate different neural systems.

Maintaining foresight and achieving long-term goals have proven to be difficult (Latham and Seijts, 1999), especially when dealing with debt repayment (Gal and McShane, 2012; Amar et al., 2011). Smaller debts are repaid quicker, perhaps due to “small victories” giving quick motivational gains (Brown and Lahey, 2015). The ability to forego immediate gratification is nevertheless hypothesized to be both specific to individuals (Frederick et al., 2002) and specific to situations (Brown et al., 2009).

Present bias is naturally related to the way in which individuals account for time and is an important predictor of decision-making (Atlas et al., 2017).¹¹ The ambiguity of accounting time has been shown to affect risk taking (Okada and Hoch, 2004), while the sunk cost effect of time invested is less prevalent than for monetary costs (Soman, 2001*b*; Soster et al., 2010). Accordingly, managing future consequences often requires an assessment of both cost and gratification. Relevant to credit cards, deferring utility gives a greater decline in discounting rates than when expediting utility (Malkoc and

¹¹The discounting rate has been shown to be time dependent, giving rise to hyperbolic functional forms of discounting (see e.g., Laibson, 1997). Models of time inconsistent preferences have even been shown to support a reversal of present bias when choices are immediately available (Sayman and Öncüler, 2009). For the sake of simplicity, the discounting factor here is expected to be linearly time-dependent, although it is expected to be heterogeneously distributed among decision makers. In other words, present bias is expected, and some are more impatient than others. Formally, the δ parameter presented here conceptually encompasses the beta-delta model suggesting hyperbolic or similar functional forms.

Zauberman, 2006). The subjectivity of time processing is likely a determinant of heterogeneous discounting (Zauberman et al., 2009), and processing time is seemingly affective (Lee et al., 2015), leading to a difficult or perhaps complex dilemma for the decision maker. Elaborating on the difficulty of understanding one’s own time preferences, O’Donoghue and Rabin (1999) show that naive people are prone to procrastination (facing a task with immediate costs). Some research has focused on a possible misalignment between the current and future self, where assisted alignment of future and current self seems to simplify the necessary empathy and understanding to make correct decisions (Hershfield et al., 2011; Bartels and Urminsky, 2015).

Using credit and debit card transaction data, Agarwal et al. (2013) illustrate the effects of mental accounting on credit card borrowing. Using data that have different statement dates, and hence different grace periods, they show that individuals treat each payment form as belonging to different categories. Credit card borrowing, specifically consumption of discretionary retail spending, is 10% higher per day of the week following the credit card statement. This trend is independent of debit card spending, supporting the notion of credit and debit being registered in different mental accounts. Of course, this is also consistent with cardholders’ propensity toward instant gratification (Meier and Sprenger, 2010). Agarwal et al. (2013) also show that, contrary to the hypothesized effects of liquidity constraints driving credit card borrowing, cardholders who generally pay off their debt when due are more affected by the statement date. Conversely, debt revolvers are less affected by statement date, supporting the mental accounting hypothesis.

Examining the decision to continue financing consumption using credit (or to commence repayment), the discounting parameter of future costs, δ , is considered:

$$V\left(C_t + d_t - \lambda r \sum_{t=1}^T \delta^t d_t - r_l \sum_{m=1}^M \alpha^m d_{t-m}\right) > V\left(C_t^* - d_t - r_h \sum_{m=1}^M \alpha^m d_{t-m}\right). \quad (2.9)$$

Given a higher discounting rate, $\delta^t = 0$, the value of future repayment is reduced, approaching irrelevancy. Given a decoupling rate α approaching 1 and disregarding differences of repayment rate when considering the option to continue borrowing, $r_l = r_h$, the decision to borrow simplifies to a valuation of current credit uptake and future repayment stream:

$$V\left(d_t - \lambda r \sum_{t=1}^T (\delta^t d_t)\right) > 0. \quad (2.10)$$

Thus, by virtue of valuing current consumption more than future repayment, individuals who exhibit a higher concern for immediate consequences are expected to carry more credit card debt (as shown with survey data in Joireman et al., 2010) and have a higher likelihood of default. For some individuals, present bias eschews future consequences and causes consistently higher rates of borrowing (Meier and Sprenger, 2010). The following effect of present bias is expected:

H3: The likelihood of delinquency increases as the rate of present bias increases.

2.1.6 Financial Risk Aversion

The final hypothesis controls for a central tenet of a rational model of decision-making, namely the stability of preferences and long-run planning suggested by the life-cycle hypothesis (Ando and Modigliani, 1963). Assuming consumption smoothing requires forward-looking individuals who anticipate difficult times and eventual income shocks. This implies some risk-aversion, at the minimum suggesting that borrowing does not exceed a crucial threshold. Individuals who are risk-averse planners (Ameriks et al., 2003) are expected to handle fluctuations in income better than individuals who have difficulty judging risk (Brockett and Golden, 2007). Risk taking itself is believed to sig-

nificantly contribute to credit scores (Brockett and Golden, 2007), explaining why the credit score is a useful heuristic of risk when assessing probability of insurance loss (Wu and Guszcza, 2003). The same mechanisms will hinder financially prudent individuals from borrowing excessively. Therefore, the following hypothesis not only directly controls for exogenous shocks to finances but also indirectly accounts for several behavioral aspects consistent with a rational model of consumer behavior:

H4: The likelihood of delinquency decreases as the rate of financial risk aversion increases.

The next section details classification methods used for behavior scoring.

2.2 Classification Methods in Credit Scoring

This section will examine credit scoring in general and behavior scoring in particular. Prevailing methods are examined, and a selection of these methods will be applied to serve as baseline comparisons to the model suggested in this thesis.

Classifying risk of default in retail credit has received mounting interest in recent years (Lessmann et al., 2015). To reiterate, credit scoring consists of classifying both credit card applicants (referred to as application scoring) and classifying existing cardholders (referred to as behavior scoring) to identify customers in risk of delinquency. While research appears to overwhelmingly focus on application scoring, this attention is likely fueled by the availability of high quality public datasets of credit applicants. Along with advancements in data mining and artificial intelligence methods, this has permitted the development of novel methods of classification while allowing convenient comparison with extant models (see Louzada et al., 2016, for a review of methods). Credit risk classifiers have evolved from classical statistical methods, such as logistic regression (Hamilton and Khan, 2001), to combinations of several artificial intelligence algorithms.

Behavior scoring has not received the amount of attention or seen the breadth of techniques as application scoring has, though the methods employed have been similar. As such, the classifiers that have been successful and prevalent in application scoring should also be of interest in behavior scoring contexts. The similarities of the applications are numerous, including similar predictor variables and observed outcome. The following section will briefly explain popular classification methods of credit risk and applications thereof to behavior scoring data. As the number of studies examining application scoring easily eclipse those of behavior scoring, classifier success and prevalence is judged based on applications in both application and behavior scoring. Accordingly, prevailing techniques that have shown success in behavior scoring and application scoring are chosen as baseline models for comparison with the proposed model. An overview of extant behavior scoring applications is presented in table 2.1.

2.2.1 Single Classifiers

Regression techniques employed in behavior scoring include logistic regression and Tobit regression. In application scoring, logistic regression is often used as a baseline comparison model when predicting default (Louzada et al., 2016), while early applications of logistic regression in behavior scoring include examples such as identifying debt revolvers (Hamilton and Khan, 2001). Tobit regressions have been applied in a similar way. Min and Kim (2003) employ Tobit regressions to predict which households will choose to borrow and how much they will borrow. Extending the classical Tobit II specification, Zhao et al. (2009) provide a more flexible model, incorporating dynamics and consumer heterogeneity. Using proprietary monthly credit card data, including cash advances, expenses, credit limit, total balance, and delinquency (defined as not meeting the minimum payment in a month), these usage variables are linked to the actual amount paid. Using individual-level data with monthly credit card statements, they specify a Tobit II model

such that delinquent cardholders can be separated into high-risk and low-risk groups. Their results support the model specification, showing substantial learning effects (through card age), different carryover effects for parameters, and a substantial amount of unobserved heterogeneity. Consumer oversight is found to cause nonpayment in approximately 5% of the cases. In a prediction and model comparison, their model specification significantly outperforms the industry standard,¹² first and foremost due to dynamics, but heterogeneity also impacts prediction accuracy.

Nearest neighbor methods also represent a popular class of baseline comparison models in credit classification applications. These techniques are instance-based learners, comparing each test case to training cases and finding the class among the k nearest neighboring (k -NN) training cases. Performance of the k -NN algorithm has, in some cases, approached classification rates similar to artificial neural networks (ANN; Koklu and Sabanci, 2016) and several other data mining techniques (Yeh and Lien, 2009). However, in application scoring, k -NN has (perhaps expectedly) been outperformed by more advanced artificial intelligence algorithms (Louzada et al., 2016; Lessmann et al., 2015), suggesting that the aforementioned results could be a statistical artifact stemming from that particular dataset.¹³

ANN represent a collection of methods that construct models in a fashion comparable to how the human brain is understood. Specifically, input variables (explanatory variables) are related to output variables (response

¹²The industry standard model as defined by Zhao et al. (2009) corresponds to a logistic regression that accounts for heterogeneity. While logistic regression and discriminant analysis was previously widely applied (Hand and Henley, 1997), the proliferation of machine learning methods (Lessmann et al., 2015) could suggest that their popularity has waned.

¹³Both Yeh and Lien (2009) and Koklu and Sabanci (2016) use the “default of credit card clients data set” available from the UC Irvine Machine Learning Repository and Kaggle. The data was collected over 12 months from 2005 to 2006 in Taiwan. The performance of data mining techniques (especially using single classifiers) can be sensitive to the data used, which could explain the exceptional performance of the k -NN algorithm when acknowledging the usual moderate performance in application scoring.

variables) through a structure of nodes. Specifying the optimal structure of a ANN involves, in particular, deciding the number of hidden layers of nodes, the number of nodes in each layer, and the activation functions for the nodes. The performance of ANN is typically sensitive to the architecture and activation function specified, which also holds true for credit risk applications (Šušteršič, Mramor and Zupan, 2009; Louzada et al., 2016). In the area of behavior scoring, ANN have repeatedly outperformed other classification methods using individual classifiers (Yeh and Lien, 2009; Koklu and Sabanci, 2016; Pasha et al., 2017; Neema and Soibam, 2017).¹⁴ ANN are also used in sequential schemes. Ha and Krishnan (2012) predict delinquency and repayment with a proportional hazard model for different segments mapped using ANN. While singular classifiers of credit risk using ANN typically entails feedforward back-propagation networks with different structures (Louzada et al., 2016; Šušteršič, Mramor and Zupan, 2009), recent developments have focused on combining classifiers (Lessmann et al., 2015). Combining ANN classifiers into ensemble learning models has been shown to improve accuracy (Yu et al., 2008).

2.2.2 Classifier Ensembles

Combinations of classifiers using ensemble methods have become the prevailing classification methods in assessing credit risk (Louzada et al., 2016). Common ensemble methods using homogeneous classifiers include bagging (Breiman, 1996) and random forest (Breiman, 2001), while boosting methods (Freund and Schapire, 1996) are most prevalent when assessing credit risk (Lessmann et al., 2015). Boosting methods employ a base learning algorithm, which is invoked many times and applied to samples of the training set. The

¹⁴As with the k -NN algorithm, the same reservations apply: these results are gleaned from the same dataset from Taiwan credit card clients available at the UCI repository and Kaggle, and prediction performance could be specific to the dataset. While Neema and Soibam (2017) found that ANN balanced risk and number of potential customers, random forest provided the lowest portfolio cost.

Adaboost family of algorithms is a prevalent boosting method that reweights each sequential subsample by giving previously misclassified cases more influence. Introduction of stochastic gradient boosting (GBM, Friedman, 2002), allowed boosting algorithms to adapt to any loss function, updating the base learning algorithm conditional on the current loss function gradient. More recently, the XGBoost algorithm implemented a regularized gradient function to increase performance and avoid overfitting (Chen and Guestrin, 2016). Casually surveying the winning solutions in machine learning competitions, such as Kaggle, suggests that XGBoost is considered the algorithm du jour for classification problems. In conjunction with the often high classification rates of boosting methods, they have also become common in application scoring (Louzada et al., 2016; Lessmann et al., 2015) and behavior scoring (Hamori et al., 2018; Khandani et al., 2010). Similar to boosting, stacking learning algorithms into either homogeneous ensembles (SE) or heterogeneous ensembles (HSE) applies several models to the training data according to a metalearning algorithm. While SE classification methods are common, applications of HSE have shown promise in predicting credit risk (Louzada et al., 2016). HSE combine diverse classifiers using different techniques to arrive at the optimal ensemble. In application scoring, HSE using a simple or weighted classifier averaging as well as more advanced sequential classifier selection have been proven successful (Lessmann et al., 2015).

While the aforementioned classification methods represent the techniques most common to behavior scoring, several others have been applied to application scoring problems. Linear discriminant analysis (along with regression) have historically been the most widely used classification techniques in application scoring (Hand and Henley, 1997). As linear discriminant analysis assumes multivariate normality and equal class variance-covariance matrices, the technique is often relegated to providing baseline classification rates.¹⁵ Until the more recent inception of ensemble methods, renditions of

¹⁵Quadratic discriminant analysis may be used when the assumption of equal class

support vector machines (SVM) were often proposed and commonly preferred (Louzada et al., 2016). However, individual SVM classifiers have exhibited moderate performance in comprehensive evaluations of application scoring classifiers (Lessmann et al., 2015), and the technique has received little attention in behavior scoring. Other possible classification methods include Naive Bayes, Bayesian network, classification trees (often used as the base learner in boosting), and extreme learning machines. As with SVM, these classifiers have been largely neglected in behavior scoring.¹⁶ Methods dealing with self-selection problems such as matching (Gensler et al., 2013), could be applied to assess repayment or cardholder value, although these issues are not considered relevant to this thesis. Table 2.1 provides an overview of behavior scoring research along with the current thesis, highlighting their classification methods and results.

In conclusion, based on reviews of extensive research in application scoring (Lessmann et al., 2015; Louzada et al., 2016), and a broad survey of behavior scoring methods, a handful of methods appear most relevant. This includes neural networks (ANN), boosting algorithms (GBM and XGBoost), and stacked ensembles combining classifiers (ANN-SE, GBM-SE and HSE). Performance of the proposed model (presented later, specified as Hierarchical LR w/ k -NN in table 2.1), will be juxtaposed with these methods. See section 3.3.2 for a detailed specification of the chosen machine learning classifiers.

variance-covariance matrices is violated. In application scoring, avoiding multicollinearity by employing correlational-based feature selection (Lessmann et al., 2015) may help linear methods, such as discriminant analysis and regression methods.

¹⁶Other aspects of the classification process, such as feature selection, feature generation and data transformation will not be described here. These issues are central to application scoring models, due to the usually large set of features available (see Oreski et al., 2012, for an implementation of feature selection).

Author(s)	Description of data	Method(s)	Dependent variable; key finding(s)
Dunn and Kim (2006)	Survey	Probit Regression	6-month delinquency; minimum required payment highly significant
Min and Kim (2003)	Survey	Tobit I & II	Borrowing decision and balance; opposing effects of variables
Tan et al. (2011)	Survey	Ordered Tobit	Yearly balance; past behavior predictive
Hamilton and Khan (2001)	Monthly credit transactions and demographics	DA, LR	Debt revolvers; behavior variables highly predictive
Zhao et al. (2009)	Monthly credit transactions	Tobit II	Delinquency and repayment; heterogeneity with dynamics improve prediction
Khandani et al. (2010)	Credit and debit transactions with financial data	GBM	90-day delinquency; model highly predictive of future delinquency
Yeh and Lien (2009)	Monthly credit transactions and demographics ^a	k -NN, LR, DA, NB, CT, ANN	Payment default; ANN outperforms other classifiers
Koklu and Sabanci (2016)	Monthly credit transactions and demographics ^a	k -NN, ANN	Payment default; ANN slightly better classification performance
Ha and Krishnan (2012)	Monthly credit transactions and demographics	ANN segments, Cox regression	Repayment of delinquent debt; model predicts recovery and repayment
Pasha et al. (2017)	Monthly credit transactions and demographics ^a	k -NN, ANN, DA, NB, LR, CT	Payment default; ANN outperforms other classifiers
Neema and Soibam (2017)	Monthly credit transactions and demographics ^a	ANN, k -NN, DA, LR, CT, NB, RF	Payment default; RF most cost-effective, ANN most accurate
Current thesis	Credit and debit transactions	Hierarchical LR w/ k -NN, ANN, ANN-SE, GBM, GBM-SE, XGBoost, HSE	90-day delinquency; proposed Hierarchical LR w/ k -NN classifier outperforms other classifiers

Table 2.1: Behavior scoring research. Studies marked ^a use the “default of credit card clients” data available at the UCI repository. LR = logistic regression, NB = Naive Bayes, DA = linear discriminant analysis, k -NN = k nearest neighbor classifier, CT = classification trees, RF = random forest, ANN = artificial neural network, GBM = gradient boosting machines, SE = homogeneous stacked ensemble, XGBoost = extreme gradient boosting, HSE = heterogeneous stacked ensemble.

3 — Methodology

The following chapter presents the process of cleaning and filtering the data, graphical descriptions of the data, and the construction of variables which are operationalized to capture the hypothesized effects. Specification and estimation of the proposed model is presented in section 3.3, along with specification of the competing machine learning algorithms and classification performance metrics.

3.1 Data and Selection of Variables

Credit card data was collected from a large Norwegian credit card lender. Customers were randomly drawn, although the sample was constructed to contain a higher ratio of customers who defaulted on their credit card account. Specifically, data was extracted from two databases: a database of credit card transaction information and a database with monthly cardholder status and information. In total, 90,247 unique accounts were selected, held by 31,271 unique customers. The daily transaction database consisted of 17,143 unique customers and 25,416 unique credit card accounts, where 2,762 accounts were classified as being delinquent at least once.¹ This gives a delinquency rate of 10.8%, which is considerably higher than actual delinquency

¹An account is classified as being delinquent when the cardholder carries a positive balance, and neglects payment for three calendar months.

rates among cardholders.² The deliberately high proportion of delinquent cardholders allows for a more sophisticated sample selection, without concerns about too few observations.

3.2 Sample Selection

The raw sample of cardholder usage data consisted of 1,883,799 transactions. This encompassed all transactions made in the period June 1, 2008 through June 31, 2011: 51 transaction types in total. Of these, 344,156 transactions were repayments, reimbursements, or compensation, reducing the current balance. Transactions such as yearly credit card fees, card replacement fees, bill payment costs, account transfers, and a multitude of others were filtered out. Only transactions in which the cardholder’s behavior likely had a measurable impact on the monetary amount, such as point of sale payments, bill payments, or teller transactions, were kept for calculating expenses.

These databases were then combined: the monthly data that includes end of month balance, beginning balance, credit limit, default status, payment status, and credit card type, with the daily transactional data. Unique monthly customer observations were then created and joined with aggregated monthly transactional data. In other words, the accumulated expenses were calculated in months when the cardholder had multiple expenses. This was also done for interest, fees, and repayment transactions.

Data was then split into two groups: accounts that were delinquent at least one month and accounts that had no delinquent months. The non-delinquent cardholders were then filtered by removing months when the cardholder carried no balance at the month’s end or beginning. This left 316,177 monthly account observations for non-delinquent customers. In ad-

²As noted in the introduction, current delinquency rates are 3.6%, reported in the fourth quarter of 2018. See the Statistical Release “Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks” by the Federal Reserve at <http://www.federalreserve.gov/releases/chargeoff/> for current delinquency rates.

dition, the non-delinquent accounts were filtered by retaining accounts that resemble “bank credit cards” or “retail credit cards.” This removes accounts that were in reality down payment plans or store-granted credit, where the credit card provider supplied the customer with a credit card in the hopes of increasing borrowing. Finally, accounts that had a shorter than 16-month history were removed.

As noted, an account is classified as delinquent when the cardholder neglects sufficient payment for three consecutive calendar months when carrying a positive balance. In addition to this, repayment status is recorded by tracking how many payment reminders are sent. The cardholder receives two reminders before being classified as delinquent, and the account is then transferred to a collection agency. Delinquency status also renders the account frozen, as all transactions and usage are suspended; even data registration is halted for that particular account. This complicates variable development and modeling, as account histories only provide information on what happens preceding delinquency. Table 3.1, in which a cardholder is declared delinquent in December 2009, illustrates this.

As can be seen in table 3.1, current balance, credit card expenses, fees, and repayment are frozen, while opening balance fluctuates after the account is declared delinquent. After removing accounts with missing credit limits and inactive accounts, the dataset with delinquent accounts totaled 994 unique customers and 1,234 accounts. Account histories were then filtered, removing accounts that had been marked as delinquent within the first five months during the recorded period. Account histories with no variation in balance were also removed, along with accounts that were not bank or retail credit cards. This left a total of 592 accounts across 510 customers as the final sample of delinquent cardholders for the credit transaction data.

While matching the credit and debit data, it became clear that using the credit account as the unit of analysis was misguided. Over half of the customers held at least two credit cards, and some customers even had eight

Table 3.1: *Sample delinquent cardholder monthly transaction history*

Month	Current Balance	Opening Balance	Credit Limit	Expenses	Repayment	Fees
02-2009	11763	12495	25000	0	0	250
03-2009	12073	12752	25000	0	0	309
04-2009	12329	13071	25000	0	0	256
05-2009	11788	13338	25000	0	-797	256
06-2009	10098	11801	25000	0	-2000	310
07-2009	10323	12056	25000	0	0	225
08-2009	10549	11298	25000	0	0	225
09-2009	10778	11536	25000	0	0	229
10-2009	11012	11282	25000	0	0	233
11-2009	11309	11525	25000	0	0	296
12-2009	11552	11829	25000	0	0	242
01-2010	11493	12081	25000	0	-59	0
02-2010	11493	11512	25000	0	0	0
03-2010	11493	11763	25000	0	0	0

This table presents a typical credit card history for a delinquent account. There are few expenses and repayments are sporadic. Fees are consistent, but not necessarily increasing with time. All amounts in NOK. Decimals were removed for readability.

cards registered to their name.³ In addition, the account number is sometimes changed due to database maintenance, such that accounts had to be matched with their physical cards. As a consequence, this model will utilize summed customer data: if a customer holds two cards, each with a balance of 5,000, this is summed to 10,000. This not only makes fitting a useful model easier, it also replicates customer debt handling in a more realistic manner.

After the selection of cardholders, debit data was provided for the portion of cardholders that also used the credit card provider as their primary bank. The debit sample consisted of 3,513 unique cardholders, totaling

³Customers could also hold and borrow on credit cards from other institutions, which would not be accounted for in this dataset.

11,055 unique credit card accounts. The debit data included transaction date, amount, and a significant amount of information regarding the actual transaction, including transaction codes and a variety of character strings. This latter information, though mostly useless for analytical purposes, is employed in creating bank statements and providing information to the account holder while using Internet banking.

The daily debit transactions, 5,742,558 rows in total, were then aggregated to create monthly variables. A total of 114,738 unique monthly cardholder observations were constructed, and monthly sums of payments and income were created. Monthly income was found using the highest observed positive transaction. This precludes adding monthly positive transfers that are not anticipated. Thus, monthly income attempts to capture the most significant positive transaction, as the timing and amount of this transaction is thought to influence cardholder behavior. The income date was then used to calculate payments made in the subsequent 10 days after the highest observed income, in addition to total payments between incomes. Rather than calculating monthly payment sums, this was done to better replicate how cardholders react when receiving their paycheck (or monthly income transfer). The final filtering process, moving from 220 to 183 delinquent cardholders, consisted of removing individuals who were delinquent a given month, but recovered in subsequent months. This might indicate that the delinquency status was caused by an oversight, rather than a result of deteriorating financial clout. The final dataset, which combined credit and debit transactions, contained 1,974 non-delinquent cardholders covering 65,501 months, and 183 delinquent cardholders covering 3,547 months.⁴ To avoid a disproportionate amount of non-delinquent months in the final sample (to facilitate estimation), 250 of the non-delinquent cardholders were selected and combined with the 183 delinquent cardholders.

⁴Non-delinquent cardholders have longer transaction histories, on average, because accounts for delinquent cardholders are eventually frozen.

The full dataset with 11,902 observations was employed to estimate the different permutations of the proposed model. As the amount of delinquent cardholders was limited, splitting the data was only carried out when the proposed model was compared to other classifiers. The intention was to alleviate concerns of spurious findings or other small sample issues when examining model structure and carrying out tests of hypotheses.

For comparisons of predictive ability, the data was split into training and test samples. The training sample contained 10,470 months and 357 unique cardholders, of which 913 months and 107 cardholders were delinquent. The test sample consisted of 76 cardholders, which were eventually classified as delinquent, covering 1,432 months, where 595 months were marked as delinquent. Delinquent months were then removed, such that the test sample included all months up to and including the month preceding delinquency classification. This dataset was constructed to measure prediction one month prior to delinquency designation and covers 837 months. Similarly, a dataset for two months prior to delinquency containing 761 months and a dataset for three months prior to delinquency containing 685 months were constructed. The month in question was then marked as a positive, which the classification algorithms attempted to correctly predict. The number of positives in the training sample and the one-month test sample was proportional, while the portion was slightly higher in the two-month and three-month test samples.⁵

Class imbalance was corrected for using various methods and selected using an adaptive racing algorithm (Dal Palozzolo et al., 2013). The racing algorithm suggested synthetic minority over-sampling (SMOTE; see Chawla et al., 2002) for ANN, k -NN, and XGBoost classifiers, while under-sampling was recommended for GBM. SMOTE generates new synthetic instances of the minority class (delinquent cardholders in this case) based on the actual observed cases, while under-sampling randomly removes cases from the ma-

⁵The portion of positives in the training sample was 8.7%, while the portion of delinquent months in the one-, two-, and three-month test samples was 9.1%, 10.0%, and 11.1%.

jority class. For the heterogeneous classifier combining ANN and GBM, the data generated by SMOTE was used. While SMOTE was recommended for the proposed model using k -NN, the instances created by SMOTE prohibited matching cases with parameter values. Random oversampling was used to draw delinquent samples to create a balanced dataset. Classification performance was evaluated for the original proportional data and balanced data. To compare classifiers, the proposed model was re-calibrated using the training data to avoid using the same data twice. A variation of the training data, which mimicked the process of creating the test data, was also constructed, referred to as the congruent training data. For this dataset, delinquent months were removed, such that the month preceding delinquency was marked as delinquent.⁶

3.2.1 Descriptive Analysis

Examining credit card delinquency among Norwegian cardholders reveals a few idiosyncrasies in the data. Mainly, the frequency of transactions registered to credit cards is significantly lower in Norway compared to most other countries. Credit cards serve as a secondary payment instrument, used in 9.2% of card transactions in 2017 (Norges Bank, 2018), which is also evident in the dataset.⁷ Closer examination of the dataset bears witness to this phenomenon when inspecting how repayments and expenses charged are distributed prior to delinquency. As can be observed in figure 3.1, the portion of (eventually delinquent) cardholders charging expenses to their card is around 27%, but this drops abruptly the two months leading up to delinquency. The median amount of expenses charged shows a similar downward tendency (albeit more volatile). The portion making repayments is consistently between

⁶Training data constructed similarly to the congruent sampling data was also constructed, whereby the month preceding delinquency was oversampled. This dataset is labeled the manually resampled training data.

⁷For comparison sake, credit card transactions in the United States accounted for 33.6% of all card transactions (Federal Reserve, 2018).

40%-50%, but also exhibits a marked drop to under 20% before delinquency is declared. The median amount repaid is relatively stable, around 1,000NOK.

This illustrates how most cardholders moderate their spending behavior when they expect financial difficulties. Interestingly, most cardholders also cease repayment of their debt. Intuitively, the opposite would be expected; facing accelerating fees and restrictions to borrowing should prompt rational cardholders to accelerate repayment. These final months usually include a payment reminder and a final reminder, although this does not seem to beget repayment among cardholders. Behavioral explanations notwithstanding, this has implications for constructing variables and model specification when attempting to understand delinquency. Understanding delinquency probability seemingly necessitates capturing cardholder histories in a meaningful way, as opposed to probing monthly states independently.

Figure 3.2 illustrates how debit data add explanatory value when combined with credit data. Monthly income is clearly, and naturally, an important ingredient in predicting delinquency. Cardholders who are eventually declared delinquent carry more debt as a ratio of their credit limit and spend more of their income in the subsequent 10 days of receiving their paycheck. Considering the consistently high rates of spending in addition to the striking increase in monthly balance as a ratio of income suggests that delinquent cardholders continue borrowing and spending unabated when their income falls.

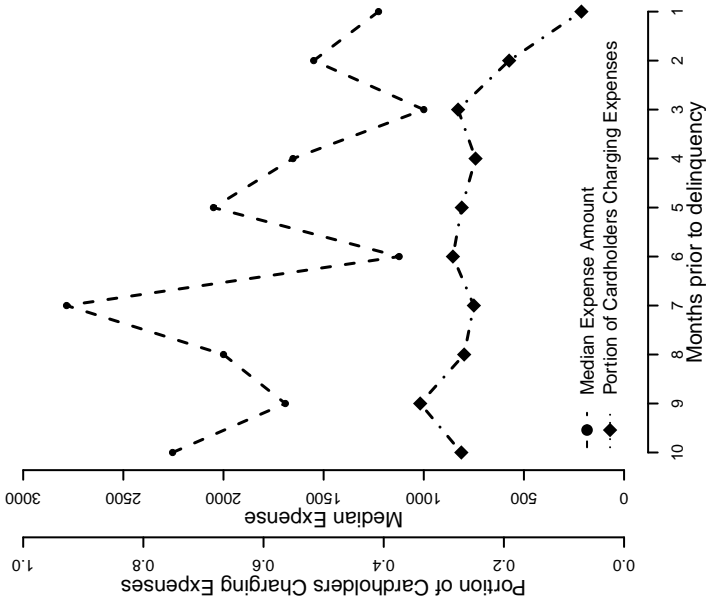
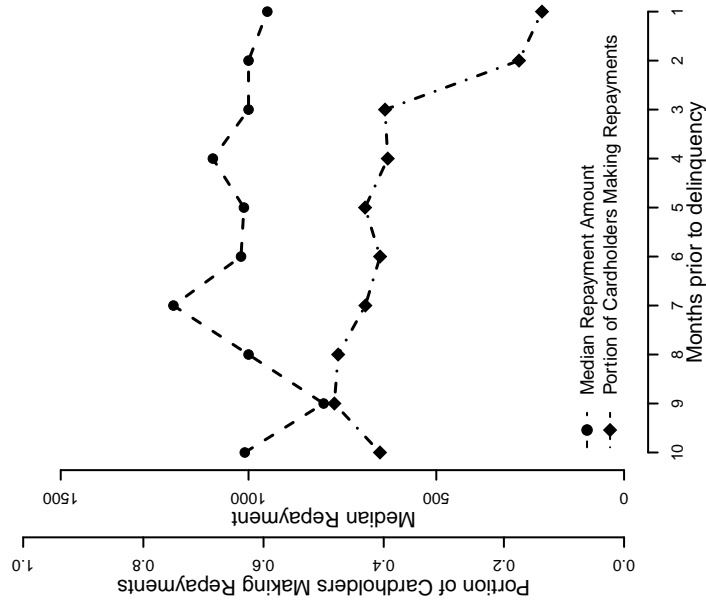


Figure 3.1: Expense and repayment transactions prior to delinquency. Median transaction amount and the portion of cardholders charging expenses or making repayments in a given month are displayed.

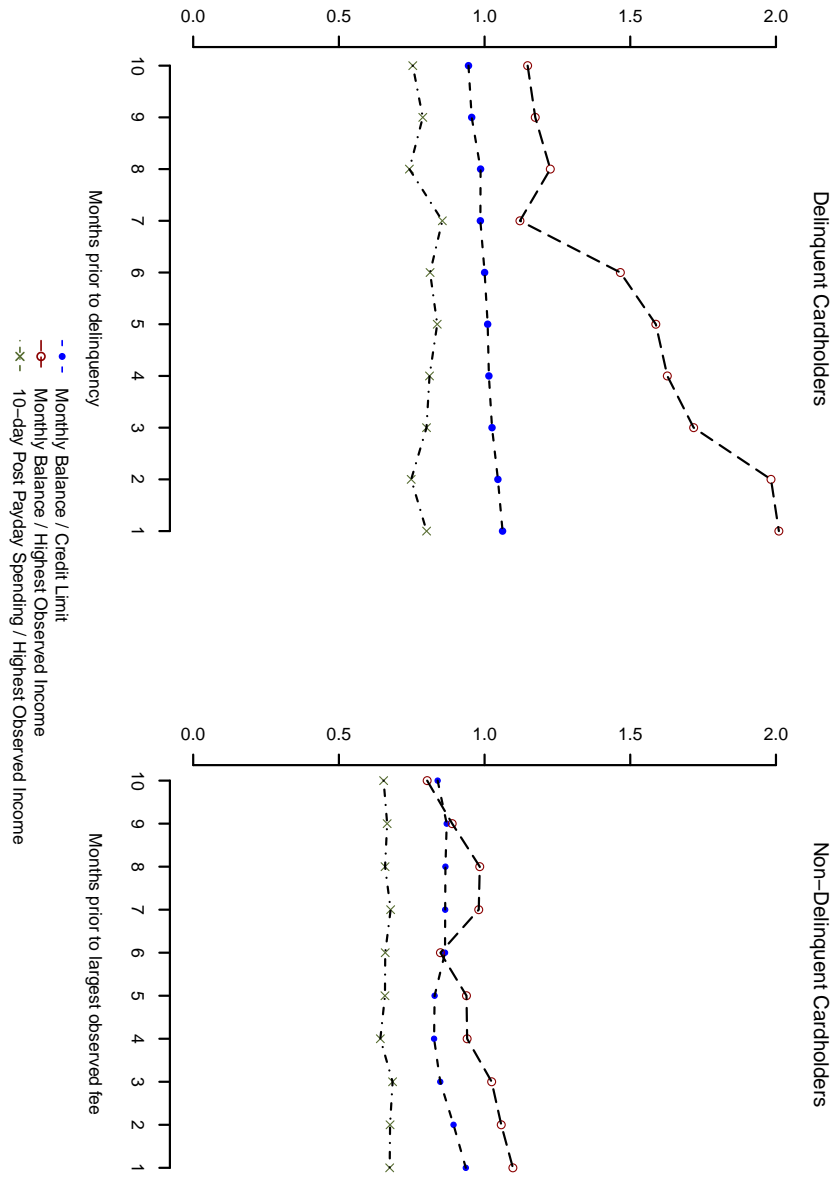


Figure 3.2: *Debit and credit ratios for delinquent and non-delinquent cardholders. Largest observed fee is used as a marker of financial distress for non-delinquent cardholders. Highest observed monthly income attempts to capture paycheck transfers or other income and the subsequent effect on financial behavior.*

The following section details how ratios are constructed to capture the effects posited in the hypotheses. Information from the credit data includes repayment, balance, and credit limit, while monthly income and 10-day pay-day spending are extracted from the debit data.⁸

3.2.2 Selection and Construction of Variables

The dependent variable of interest is delinquency, where delinquency is observed when the account is marked delinquent by the lending institution. In this case, cardholders are served with two payment reminders when payment is neglected. Delinquency is declared three calendar months after the initial expense is registered, including exhaustion of a grace period.⁹ As such, the dependent variable is binary $\{0, 1\}$, where 1 indicates that the cardholder is delinquent.

Payment Decoupling

As previously proposed in equation 2.3, the effect of decoupling entails that cardholders with a high decoupling rate $a \rightarrow 0$ are more likely to be delinquent. A higher decoupling rate leads to a lower valuation of debt reduction in place of continued consumption, suggesting that individuals will repay at lower levels. Also revealed in equation 2.3 is the lagged effect of the decoupling rate. Individuals with consistently high decoupling rates eventually face delinquency as debt accumulates by virtue of low repayment rates.

As a corollary, if the effect of repayment is only significant for the current month, support is found for the traditional life-cycle behavioral hypothesis.

⁸Cash withdrawals are excluded, as these transactions are exceedingly rare in this dataset. Expenses are captured in the monthly balance.

⁹The grace period varies, though the usual grace period is 40 days. This does not affect delinquency, because it is calculated based on calendar months after borrowing and the grace period only pertains to interest charged. After a cardholder is marked as delinquent, the account(s) are sent to a debt collection agency, which either purchases the debt outright or reimburses the lending institution after collection. Delinquency entails a cost in the form of debt that is written off for the lending institution.

Assuming a rational borrower, the current level of repayment is the result of a careful analysis of the current financial situation and expected outlays. Past financial states are not assumed to influence the current situation, as the current state captures the sum of all previous spending and income. For example, consistently low repayment caused by a worsened financial situation is captured by the current balance. As such, the likelihood of delinquency will not depend on past repayment levels.

Using the credit data, the first independent variable is the ratio of monthly repayments to monthly balance. The effect of decoupling is captured by the lagged repayment ratios.

$$PAYBAL_{it} = \frac{REPAY_{it}}{BALANCE_{it}}, \quad (3.1)$$

where i denotes the individual cardholder and t indicates time in months.¹⁰ Note that the number of months that enter into the lagged term varies among individuals for all variables, according to the model specification.

Decision-making Ineptitude

Financial decision-making incompetence, caused by effects such as low financial knowledge or low self-efficacy, suggests that cardholders will continually spend to uphold consumption levels. As shown in equation 2.7, higher levels of habitual decision-making ineptitude, h , will increase the likelihood of delinquency when desired consumption requires borrowing. Financial decision-making ineptitude is observed when borrowing persists at higher levels.

Similarly to repayment, sole support for the concurrent effect will support the traditional life-cycle behavioral hypothesis. Current debt is an accumulation of previous decisions to borrow, and the likelihood of delinquency should be independent of past borrowing levels.

¹⁰The monthly balance used in all independent variable ratios is the balance cardholders are billed in that given month, not outgoing balance.

The second independent variable is a ratio of monthly balance and credit limit. The effect of decision-making ineptitude is captured by the lagged rates of borrowing:

$$BALLIM_{it} = \frac{BALANCE_{it}}{LIMIT_{it}}. \quad (3.2)$$

Present Bias

Individuals who have a higher concern for immediate consequences, as revealed in equation 2.1.5, disregard future payment streams. This promotes higher levels of debt and lower repayment, and will eventually lead to delinquency. A history of accelerated spending of disposable income is suggested to indicate present bias. The rate of spending and subsequent remaining disposable income after the monthly paycheck is used to capture the effect (as in Agarwal et al., 2013).

The variable measuring spending rate is constructed as the amount of monthly income spent as a ratio to balance:

$$CASHBAL_{it} = \frac{INCOME_{it} - NEXTPAY_{it}}{BALANCE_{it}}, \quad (3.3)$$

where $NEXTPAY_{it}$ is the sum of debit expense transactions in the 10 subsequent days following the monthly observed highest income.¹¹

While a concurrent effect of spending rate could be evidence of present bias, the effects are expected to persist. If only current effects are significant, this lends stronger support to the life-cycle behavioral hypothesis as evidence of intermittent myopia as an explanation of financial distress. Lagged effects of the variable captures persistent present bias, resembling a personality trait, as hypothesized.

¹¹Credit card repayments were excluded when calculating this variable.

Financial Risk Aversion

The final hypothesis is consistent with a rational choice model, suggesting that individuals with the ability to anticipate shocks to future finances have a lower likelihood of delinquency. These cardholders are also expected to dutifully plan, avoid risk, and handle fluctuations in income better. These behavioral aspects are also expected to persist and should be observed in lagged effects.

The effect is operationalized as debt clearing ability and is measured as the ratio of monthly income to balance:

$$INCBAL_{it} = \frac{INCOME_{it}}{BALANCE_{it}}. \quad (3.4)$$

Synopsis of Variables

Variations of the variables were also calculated using standardized scores, where the variable score indicates the number of standard deviations from the mean for that particular individual for a given month. These standardizations were employed to combat the lack of variation in the data. The standardized versions of the variables were used in separate implementations. While standardizing variables will not solve issues of causality or relative importance of variables in any econometric model (Gelman, 2008), it might facilitate estimation in hierarchical MCMC (Markov chain Monte Carlo) models. As Kruschke (2011) notes, the viable values of the intercepts and slopes, which the MCMC algorithm explores, are possibly highly correlated, which could lead to Gibbs samplers getting stuck, and Metropolis-Hastings algorithms could fare even worse.¹² Eventual differences in model

¹²Kruschke (2011) denotes the intercept and slope values being explored as “the zone of believability,” and Gibbs samplers could “keep bumping into the walls.”

fit are expected to be a result of this dilemma. The individual-level standardized variables were calculated as follows:

$$\begin{aligned}
 PAYBALscore_{it} &= \frac{PAYBAL_{it} - mean(PAYBAL_i)}{sd(PAYBAL_i)} \\
 BALLIMscore_{it} &= \frac{BALLIM_{it} - mean(BALLIM_i)}{sd(BALLIM_i)} \\
 CASHBALscore_{it} &= \frac{CASHBAL_{it} - mean(CASHBAL_i)}{sd(CASHBAL_i)} \\
 INCBALscore_{it} &= \frac{INCBAL_{it} - mean(INCBAL_i)}{sd(INCBAL_i)},
 \end{aligned} \tag{3.5}$$

where $mean()$ and $sd()$ indicate the intra-individual mean and standard deviation.

Table 3.2 provides a summary of the hypothesized effects, variable names, verbal description of measurement, calculation of, and contingency logic that indicates observation of the hypothesized effect. Table 3.3 shows the correlations between the variables, in addition to the variance inflation factors for each variable, calculated using a fixed effects regression and a “pooled” OLS (ordinary least squares) regression.

Table 3.2: Summary of variables

Hypothesized Effect	Variable Name	Verbal Description	Calculation	Effect Observation Logic
Decoupling	$PAYBAL_{it}$	Rate of debt repayment	$\frac{REPAY_{it}}{BALANCE_{it}}$	Decoupling observed if lagged effects increase delinquency probability
Decision-making ineptitude	$BALLIM_{it}$	Ratio of borrowing to credit limit	$\frac{BALANCE_{it}}{LIMIT_{it}}$	Effect of decision-making ineptitude if lagged effects increase delinquency probability
Present bias	$CASHBAL_{it}$	Post income spending rate	$\frac{INCOME_{it}-NEXTPAY_{it}}{BALANCE_{it}}$	Present bias observed if lagged effects increase delinquency probability
Risk aversion	$INCBAL_{it}$	Ratio of monthly income to balance	$\frac{INCOME_{it}}{BALANCE_{it}}$	Risk aversion observed if concurrent or lagged effects reduce delinquency probability

Table 3.3: Correlations and multicollinearity statistics (VIF) – balanced data

	PAYBAL	BALLIM	CASHBAL	INCBAL	I.PAYBAL	I.BALLIM	I.CASHBAL	I.INCBAL
PAYBAL	1.00							
BALLIM	-0.19	1.00						
CASHBAL	0.27	-0.21	1.00					
INCBAL	0.26	-0.25	0.86	1.00				
I.PAYBAL	0.10	-0.13	0.01	0.02	1.00			
I.BALLIM	-0.00	0.36	-0.07	-0.09	-0.26	1.00		
I.CASHBAL	0.03	-0.14	0.04	0.08	0.19	-0.36	1.00	
I.INCBAL	0.04	-0.16	0.04	0.08	0.19	-0.44	0.88	1.00
VIF1	1.11	1.16	3.78	3.80	1.09	1.29	4.04	4.23
VIF2	1.12	1.24	3.78	3.83	1.10	1.44	4.41	4.71

Sample correlations, balanced data using oversampling method. Lagged variables, indexed with the prefix “l,” are estimates from the proposed model, included here as they enter into the classification method using k-NN. VIF1 is the variance inflation factor constructed from a fixed effects regression, while VIF2 is the same measure constructed using a “pooled” OLS.

3.3 Model Specification and Estimation

Delinquency probability is modeled as a hierarchical logit regression. The coefficients of interest are the aggregate B coefficients on the independent variables X_k , along with the individual-level coefficients b_i . In addition, the segment assignment s_i for each individual and the p lag coefficients for each segment, which specify the lag weights, are central to the following model. The logit of the probability π_i is thus given by:

$$\text{logit}(\pi_i) = (B_0 + b_{i,0}) + (B_k + b_{i,k})(X_{i,k} + A_{i,k}), \quad (3.6)$$

where $A_{i,k}$ is the distributed lag term of the k independent variable, specified by the following equation:

$$A_{i,k} = X_{i,k,t-1}(p_{i,s_k})^1 + X_{i,k,t-2}(p_{i,s_k})^2 + X_{i,k,t-3}(p_{i,s_k})^3 + X_{i,k,t-4}(p_{i,s_k})^4 + X_{i,k,t-5}(p_{i,s_k})^5, \quad (3.7)$$

where p_{i,s_k} indicates the lag weight for individual i assigned to segment s .¹³ The number of lags is initially fixed at five lags, which corresponds to the maximum lag length.¹⁴ The number of segments is restricted, as is the lag weight space, to facilitate estimation. The lag weight parameters, p , are restricted between $(0 < p < 0.5)$ and $(0.5 < p < 1)$, such that individuals are assigned to a segment with a low or high lag weight. For $s = 4$ segments, this results in the following setup for the credit-only models (with

¹³This resembles the infinite distributed lag model, in which the Koyck (1954) transformation to an ARMAX model specification is commonly applied. However, as the number of distributed lags here is finite, estimation is feasible.

¹⁴As previously explained, the data were filtered to include only cardholders with five months in which they were not delinquent. The minimal amount of data for an individual translates to five months of credit card transactions, and the cardholder would be declared delinquent at month six.

two independent variables, k):

$$s_i \in \{1, 2, 3, 4\} \quad (3.8)$$

$$P_{s,k} = \begin{pmatrix} 0.5 < p < 1 & 0 < p < 0.5 \\ p_{s=1,k=1} & p_{s=2,k=2} \\ p_{s=1,k=2} & p_{s=3,k=1} \\ p_{s=2,k=1} & p_{s=4,k=1} \\ p_{s=3,k=2} & p_{s=4,k=2} \end{pmatrix}, \quad (3.9)$$

where each individual i is assigned to one of four segments. This reveals a total of eight lag weight parameters $p_{s,k}$ to estimate, four permutations of segments for the $k = 2$ independent variables. For the full model, with credit and debit data and four independent variables, only two segments were used, as opposed to the fully parameterized option used for the credit-only models. Lag weight assignment is restricted to either high or low lag weights for all four variables. This was done to facilitate estimation. While some precision is lost, the results shown for the credit-only models suggest that this is not a significant problem, as most cardholders were either assigned to a low or high lag weight segment when combinations were permitted.

$$s_i \in \{1, 2\} \quad (3.10)$$

$$P_{s,k} = \begin{pmatrix} 0.5 < p < 1 & 0 < p < 0.5 \\ p_{s=1,k=1} & p_{s=2,k=1} \\ p_{s=1,k=2} & p_{s=2,k=2} \\ p_{s=1,k=3} & p_{s=2,k=3} \\ p_{s=1,k=4} & p_{s=2,k=4} \end{pmatrix} \quad (3.11)$$

The $s = 2$ segments and $k = 4$ independent variables generates eight lag weight parameters $p_{s,k}$ to estimate.

Next, the assumption of a uniform number of lags is relaxed, allowing lag lengths to vary between individuals. The individual A variable lag is then dependent on the J lag length:

$$A_{i,k,j} = \sum_{j=1}^J (X_{i,k,t-j}(p_{i,s_k})^j). \quad (3.12)$$

3.3.1 Estimation of Proposed Model

The procedures for Bayesian specification of a logit with a continuous mixing distribution has been formulated in detail in several previous publications (Train, 2003; Rossi et al., 2005). Allowing for individual heterogeneity where the individual-level parameters b_i follow a normal distribution, $b_i \sim \mathcal{N}(u, W)$, and where parameters u and W are given diffuse priors, $u \sim \mathcal{N}(b_0, s_0)$, $W \sim IW(\nu_o, s_o)$ ¹⁵, we end up with a posterior distribution that is relatively easy to sample from:

$$p(u, W, b_i|Y) \propto \prod_i L(Y_i|b_i)\phi(b_i|u, W)p(u, W), \quad (3.13)$$

for n individuals i . Sampling from this posterior entails sampling from the conditional distributions. Given a diffuse prior for $W \sim IW(\nu_o, s_o)$, the posterior distribution becomes:

$$IW \sim (k + N, \frac{\nu_o s_o + N\bar{S}}{\nu_o + N}) \quad (3.14)$$

$$\bar{S} = (1/N) \sum_{i=1}^n (b_i - u)(b_i - u)',$$

¹⁵ \mathcal{N} denotes a normal or multivariate normal distribution and IW denotes an inverse Wishart distribution. b_0 is typically set to 0, with a sufficiently large variance s_0 . ν_o is the degrees of freedom for the inverse Wishart, and s_o is the scale matrix that is typically a $k \times k$ identity matrix.

where N is the total number of observations and k is the number of B aggregate parameters estimated. Given a diffuse normal prior distribution for u , the posterior distribution for the parameter vector B with k dimensions is:

$$B \sim \mathcal{N}(\bar{b}, W/N), \quad (3.15)$$

where \bar{b} is the sample mean of the coefficients b_i plus the Cholesky factor of W/N multiplied by a vector of k standard normal draws. Drawing values for B and W is implemented with a Gibbs sampler (Geman and Geman, 1990), while drawing from the posterior of the b_i parameters entails a Metropolis-Hastings step. The b_i parameters are drawn using a Metropolis-Hastings step using the normal density:

$$b_i \sim \mathcal{N}(0, \rho^2 W), \quad (3.16)$$

where ρ is specified by the researcher, and determines the size of each jump for the Metropolis-Hastings algorithm. Lag weight segments are also assigned according to an accept-reject algorithm. Individual cardholders are assigned to a different segment if the log likelihood ratio for the new segment is greater than a uniform draw $(0, 1)$. The lag weight values for each segment are then subjected to a random walk, and the candidate values are accepted using the same accept-reject rule as for the segment assignment. The full routine can be summarized thus:

$$s_i = s_i^* \text{ if } \frac{\sum_{i=1}^N \pi(s_i^*, Y_i | \beta, b_i, X_i, A_i) * q_{switch}}{\sum_{i=1}^N \pi(s_i, Y_i | \beta, b_i, X_i, A_i) * q_{stay}} > \log(U(0, 1)) \quad (3.17)$$

$$p_l^* = p_l + \alpha$$

$$p_l = p_l^* \text{ if } \frac{\sum_{l=1}^4 \pi(p_l^* s_i, Y_i | \beta, b_i, X_i, A_i)}{\sum_{l=1}^4 \pi(p_l, s_i, Y_i | \beta, b_i, X_i, A_i)} > \log(U(0, 1)),$$

where α is a uniform draw of a sufficiently small value, p^* is the candidate lag weight vector, π is the log likelihood function, s^* is a different segment than the current one at a given iteration, and the q values are the probabilities assigned to switching or staying at a given segment. Switching probabilities are introduced to penalize switching, tuning the sampling algorithm as iterations increase. The model evaluates all segments for each iteration, which in the credit-only model equates to four segments and two segments for the full model. The q values are assigned according to transition matrices, and weighted according to the segment variability for a particular individual:

$$\begin{aligned} l_{i,switch} &= ks_i * SS_i^* \times T_i + (1 - ks_i) * SS_i^* \times AT \\ l_{i,stay} &= ks_i * SS_i \times T_i + (1 - ks_i) * SS_i \times AT, \end{aligned} \quad (3.18)$$

where T_i is the individual transition matrix, AT is the aggregate transition matrix, and ks_i is an individual weight. SS_i and SS_i^* are vectors indicating current and candidate segments for each individual, and the transition matrices are calculated based on previous segment transitions.

The number of lags is again selected in a Metropolis-Hastings step with an accept-reject algorithm. Thus, finding the optimum number of lags at a given iteration follows this procedure:

$$A_{i,J} = A_{i,J}^* \text{ if } \frac{\sum_{i=1}^N \pi(A_{i,T}^*, Y_i | p_i, s_i, \beta, b_i, X_i, A_i)}{\sum_{i=1}^N \pi(A_{i,T}, Y_i | p_i, s_i, \beta, b_i, X_i, A_i)} > \log(U(0, 1)). \quad (3.19)$$

An individual is assigned to a different lag from the set $J = \{1, 2, 3, 4, 5\}$, where all lags are considered for each parameter k and individual i .

3.3.2 Specification of Classification Algorithms for Prediction

As mentioned in section 2.2, the model proposed in this thesis will be compared to prevailing methods in credit scoring. To reiterate, the chosen methods include neural networks (ANN), boosting algorithms (GBM and XGBoost) and stacked ensembles (homogeneous ensembles, SE, and heterogeneous ensembles, HSE). The specification of these classification algorithms is detailed below. In addition, the specification of the routine to estimate predictive performance of the proposed model is presented.

Implementation of the machine learning algorithms is carried out using the H2O packages for ANN (Arora et al., 2015) and boosting (Click et al., 2015), while XGBoost is implemented using the Caret package (Kuhn, 2008). Grid searches are used to tune hyperparameters using a random discrete search strategy, where the area under the ROC (receiver operating characteristic) curve (AUC; see Bradley, 1997) is used as search and stopping criteria for all algorithms. Specifically, the ANN estimated is a multi-layer feedforward network (also referred to as a deep neural network, DNN), which is trained using stochastic gradient descent. Specification involves deciding the number of neurons, hidden layers, and activation function in addition to tuning parameters, such as learning rate and regularization (L1, L2, and dropout). Tuning the boosting algorithm involves a grid search of parameters such as tree depth, learning rate, and sampling rate (rows and columns). Stacked ensembles were also created from the final grid specifications for both DNN and GBM, yielding two homogeneous ensembles (SEs) and one heterogeneous ensemble (HSE). The grid search for XGBoost involves specifying parameters similar to GBM. All classifiers were trained using k -fold cross validation. Where applicable, threshold values were selected using the $F_{0.5}$ metric, which gives more weight to precision than to recall.¹⁶ The stacked

¹⁶Precision is the portion of true positives of all positive cases, while recall is the portion of true positives of all positive cases. Given the large number of negatives in the training

ensembles of SE and HSE require a metalearning (or superlearner) algorithm. Here, the metalearning algorithm for all ensembles is GBM.

For the proposed model, prediction of delinquency is calculated using a k -NN routine. The procedure uses a nearest neighbor classification scheme to select the B_k and $b_{i,k}$ parameter coefficients from the most similar observed financial history in the training sample. These coefficients are then used in the model equation along with the independent variable values from the cardholder being examined to determine the delinquency probability of the individual.

For all prediction algorithms, probabilities were rescaled as specified by Platt (1999). This creates probabilistic outputs that are needed to calculate one of the classifier performance indicators (namely the Brier score).

3.3.3 Classifier Performance Indicators

Comparing classification performance of the proposed model with the machine learning algorithms will be carried out using several metrics. Each metric examines different aspects of model classification, allowing for a robust interpretation of performance. The specification of these metrics is presented below.

As recommended by Lessmann et al. (2015), classifier performance is evaluated with the AUC score, Brier score (BS) and a partial Gini index (PG). AUC is a global evaluation of classifier performance that is widely used and highly correlated with other similar performance metrics.¹⁷ AUC gives the probability that a randomly chosen positive observation is ranked

sample and test sample, most metrics will produce thresholds that are imprecise. A single threshold could also have been set *a priori*, such as the prior class probability alone or combined with the misclassification cost (Lessmann et al., 2015). For the dataset examined here, the $F_{0.5}$ metric generally produces thresholds closer to the prior class probability.

¹⁷Lessmann et al. (2015) note that the AUC score is highly correlated with the H-measure, PCC (percent correctly classified), and KS (Kolmogorov-Smirnov) statistic. They recommend choosing one of either the AUC or H-measure, despite recent criticism of the AUC metric (Hand and Anagnostopoulos, 2013).

higher than a randomly chosen negative observation. Formally, the AUC is calculated as the area under the curve indicating the false positive rate and the true positive rate at different thresholds for a classifier. As such, it is invariant to the chosen threshold and predicted probability scale. As the formula for the partial Gini index is presented below, be mindful of the relation between the two scores: $AUC = (PG + 1)/2$ for a specified cutoff value of $b = 1$ for the partial Gini index. The Brier score is commonly defined as mean squared error of the forecast:

$$BS = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2,$$

where p_i is the probability of delinquency for cardholder i and o_i is the observed indicator of delinquency. BS is essentially a measure of accuracy of predicted probabilities. In contrast to AUC and BS metrics, the partial Gini index relaxes the assumption that prediction performance across the entire distribution of predictions is equally important. Although the Gini index is closely related to the AUC in construction, the partial Gini index is calculated among the lower tail of prediction scores, which are especially crucial to correctly predict in credit scoring. Following Pundir and Seshadri (2012), the PG index is calculated as:

$$PG_b = 1 - \frac{2 * \int_0^b L(p) dp}{b^2},$$

where $L(p)$ is the Lorenz function of the predictions, and b is a cutoff for a portion of cases for which the Gini index is calculated. Here the cutoff is set at $b = 0.4$ in concordance with Lessmann et al. (2015) and Pundir and Seshadri (2012).¹⁸

¹⁸The cutoff $b = 0.4$ could be set at any arbitrary value, although it should reflect the portion of interesting cases. The equation for PG is adapted and simplified to credit scoring classification, as the starting point is 0. In other applications, the PG is presumably specified to examine a variety of other portions of cases.

Collectively, the AUC, BS, and PG evaluate aspects of global performance, accuracy of probability predictions and local performance of the classifiers. These metrics will serve as indicators of classification performance to allow for a comprehensive comparison between the proposed model and the machine learning algorithms.

4 — Results and Analysis

This chapter will present the parameter estimates for the proposed model and its precursors, model fit for the proposed model and hypothesis tests, and comparisons of predictive ability between the proposed model and the machine learning algorithms. The summary of parameter estimates in section 4.1 are shown in a sequential manner, illustrating the incremental value of different variables and layers of model complexity. The proposed model, shown last in section 4.1, exhibits better global fit according to several measures of fit, as shown in section 4.2. Lastly, section 4.3 investigates the practical usefulness of the proposed model by comparison of predictive performance with the chosen machine learning algorithms.

4.1 Model Estimates

Model estimates for the different models are presented, starting with the naive heterogeneity model. Posterior medians of the aggregate B_k values for the independent variables are reported along with the corresponding posterior standard deviations. The estimates represent the posterior distributions of the parameters; as such, the classical test of significance can be approximated using the ratio heuristic of point estimate and standard deviation, though this will not be explicitly presented. The first model presented is estimated using credit data while allowing for individual heterogeneity. Model 2 adds segmented random lag weights at fixed lag lengths, while model 3 allows the

lag lengths to vary. Finally, the proposed model introduces the debit data variables. Especially noteworthy quantities include the parameter estimates, classification rates, lag weights and segment memberships, and lag lengths.

Table 4.1 displays the (aggregate) B_k estimates for the naive model; using credit data estimated with a continuous mixing distribution. The naive model corresponds to benchmark model 2 in Zhao et al. (2009), although the basic structure employed here is a simple logit specification (compared to a simple Tobit II). This model captures heterogeneity, though dynamic effects are not included. The three columns represent different variable specifications: simple ratios, standardizing the *PAYBAL* variable, and standardizing both variables. *INT* is the aggregate intercept estimate. As is expected, an increase in the ratio of monthly balance over credit limit, *BALLIM*, leads to a higher probability of delinquency. Also, an increase in the ratio of monthly payment over monthly balance, *PAYBAL*, decreases the probability of default. The standardized variables perform better in terms of classification, and the *PAYBAL* variable has no effect when not standardized. The individual-level estimates, $b_{i,k}$, are not presented for this model specification. Examining the classification rate, this specification performs poorly regardless of which variables are employed. Type I errors are prevalent even when using the standardized variables, where 356 months were misclassified as delinquent. Likewise, a high total of 265 months were overlooked, being misclassified as non-delinquent when the opposite is true.

Table 4.1: *Naive model: continuous heterogeneity, credit data variables*

B_k Regression Coefficients			
	Simple Ratio Variables	Standardized Payment Variable	Standardized Variables
<i>INT</i>	-0.006 (0.008)	-0.033 (0.012)	0.162 (0.010)
<i>BALLIM</i>	0.044 (0.010)	0.134 (0.010)	0.044 (0.008)
<i>PAYBAL</i>	0.000 (0.000)	-0.034 (0.007)	-0.030 (0.008)
Classification Rates of Customer-Months Defaulted			
	Simple Ratio Variables	Standardized Payment Variable	Standardized Variables
Type I error	405	383	356
Type II error	433	391	265
Correctly classified	11049	11113	11266

Type I error is understood as misclassifying a customer as delinquent in a month in which he/she was not. Correctly classified months contain both delinquent and non-delinquent months. Parameter estimates are posterior medians.

Table 4.2 presents estimates for model 2, which captures heterogeneity and introduces lagged independent variables for the credit data. This model includes both heterogeneity and a simple structure to capture dynamic effects, mirroring the proposed model in Zhao et al. (2009).¹ Examining the aggregate B_k posterior median estimates for the independent variables with fixed lag length of five months, the effect of former monthly financial status is noticeably larger than current monthly financial status. As explained in the previous chapter, the lagged independent variable value in a given month for

¹Keep in mind that the basic structure here is a logit, as opposed to a Tobit II specification used in Zhao et al. (2009).

Table 4.2: *Model 2: continuous heterogeneity with segmented random lag weight and fixed lag length, credit data variables*

B_k Regression Coefficients				
	Simple Ratio Variables	Standardized Payment Variable	Standardized Variables	
<i>INT</i>	-0.005 (0.004)	-0.085 (0.010)	0.176 (0.003)	
<i>BALLIM</i>	-0.004 (0.003)	0.020 (0.003)	0.025 (0.006)	
<i>PAYBAL</i>	0.000 (0.000)	-0.027 (0.005)	-0.025 (0.005)	
<i>LagBALLIM</i>	0.131 (0.004)	0.091 (0.003)	0.055 (0.002)	
<i>LagPAYBAL</i>	0.000 (0.000)	-0.102 (0.021)	-0.025 (0.004)	
Classification Rates of Customer-Months Defaulted				
	Simple Ratio Variables	Standardized Payment Variable	Standardized Variables	
Type I error	232	183	145	
Type II error	189	168	101	
Correctly classified	11466	11536	11641	
Standardized Variables: Lag Weight Segment Estimates and Segment Membership				
	Segment 1	Segment 2	Segment 3	Segment 4
Lag weight estimates	0.99/0.99	0.820/0.320	0.370/0.870	0.226/0.226
Median segment: delinquents	111	45	54	10
Median segment: non-delinquents	1	12	11	226

the fixed lag specification is a sum of the discounted values of the previous five months. The lag weight is assigned using a segment-based routine. The lag parameters are estimated using a random walk, and parameter estimates for each segment for the two credit variables using the standardized scores are presented in table 4.2.²

²The setup used here for the segments is presented in the previous chapter. Simply put, segment 1 has high lag weights for both variables, segment 2 has a high lag weight

Also reported is the sum of the posterior median segments for the delinquent and non-delinquent cardholders. The delinquent cardholders are predominantly assigned to segment 1, which gives large lag weight values to both variables. Delinquent cardholders are also assigned to segments 2 and 3, which give a large lag weight to one of the two variables. Only 10 delinquent cardholders are assigned to segment 4, which gives low lag weights to both independent variables. However, the non-delinquent cardholders are predominantly assigned to the segment giving low lag weights for both variables. Using the standardized variables, 226 of the 250 cardholders are assigned to segment 4. Comparing the classification rates, the model that includes lagged independent variables clearly outperforms the model that only controls for individual-level heterogeneity. The false positive classifications for the standardized variables is improved from 356 customer-months to 145, and false negatives are improved from 265 to 101. Accounting for previous financial status is obviously a large improvement in model specification.

Demonstrating the effect of adding dynamic effects, model 3 displayed in table 4.3 relaxes the constraint of fixed lag lengths, allowing for varying lag lengths for each individual. Although the effect of the previous financial state is largely governed by the lag weight, it is expected that the lag length will also differ between cardholders. Some cardholders will be affected by financial states or shocks several months prior to the current month, while other cardholders are only affected by the last month's financial state.

Also displayed in table 4.3 are the relevant parameter estimates, which include individual-level heterogeneity, random segmented lag weights, and random lag lengths. Referencing the estimates for the standardized variables (right-hand column), the effects of the independent variables are simi-

for the BALLIM variable and a low lag weight for the PAYBAL variable, segment 3 is the opposite of segment 2, while segment 4 assigns low lag weights to both variables. The lag weight can be interpreted as an autoregressive parameter; a high decay value translates to a larger effect from month $t_{-1:T}$, where the number of lags T in the fixed lag model is five.

Table 4.3: *Model 3: continuous heterogeneity with segmented lag weight and random lag length, credit data variables*

B_k Regression Coefficients					
	Simple Ratio Variables	Standardized Payment Variable	Standardized Variables		
<i>INT</i>	0.002 (0.004)	-0.098 (0.014)	0.183 (0.007)		
<i>BALLIM</i>	-0.007 (0.004)	0.012 (0.004)	0.026 (0.005)		
<i>PAYBAL</i>	0.000 (0.000)	-0.024 (0.004)	-0.030 (0.005)		
<i>LagBALLIM</i>	0.120 (0.003)	0.097 (0.002)	0.065 (0.006)		
<i>LagPAYBAL</i>	0.000 (0.000)	-0.094 (0.007)	-0.088 (0.005)		
Classification Rates of Customer-Months Defaulted					
	Simple Ratio Variables	Standardized Payment Variable	Standardized Variables		
Type I error	234	183	131		
Type II error	199	137	82		
Correctly classified	11454	11567	11674		
Standardized Variables: Lag Weight Segment Estimates and Segment Membership					
	Segment 1	Segment 2	Segment 3	Segment 4	
Lag weight estimates	0.99/0.99	0.72/0.22	0.24/0.74	0.33/0.33	
Median segment: delinquents	106	39	36	2	
Median segment: non-delinquents	3	17	21	209	
Standardized Variables: Median Lag Length					
	1 Lag	2 Lags	3 Lags	4 Lags	5 Lags
Delinquents	6	17	35	26	104
Non-delinquents	238	6	1	1	4

lar to model 2. The lagged effects of the independent variables are different, which is indicative of the model assigning different lag lengths than the fixed value of five. The posterior median estimates for the *LagBALLIM* and *LagPAYBAL* variables are 0.065 and -0.088, respectively, while the corre-

sponding estimates in the fixed lag length model (model 2) are 0.055 and -0.025, respectively.

Classification rates are improved, as only 131 customer-months are misclassified as false positives, while 82 are false negatives (compared to 145 and 101, respectively, for the fixed lag length model). Segment assignment is also similar, where delinquent cardholders are assigned to segments 1 to 3, while the non-delinquent cardholders are primarily assigned to segment 4. The assigned lag lengths also seem to indicate that cardholders who experience financial distress are more affected by the previous financial states. Delinquent cardholders are assigned to higher lag lengths (five is the limit), while non-delinquent cardholders are primarily assigned to one lag length. The segments assigned are similar to the trends observed in table 4.2, where delinquent cardholders are assigned higher lag weights, while non-delinquent cardholders are assigned to lower lag weights. According to the lag lengths and lag weights assigned, delinquent cardholders are clearly more affected by previous financial states than their non-delinquent counterparts.

Table 4.4 presents the proposed model where both credit and debit variables are used, while heterogeneity and dynamic effects are accounted for. The parameter estimates for the credit variables are similar to the estimates found in previous model specifications, where an increase in the ratio of balance carried to credit limit will promote delinquency, while the opposite will happen if a repayment happens (measured as a ratio of repayment over balance carried). Surveying the results, an increase in the debit variable *CASHBAL* decreases the probability of delinquency. Less expected, a higher income (*INCBAL*) does not reduce the likelihood of delinquency.

As observed in model 3, which employs credit data, the lagged variables effects are more prominent than their counterparts of current measures for the full dataset. Predicting delinquency is more dependent on previous financial states than on current finances. The lagged income variable *LagINCBAL* is curiously significant, though the sign suggests that an increase in income

will slightly increase the likelihood of delinquency.

The number of lag weight segments is restricted to two, as a fully parameterized segment solution would necessitate 16 segments with 64 lag weights (an intractable or near-intractable solution). The two segments feature either high lag weights or low lag weights for all variables, and the previously observed trends still apply; delinquent cardholders are assigned higher values, while non-delinquent cardholders are assigned to low lag weights. Lags are estimated as before, and the distribution of lags among cardholders is similar. Classification is greatly improved; when viewing the model that includes the *PAYBAL* variable, only 44 months are misclassified as delinquent, while 45 months are misclassified as non-delinquent. Hypothesis tests for this model, and model comparisons are found in subsection 4.2.

The modeling scheme used here, a hierarchical Bayesian logit, draws aggregate B_k values and allows for heterogeneity on an individual level. Thus, the individual-level estimates, denoted $b_{i,k}$, are usually centered around the aggregate parameter estimates.

Table 4.4: *Proposed model: continuous heterogeneity with segmented random lag weights and random lag length, debit and credit variables.*

B_k Regression Coefficients					
<i>INT</i>	0.126 (0.003)				
<i>BALLIM</i>	0.009 (0.003)				
<i>PAYBAL</i>	-0.003 (0.005)				
<i>CASHBAL</i>	-0.007 (0.004)				
<i>INCBAL</i>	0.005 (0.004)				
<i>LagBALLIM</i>	0.051 (0.004)				
<i>LagPAYBAL</i>	-0.027 (0.003)				
<i>LagCASHBAL</i>	-0.031 (0.002)				
<i>LagINCBAL</i>	0.012 (0.002)				
Classification Rates of Customer-Months Defaulted					
Type I error	44				
Type II error	45				
Correctly classified	11813				
Lag Weight Segment Estimates and Segment Membership					
	Segment 1	Segment 2			
Lag weight estimates	0.97	0.146			
Median segment: delinquents	161	22			
Median segment: non-delinquents	6	244			
Median Lag Length					
	1 Lag	2 Lags	3 Lags	4 Lags	5 Lags
Delinquents	15	18	24	32	95
Non-delinquents	244	5	3	0	0

Only standardized variables are employed.

As a consequence of this shrinkage effect, individuals who have a low probability of defaulting (and do not default) have individual-level estimates that are centered around zero. To illustrate this effect, figure 4.1 plots the posterior median $b_{i,k}$ values for the delinquent and non-delinquent cardholders for the credit variables. The cardholders who are delinquent (to the left of the vertical line in the plots) show a large variance in individual-level estimates. The non-delinquent cardholders (to the right of the vertical line) show a comparably smaller variance, as visible in the b_i estimates for the intercept.

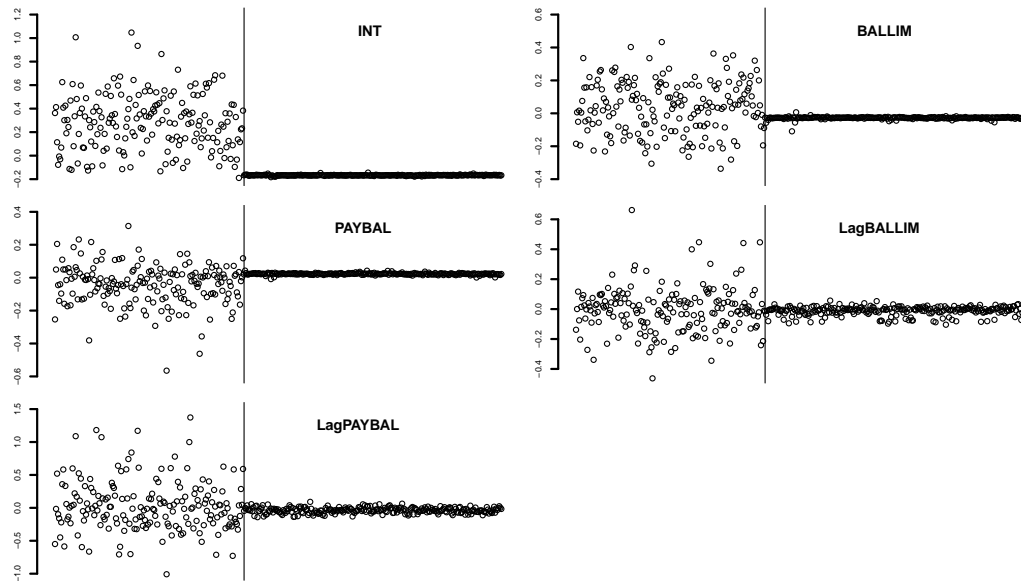


Figure 4.1: Individual-level $b_{i,k}$ posterior median estimates. The vertical line draws the distinction between delinquent cardholders on the left side and non-delinquent cardholders on the right side. Estimates are based on the standardized variables with segmented decay and random lag lengths.

4.2 Model Fit and Hypothesis Tests

The debit data clearly improves classification, demonstrating superior explanatory power. Compared to the model presented in Zhao et al. (2009), with the addition of segmented lag weights, random lag selection already improves classification.³ The improvements to estimation found in the random lag segmented lag weight structure and the addition of debit data allows for the classification of potentially problematic cardholders.

Table 4.5: *Model fit summary*

Classification of Unaltered Unbalanced Training Data				
	Naive Heterogeneity	Heterogeneity, Fixed Lag	Heterogeneity, Random Lag, Credit Only	Heterogeneity, Random Lag, Credit and Debit
Tjur R^2	0.6256	0.8151	0.8324	0.9024
Classification	0.9478	0.9793	0.9793	0.9925
Delinquency classification	0.8243	0.9330	0.9456	0.9702

The Tjur (2009) R^2 is an analogue to the similar coefficient found in ordinary regression. Classification is the sum of correctly classified, delinquent and non-delinquent, divided by the sample. Delinquency classification is the proportion of correctly classified delinquent months.

Hypothesis tests are carried out by evaluating the posterior distributions of the B_k parameters as well as the lag weight segment assignment and lag lengths found in the proposed model (shown in table 4.4). For hypotheses related to the independent variables, a hypothesis is supported if 95% of the credible interval region for the parameter posterior distribution does

³The specification in Zhao et al. (2009) did include autoregressive parameters, but only individual-level b_i estimates for the lagged dependent variables.

not include the corresponding null hypothesis value.⁴ Through evaluation of the proposed model, which includes debit and credit data with the complete structure of segmented random lags, several of the hypotheses are supported. For **H1**, the model shows that a lower ratio of repayment in previous months will increase the likelihood of delinquency. Support is also found for hypothesis **H2**, as the model shows that a higher ratio of borrowing in previous months will increase the current likelihood of delinquency.

Hypothesis **H3** is also supported; a higher payday spending rate in previous months will increase the likelihood of delinquency. Hypothesis **H4** is not supported. The ratio of income to balance in previous months will not reduce the likelihood of delinquency. This hypothesis examines the effect of debt clearing ability, suggesting that the amount of debt carried relative to monthly income is not a significant predictor of delinquency. Further inference based on the non-significant finding of hypothesis **H4** implies that the rational economic logic of a higher monthly income will safeguard against delinquency does not hold. This could imply that effects suggested by present-biased preferences and mental accounting are not only necessary, but sufficient in explaining delinquency. However, this interpretation commands caution, as the effects are likely highly dependent on the operationalization of the variables.

4.3 Predictive Performance

Predictive performance is employed as an omnibus test of model fit and usefulness. Following recent recommendations in an extensive review of credit scoring (Lessmann et al., 2015), the proposed model is compared with multiple state-of-the-art classification algorithms (as discussed in section 2.2) and tested with a variety of suitable performance metrics (as presented in sec-

⁴The Bayesian credible interval is usually statistically similar to a classical confidence interval, though philosophically they are different. Differences will not be expounded upon here.

tion 3.3.3). In addition, the proposed model and other classifiers are trained and tested using unmodified data and imbalance-corrected data. The same comparisons are carried out for the congruent sampled training data, where months prior to delinquency are removed from the training dataset (see section 3.2). Superior predictive performance would indicate that the proposed model is more adept at classifying potentially delinquent cardholders than the prevailing classification algorithms used in credit scoring.

Table 4.6 presents predictive performance of the different algorithms for the default-only and balanced test sample, using unaltered and balance-corrected training data. Algorithms are trained using the unaltered training sample, with class membership proportional to the test samples. Concerning the classification metrics, the proposed model with k -NN selection outperforms the machine learning algorithms when examining AUC and the partial Gini index. This suggests the proposed model has the highest global predictive performance and the highest predictive performance for a crucial portion of cases. Assessing probability accuracy, the Brier score ranks XGBoost as the top performing algorithm for the default-only test sample, though the proposed model ranks highest when examining the balanced test sample. Concerning overall performance for the unaltered training sample, the proposed model predicting the balanced test sample achieves the highest scores for all metrics. Also, when the imbalance-corrected training sample is used, the predictive results generally improve for most of the machine learning algorithms (though not for the proposed model).

Assessing the results when using the congruent sampled training sample in table 4.7, the machine learning algorithms generally improve. However, the proposed model still achieves the highest scores overall, though not exclusively for the balanced test sample. Similar to the results for the unaltered training sample, predictive accuracy is challenged when examining the default-only test sample, although the top performing algorithms vary. Comparing results for the two training samples reveals similar scores from the top

performing classifiers, though it should be noted that predictive accuracy for the proposed model is significantly higher for the congruent training sample.⁵ The results suggest that the proposed model outperforms the machine learning algorithms for global prediction, local prediction, and probability prediction accuracy for the different training and testing samples.

Examining the expenses and repayment observed for the eventually delinquent cardholders in the test sample reveals the potential savings of implementing a useful behavior scoring model. Using the three-month prediction rates for the proposed model with k -NN and assuming that all delinquent accounts are eventually written off, the gains of freezing likely delinquent accounts subtracted by the loss of misclassifying eventually delinquent accounts amounts to a profit of NOK 211,356 for a hold-out sample of 76 delinquent cardholders. This becomes an average savings of NOK 2,781 per delinquent cardholder, due to an average increase in borrowing in the three months leading to delinquency of 15.76%. As noted, the Federal Reserve Bank of New York estimates that Americans presently hold a total of \$684 billion in credit card debt, suggesting that this behavioral scoring routine could save the industry \$15.8 billion, assuming no behavioral scoring models are currently implemented.⁶

In conclusion, the predictive performance of the model underscores the viability of the model, demonstrating the predictive power of the hypothesized effects combined with individual-level heterogeneity and dynamics. The model predictive performance exceeds that of machine learning algorithms, indicating that the procedure has both academic and practical utility.

⁵Predictive performance was also examined for a manually resampled training sample. The overall predictive performance was poorer for all algorithms, see table A.1 in the appendix for results.

⁶This assumption is of course unrealistic, and the incremental savings of the proposed model presented here is significantly lower. However, the incremental improvements provided by the proposed model are likely exceptionally valuable given the large scale of credit card lending operations. Comparisons of simulations is considered beyond the scope of this thesis.

Table 4.6: Comparison of predictive performance, unaltered training data

	Default-only Test Sample, Unbalanced Training Data								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.534	0.232	-0.076	0.525	0.227	-0.061	0.564	0.221	-0.127
DNN ensemble	0.464	0.203	-0.116	0.522	0.201	-0.023	0.554	0.201	-0.100
GBM	0.513	0.207	0.103	0.554	0.214	0.129	0.539	0.189	0.107
GBM ensemble	0.521	0.216	0.132	0.547	0.222	0.167	0.544	0.198	0.076
DNN-GBM ensemble	0.509	0.208	0.075	0.530	0.217	0.090	0.547	0.221	0.112
XGBoost	0.529	<i>0.174</i>	0.071	0.596	<i>0.162</i>	0.092	0.555	<i>0.168</i>	0.081
Proposed model w/ k NN	<i>0.707</i>	0.218	<i>0.320</i>	<i>0.715</i>	0.230	<i>0.303</i>	<i>0.717</i>	0.236	<i>0.375</i>
	Default-only Test Sample, Balanced Training Data								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.540	0.354	0.006	0.497	0.267	0.010	0.513	<i>0.156</i>	0.024
DNN ensemble	0.422	0.272	0.015	0.541	0.283	0.102	0.494	0.158	0.009
GBM	0.531	0.291	0.135	0.546	0.283	0.052	0.499	0.163	0.115
GBM ensemble	0.539	0.282	0.174	0.549	0.290	0.075	0.515	0.163	0.097
DNN-GBM ensemble	0.539	0.254	0.138	0.539	<i>0.267</i>	0.074	0.528	0.164	0.145
XGBoost	0.519	0.327	0.028	0.576	0.314	0.134	0.507	0.163	0.160
Proposed model w/ k NN	<i>0.634</i>	<i>0.184</i>	<i>0.255</i>	<i>0.634</i>	0.269	<i>0.177</i>	<i>0.625</i>	0.266	<i>0.254</i>
	Balanced Test Sample								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.513	0.258	-0.122	0.550	0.254	0.139	0.493	0.252	-0.021
DNN ensemble	0.508	0.232	-0.058	0.575	0.228	0.178	0.535	0.227	-0.020
GBM	0.535	0.275	0.016	0.551	0.272	0.012	0.506	0.268	-0.028
GBM ensemble	0.538	0.272	0.034	0.550	0.274	-0.011	0.507	0.269	-0.042
DNN-GBM ensemble	0.530	0.267	0.018	0.539	0.271	0.004	0.504	0.267	-0.025
XGBoost	0.532	0.266	-0.061	0.615	0.261	0.123	0.459	0.261	0.093
Proposed model w/ k NN	0.741	0.124	0.481	0.823	0.128	0.621	0.783	0.130	0.572

Classification metrics at 1, 2, and 3 months prior to observed delinquency. Training data for balanced test sample is chosen based on classification performance for default-only test sample. Bold font indicates best global classifier. Italic font indicates sample-specific best classifier. Algorithms as specified in section 3.3.2, classification metrics calculated as outlined in section 3.3.3: AUC = Area under the ROC curve, BS = Brier score, PG = partial Gini index. A perfect classifier would have AUC=1, PG=1, and BS=0.

Table 4.7: Comparison of predictive performance, congruent sampled training data

	Default-Only Test Sample, Unbalanced Training Data								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.646	0.124	0.237	0.645	0.120	0.196	0.562	0.134	0.250
DNN ensemble	0.563	0.087	0.144	0.561	<i>0.090</i>	0.116	0.504	<i>0.100</i>	0.028
GBM	0.537	0.096	0.133	0.581	0.101	0.033	0.519	0.109	0.033
GBM ensemble	0.522	<i>0.086</i>	0.164	0.542	0.094	-0.040	0.510	0.103	0.098
DNN-GBM ensemble	0.538	0.108	0.173	0.607	0.092	0.161	0.514	0.101	0.092
XGBoost	0.587	0.091	0.191	0.604	0.100	0.078	0.485	0.111	0.074
Proposed model w/ <i>k</i> NN	<i>0.760</i>	0.158	<i>0.503</i>	0.813	0.153	0.601	0.786	0.168	0.552
	Default-only Test Sample, Balanced Training Data								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.558	<i>0.296</i>	0.103	0.640	<i>0.266</i>	0.195	0.557	0.412	-0.053
DNN ensemble	0.527	0.378	0.049	0.632	0.284	0.045	0.521	0.375	0.030
GBM	0.607	0.494	0.155	0.645	0.472	0.091	0.482	0.529	0.114
GBM ensemble	0.615	0.477	0.216	0.605	0.480	0.115	0.540	0.528	0.181
DNN-GBM ensemble	0.606	0.454	0.171	0.645	0.434	0.108	0.540	0.504	0.123
XGBoost	0.612	0.488	0.151	0.680	0.463	0.230	0.462	0.532	0.081
Proposed model w/ <i>k</i> NN	<i>0.691</i>	0.298	<i>0.406</i>	<i>0.726</i>	0.300	<i>0.413</i>	<i>0.681</i>	<i>0.347</i>	<i>0.357</i>
	Balanced Test Sample								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.590	0.272	0.069	0.657	0.327	0.252	0.523	0.340	-0.026
DNN ensemble	0.607	0.265	0.146	0.652	0.283	0.267	0.524	0.293	0.011
GBM	0.704	0.362	0.302	0.739	0.353	0.389	0.647	0.364	0.199
GBM ensemble	0.704	0.372	0.315	0.696	0.370	0.307	0.666	0.380	0.243
DNN-GBM ensemble	0.693	0.345	0.299	0.725	0.337	0.374	0.633	0.343	0.197
XGBoost	0.717	0.345	0.307	0.768	0.347	0.429	0.670	0.361	0.223
Proposed model w/ <i>k</i> NN	0.761	0.055	0.505	<i>0.775</i>	0.052	<i>0.539</i>	<i>0.731</i>	0.055	<i>0.425</i>

Classification metrics at 1, 2, and 3 months prior to observed delinquency. Training data for balanced test sample is chosen based on classification performance for default-only test sample. Bold font indicates best global classifier. Italic font indicates sample-specific best classifier. Algorithms as specified in section 3.3.2, classification metrics calculated as outlined in section 3.3.3: AUC = Area under the ROC curve, BS = Brier score, PG = partial Gini index. A perfect classifier would have AUC=1, PG=1, and BS=0.

5 — Discussion and Conclusion

This chapter discusses the results related to specific theoretical concepts in section 5.1, while general theoretical and methodological implications are detailed in sections 5.1.4 and 5.1.5. Implications for management are delineated in section 5.2, followed by the conclusion and limitations of the thesis.

5.1 Discussion of the Results

The main research objective in this thesis was constructing a fundamental model for behavioral scoring that includes debit data. The mental accounting concepts suggested to influence cardholder delinquency were decoupling, persistent habits indicating poor financial understanding, and present bias. Conceptually, the effects were entered into a model of the value function that a cardholder faces when deciding to repay or continue borrowing. Equation 2.7 presents the chorus of effects, where the parameters for decoupling rate, persistence of habit and financial aptitude, and present bias reveal the proposed effects on the decision to borrow.

Cardholders choose to borrow when the value of continued borrowing is higher than the value of repaying at higher levels (and lower consumption levels). The theory chapter argued that cardholders who are more likely to become delinquent are, on average, more prone to all the purported effects. In essence, the value of continuing to borrow is higher when the cardholder exhibits higher rates of decoupling, decision-making ineptitude, and present-

bias. While these effects cannot be observed and measured outright, the borrowing behavior, which is implied, can be approximated. Higher decoupling rates lead to lower tendencies of repayment, persistence of unsound habits will lead to sustained high borrowing levels, while present-bias is revealed through accelerated spending. The results confirm these hypotheses, while the hypothesis involving debt-clearing ability as implied by a rational choice model is not confirmed. In addition, the concurrent effects do not significantly explain delinquency, further undermining the rational choice model of financial behavior. The following sections discuss these theoretical findings separately, as well as the joint ramifications to theory and the practice of delinquency prediction.

The hypothesized effects from decoupling (section 5.1.1), financial understanding and spending habits (section 5.1.2), and present bias (section 5.1.3) are discussed in the following sections.

5.1.1 Decoupling

Decoupling of payment and consumption leads to a perceived separation of the cost and benefits. Concerning the hedonic editing concept, credit card payments are likely difficult to integrate with the utility the purchases provided. As individuals will frame a car loan and the utility from that car in different accounts (Kamleitner and Kirchler, 2006), the utility of past discretionary spending (as credit card expenses are likely to be) is decoupled from the disutility of paying the credit card bill. The pain of paying (Kivetz, 1999) is lower for credit payments than for cash payments, and the salience of expenses is lower (Soman, 2001*a*). As set forth by Thaler (1980), when consumption is not immediate following payment (given a reasonable cost), the experience of consumption and payment is often in disagreement with economic theory. When assessing credit card consumption, the separation of payment and consumption makes the cost less salient (Soman, 2001*a*), promoting usage (Thaler, 1999).

While credit cards decouple the costs and benefits of consumption by construction, the tendency to decouple is likely heterogeneously distributed. As suggested in this thesis and by Prelec and Loewenstein (1998), these tendencies are likely part of the explanation of why we observe individual differences in spending habits. Some individuals spend indiscreetly without deliberation, while others are exceedingly mindful of past expenses. Empirically, the extent of decoupling has been shown to vary between individuals for home loans (Hoelzl et al., 2009), but not explicitly for credit card debt.

This thesis presents a model in which decoupling enters the borrowing decision, where decoupling will promote continued borrowing through reduced salience of past expenses. In turn, reduced salience of past expenses reduces the propensity to repay at sufficient levels. A cardholder with higher decoupling rates will value continued consumption (and borrowing) higher than repayment (and reduced consumption). While measuring decoupling is impractical, past repayment data are readily available. Accordingly, higher tendencies to decouple (operationalized as lagged repayment) will lead to eventual delinquency, which the results confirm. In addition, the effect of lagged repayment is almost entirely relegated to delinquent cardholders, suggesting that behavior of non-delinquent cardholders follows the conventional truism that only current (and future) assets should enter into financial decisions. Repaying credit card debt is reportedly highly prioritized (Prelec and Loewenstein, 1998), and the corollary is that low repayment is a useful indicator of behavior. Here, this behavior captures the level of decoupling among cardholders. Although paying off credit card debt appears to provide high utility, the disutility of foregoing present utility from continued interrupted consumption is somehow given more weight among eventually delinquent cardholders.

5.1.2 Decision-making Ineptitude

Empirically, numerous personality traits have been proposed to influence borrowing. Pirog and Roberts (2007) investigate four elemental personality traits, of which impulsiveness is estimated to be the central trait governing credit card misuse.¹ In addition to impulsiveness, external locus of control, self-efficacy, and self-esteem are also related to revolving credit use (Wang et al., 2011). External locus of control could be considered the inverse of willpower, as it measures the feeling an individual has of control over what happens in his life. Consistent with lack of willpower, individuals with an external locus of control also experience a low ability to manage finances (Perry and Morris, 2005). However, general personality traits have also shown weaker associations when examined with actual debt carrying among students (Norvilitis et al., 2003), although the same study indicated that personality is correlated with attitudes toward money management. This suggests that the subtle individual differences are temporally mediated by money management, subsequent to impacting our actual financial status.

Deconstructing the effects of personality traits reveals how financial literacy is a significant predictor of borrowing and financial distress (Hilgert et al., 2003). Financial literacy indicates the level of understanding regarding personal finances and financial decision-making in general in individuals. Financial aptitude and financial knowledge are naturally related to financial decision-making (Disney and Gathergood, 2013; Klapper et al., 2013; Perry and Morris, 2005) and provide useful predictors of debt collection in concert with general personality traits (Gathergood, 2012). While low financial literacy could likely be improved by experience or tutoring, the aptitude of making and understanding financial decisions is likely an enduring trait (Kamleitner and Kirchler, 2007). Certain individuals likely have a lower predilection of making sound financial decisions.

¹Credit card misuse was measured with a self-reported likert scale. Validity of results may be questioned, as all measures were conducted simultaneously in a single survey.

As a consequence, the heterogeneously distributed capability of making sound financial decisions is likely to impact delinquency. Some cardholders will, perhaps perilously, attempt to uphold consumption by continually borrowing on their credit card. This is indicated by the parameter measuring decision-making ineptitude, where a higher value would signify low financial competence. This behavior is akin to a trait, in that it engenders some permanence. Continued onerous consumption in spite of a worsening financial outlook suggests a lack of financial understanding. Accordingly, these individuals will abandon comprehensive evaluations of economic decisions and adopt heuristics such as assigning costs to different mental accounts. Cardholders with high decision-making ineptitude will value their continued consumption higher than increased repayment (and proportionally lower consumption), which eventually leads to delinquency. While the presence of this tendency is unobservable, the presence of habitual borrowing is. Thus, decision-making ineptitude is measured by a lagged credit card balance ratio, which measures continued careless borrowing. The results confirm the effects of decision-making ineptitude on delinquency; a lagged balance ratio is significant, while current balance is not. As with decoupling, the effect is nearly non-existent for non-delinquent cardholders. This provides further support for the mental accounting framework and conversely undermines the rational choice model of decision-making.

5.1.3 Present Bias

Present bias is often considered a behavioral predisposition in the study of borrowing and financial prudence. The extent to which individuals are influenced by their future outcomes is specific to individuals (Joireman et al., 2008) and has been linked to less compulsive buying and long-term investments (Joireman et al., 2005), as well as saving for retirement (Howlett et al., 2008). When related to credit card debt, present bias is usually concerned with the act of borrowing or the amount of debt chosen to carry (e.g., Meier

and Sprenger, 2010; Agarwal et al., 2013; Gourville and Soman, 1998; Joireman et al., 2010).

Inherent in the valuation of future utility in monetary terms is the psychological analysis of time. Judgment of time is considerably different than consideration of money, such that accounting of time is different than accounting of money (Soman, 2001*b*). Referencing the dual processing system of cognition (e.g., Kahneman, 2003*b*), time is usually considered in affective terms, while money is considered in analytical terms (Lee et al., 2015). While monetary decisions are generally effortful and based on reasoning, the addition of time in the equation adds an emotional aspect to otherwise rational decisions.² Time adds complexity to all economic decisions, whereby sensitivity to duration will govern the eventual choice a consumer makes. In support of the apparent distortions of objective time when viewed subjectively, Zaubermaier et al. (2009) show how present bias could be explained as a manifestation of our shortfalls of assessing time when judging present utility.

Related to the concern for immediate gratification is the level of self-control cardholders exert and the amount of self-control at their disposal. Viewed as a depletable reservoir, self-control has been linked to irresponsible behavior, such as alcohol consumption (Muraven et al., 2002) and overindebtedness (Gathergood, 2012). Lack of self-control promotes the usage of costly borrowing (Gathergood, 2012) and is related to revolving credit use (Wang et al., 2011).

The concept of self-control has understandably drawn interest from an economic perspective; most notable is perhaps the seminal study by Thaler and Shefrin (1981) in which a farsighted planner has the ability to provide rules to be followed by the myopic doer. The concept of providing rules, a

²This dual processing framework is referenced as “system 1” and “system 2” by Daniel Kahneman, where the first system is based on intuition and quick, effortless decisions, while the latter system is a slower, controlled system governed by cognitive rules (Kahneman, 2003*b*).

concept related to willpower and self-regulation, guides decision-making and is supported by self-confidence (Bénabou and Tirole, 2004). Willpower is, however, also thought of as a heterogeneously distributed resource, leading to consistent differences in the propensity to plan (Ameriks et al., 2003).

Present bias enters the model presented in this thesis as a discounting factor of future debt. A present-biased cardholder will discount future debt at a higher rate, increasing the value of continued borrowing. Low consideration of future consequences promotes higher current consumption, allowing a careless cardholder to continue financing consumption with debt. As with decoupling and financial aptitude, the discounting parameter is not observable. However, spending behavior that signifies present bias is observed. Present bias is measured as accelerated spending following the monthly paycheck. Accelerated spending implies a lack of self-control or willpower, and a history of present-biased behavior is expected to promote delinquency. The results confirm this suspicion; the lagged effects of pay-day spending has a significant effect on delinquency. Again, the concurrent effect is not significant, suggesting that incidental accelerated spending is unrelated to delinquency. Present bias is likely a durable characteristic, and the results suggest it should be measured as such.

5.1.4 Implications for Theory

The findings support the notion of differing psychological traits and mental accounting strategies (Brockett and Golden, 2007; Kamleitner and Kirchler, 2006; Ameriks et al., 2003), such as subjective evaluation of time (Overton and MacFadyen, 1998). As hypothesized, these individual characteristics give rise to differing rates of borrowing and repayment, which eventually lead to delinquency. The proposed model captures these effects through prior financial states, confirming that persistent behavioral predispositions significantly affect decision-making.

Behavior scoring models benefit by including insights from mental ac-

counting theory. Data-driven approaches, such as Zhao et al. (2009) and Khandani et al. (2010), lack important theoretical insights to explain behavior and subsequent delinquency. Understanding and explaining heterogeneous behavior regarding credit card borrowing requires a framework to explain how individuals behave when spending and collecting debt. Choices are less influenced by rational economic expectations of future standing; instead, individuals deviate from rational heuristics and rules and will adapt their mental accounts to accommodate credit spending (Cheema and Soman, 2006). While the rational choice model assuming the life-cycle hypothesis of spending and borrowing predicts a sensible forward-looking individual, the results here contradict this. According to the rational choice model, the current financial situation is a result of previous spending and borrowing, such that past behavior is reflected by current assets and credit balance. However, the results demonstrate that past behavior influences current and future likelihood of delinquency. The theoretical framework provided by mental accounting explicates the psychological processes that likely permeate decision-making, providing a foundation for and supporting the findings.

Figure 5.1 attempts to visually illustrate the necessity of a financially holistic approach while displaying the mental accounting effects when predicting cardholder delinquency. Specifically, subfigure a) plots the quantiles for immediate spending, monthly income, and balance, where immediate spending is the sum of debit expenses over the 10 days following the highest observed monthly income. As the results from the proposed model indicates, credit borrowing consistently grows, reaching an undesirable level where delinquency is reached. As predicted with the present bias concept (Meier and Sprenger, 2010), cardholders facing eventual or existing delinquency uphold their current spending levels in spite of a worsening financial outlook. In isolation, monthly income and immediate spending are generally stable, suggesting that individuals do not suffer from sudden financial shocks, such as losing substantial income or sudden and devastating finan-

cial outlays. This is especially evident in the lower barplot in subfigure a), where the quantiles of all individual histories leading to a 90% estimated likelihood of delinquency are presented. Income and immediate spending are consistent, while borrowing steadily increases.³ This supports the modeling procedure twofold: first, the likelihood of delinquency will not be projected from a sudden event, and second, projecting future delinquency months in advance necessitates a behavioral understanding of the dilemma.

The importance of behavioral modeling of the delinquency dilemma is also illustrated in subfigure b), where the 25% and 75% individual quantiles for the same amounts are displayed in somewhat overlapping bands. Immediate spending is generally lower than the monthly income band for delinquent and non-delinquent cardholders. Viewed in isolation, the quantiles do not appear considerably different, though comprehensively the graphics illustrate the worsening state that faces eventually delinquent cardholders. This indicates why a thorough examination and modeling of the financial situation is a requirement for explaining delinquency. Mental accounting affords an integrative framework, and the results in this thesis suggests that the framework set forth by Thaler (1985) should hold a more central role in delinquency considerations and spending behavior in general.

However, the main tenet of the mental accounting framework also complicates empirical applications thereof. The framework seeks to expand the idea of choices made by rational individuals by admitting that evaluations are subjective and prone to suboptimal choices. While the model presented here suggests that nonstandard beliefs and decision-making in the form of payment decoupling, persistent ineptitude, and present bias all influence delinquency likelihood, the effects presented could be encompassed by other theoretical constructs. Put differently, the discriminant validity of the theoretical

³This is more evident when assessing the median quantiles, where the median amount borrowed is approximately equal to median monthly income in months 4 to 6 leading up to a high delinquency likelihood. The median amount borrowed then steadily grows, reaching levels much higher than monthly income.

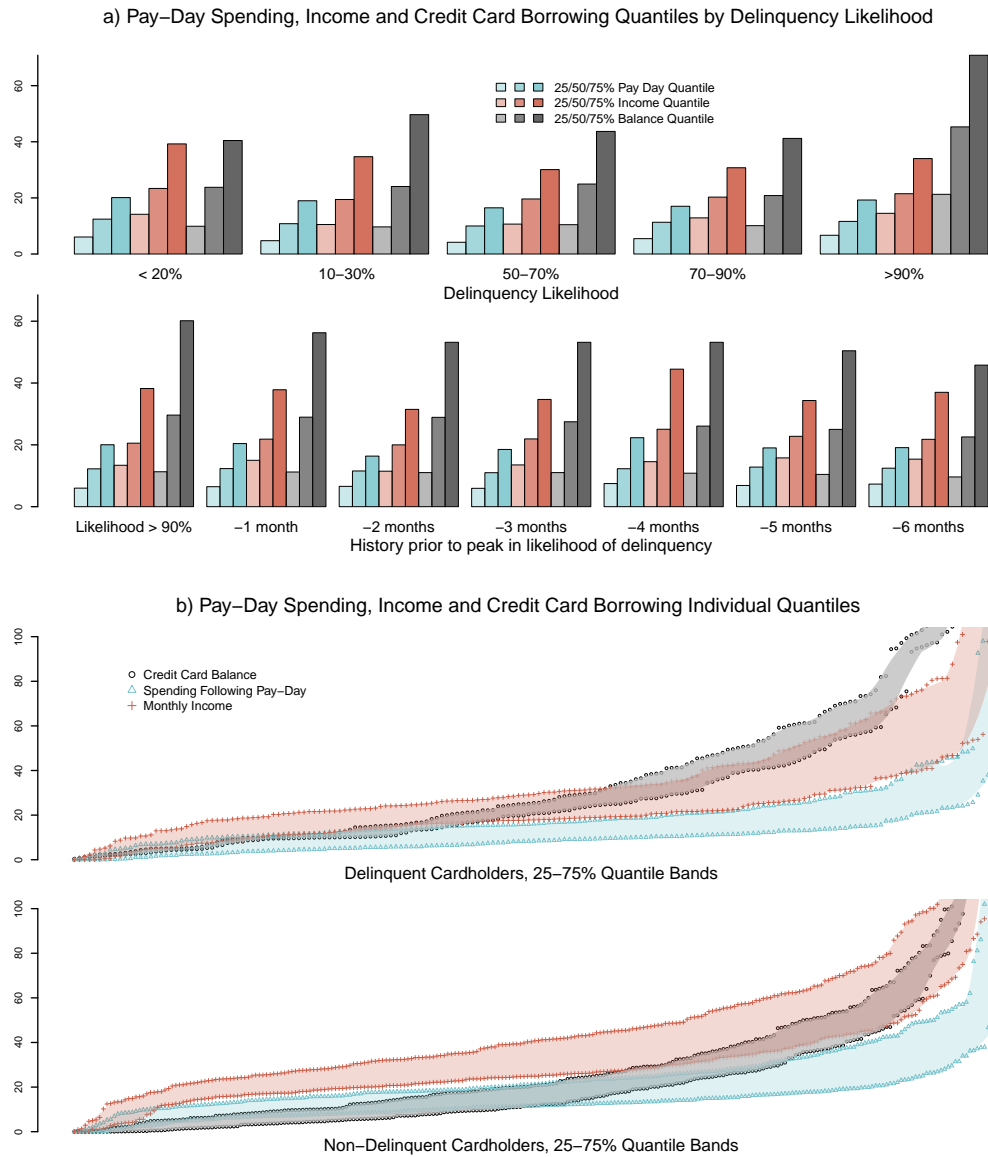


Figure 5.1: Barplots in subfigure a) report the 25%, 50%, and 75% quantiles by estimated delinquency likelihood for 10-day spending following income date, monthly total income, and credit card balance by delinquency likelihood and history prior to a 90% likelihood peak. Subfigure b) plots the same quantiles by individuals for delinquent and non-delinquent cardholders.

constructs employed and other constructs presented in behavioral economics and mental accounting is not absolutely unequivocal. This caveat also applies to other empirical applications of mental accounting and behavioral economics; the theoretical constructs are somewhat entangled and difficult to differentiate when examining several concepts simultaneously. The mental accounting framework performs well in many experimental contexts (as seen in DellaVigna, 2009), but heeds caution when applied in empirical contexts.

5.1.5 Implications for Classification Methods

Shedding light on the process leading to delinquency not only removes ambiguity regarding the statistical reliability, it also has practical implications. Perhaps unexpectedly in a machine learning application of behavior scoring, Khandani et al. (2010) relay the importance of providing meaningful models for "...the banking sector in which 'black-box' models are viewed with suspicion and skepticism." Choosing an appropriate structure is significant from not only an academic viewpoint but also from a practical stance. Comparing the proposed model with state-of-the-art machine learning algorithms in predictive performance illustrates this. When classifying an observed event that is the result of ongoing behavior, paying heed to the information provided by the structure of individual histories is a palpable prerequisite.

The modeling structure itself also provides a meaningful contribution. The proposed model expands previous research that underscores the importance of capturing heterogeneity and dynamic effects (Zhao et al., 2009), allowing for a more nuanced specification for capturing persistent behavior. The segmented lag weight approach with random lag selection not only supports the notion of prior choices affecting our current behavior, it also hints at the magnitude of individual differences in financial behavior.

Evaluating global performance, the proposed model outperforms machine learning classification algorithms that have been prevalent and successful in credit scoring (Lessmann et al., 2015; Louzada et al., 2016). The classifiers

are evaluated using recommended metrics for credit scoring (Lessmann et al., 2015), showcasing global and local predictive performance (AUC and partial Gini coefficient), and accuracy of predicted probabilities (Brier score). The hierarchical Bayesian logit scheme used sequentially with a k -NN approach illustrates the potential in structured approaches to classification when applicable. Specifically, classification problems with clearly structured data (such as credit histories nested within individuals), are better solved using classifiers that heed the inherent structure. The results suggest that the added explanatory value of incorporating individual-level heterogeneity and dynamics outweighs the raw predictive power of machine learning algorithms.

Collectively, this could point to the inherent differences between application scoring and behavior scoring. Despite the many similarities, the longitudinal structure of observations in a behavior scoring classification problem heavily favors models that replicate histories directly, as opposed to the machine learning algorithms commonly employed in application scoring. As a corollary, this slight disharmony also influences the bias towards application scoring for machine learning algorithms, though likely to a lesser degree than practical concerns of data availability. The current thesis underscores the uniqueness of behavior scoring, demonstrated by the structure and results of the proposed model.

5.2 Managerial Implications

Current customer value analysis of existing and potential cardholders is largely a function of the available data. Avoiding the adverse selection problem by applying advanced application scoring techniques has steadily evolved for several decades, from simple discriminant analysis to advanced non-parametric approaches. Valuation of existing customers through behavior scoring has quickly matured from logistic regression (Hamilton and Khan, 2001) to Tobit specifications (Zhao et al., 2009) and machine learning

(Khandani et al., 2010), while this thesis proposes a model that specifies a sophisticated dynamic structure to replicate financial behavior.

Behavior scoring attempts to predict future delinquency by examining observed credit card usage behavior. Improving the performance of behavior scoring models allows managers to have a better foundation to make decisions related to their customer portfolio of cardholders. The proposed model in this thesis shows improved accuracy over existing techniques, leading to more accurate targeting of cardholders who are potentially facing financial difficulties as well as avoiding misclassification of non-delinquent cardholders. With improved classification rates, a manager can confidently act on the insights gleaned from the analysis, choosing whether to freeze accounts, reduce limits, or instigate forms of impersonal or personalized financial advice.

The findings deduced from the improved model structure suggest a broader set of implications for practice related to financial decision-making and consumer credit usage in particular. While the study of non-mortgage consumer loans has graduated past the assumptions of hyper-rational economic theory to include behavioral aspects, such as mental accounting (Thaler, 1999), time discounting (Lee et al., 2015), and payment decoupling (Soman, 2001 *a*), the application of these insights is still beyond their present incarnations of correlational and experimental usages. Mental accounting hypotheses, such as those relied upon in this thesis, belong in analyses of financial behavior and should be treated with equal regard as conventional wisdom, such as the consumption smoothing argument for credit uptake. This thesis demonstrates the applicability of behavioral aspects to financial decision-making when taking a comprehensive approach to those decisions; not only are the behavioral aspects of financial decisions valid and significant, but they are inherently essential.

5.3 Limitations and Directions for Further Research

The main limitation of this thesis is the reduction of data to construct a model that deals with inter-monthly behavior. Transactional data is available for both credit and debit transactions, which provides a possibility for modeling more intricate decision-based models. This could possibly allow for a model in which the individual transactions provide insights into how cardholders differ in processing each expense, perhaps answering questions as to which individuals are likely to spend excessively when feeling inclined to shop and which individuals will spend excessively using a credit card when a decision to spend has been made. While the proposed model demonstrates the importance of financial behavior in predicting future delinquency, employing more precise disaggregate transactional data in a meaningful structure will likely improve prediction and realism.

The operationalization of the theoretical constructs and their subsequent relation to independent variables are somewhat convoluted. Mental accounting, present bias, and related theories are suited for experiments studying singular effects, while the construction of a realistic model of behavior demands appropriately constructed theories. Hence, the proposed effects often coincide, making theory testing problematic. For example, isolating the effect from irrational behavior due to decision-making ineptitude is quixotic when also dealing with the effect from payment decoupling. It is expected that these effects work in tandem when examining credit card spending, although the precise significance and explanatory power of the individual effects are not easily ascertained. While correlational studies using surveys and financial data have attempted to delineate psychological traits related to credit uptake with varying success (e.g., Norvilitis et al., 2003), constructing a model of cardholder behavior based on somewhat fragmented theories will eschew some epistemological precision.

Other limitations of this thesis are related to the data available. Credit application information available in this database was severely flawed, as demographic and financial information at the time of application was outdated and sporadically recorded. Also, the cardholders quite possibly held credit cards from different providers, in addition to other consumer loans and non-mortgage loans. The data assessed here is only from one lending institution, which means other important information was possibly missing.

Future research should attempt to extend the theoretical modeling structure presented in this thesis. Disaggregate financial data should allow more rigorous testing of the proposed theoretical effects. For example, while the precision of present bias with regard to measurement, composition, and effect has steadily increased (Lee et al., 2015; Malkoc and Zauberma, 2006; Zauberma et al., 2009), the practical significance of the concept is less scrutinized. Connecting and comparing present bias with mechanisms, such as the pain of paying (Prelec and Loewenstein, 1998) or payment decoupling (Thaler, 1999), disaggregate financial data will allow construction of models that explicitly consider these effects in tandem, in isolation, or in conjunction. This thesis firmly contends that inclusion of these mechanisms is crucial in assessing delinquency, while future studies should investigate the relatedness and interplay of these concepts.

Finally, a comprehensive model of payment choice could be constructed. Combining disaggregate financial data with panel shopper data should allow modeling of not only shopping decisions but also of choice of payment for each shopping occasion. Choosing debit or credit payment is likely done at different stages, which in turn will affect the amount spent. Increased nuance of financial data registered could possibly allow such an investigation, although current information is inadequate.

5.4 Conclusion

This thesis provides a fundamental model of credit card customer profitability. Utilizing insights from mental accounting (Thaler, 1985), a model is constructed to capture the effects of payment decoupling (e.g., Thaler, 1999; Zeelenberg and van Dijk, 1997), decision-making aptitude (Agarwal and Mazumder, 2013), and present-biased preferences (Meier and Sprenger, 2010; Zauberman et al., 2009). The thesis provides an extension of previous work into behavior scoring (Zhao et al., 2009). The proposed model outperforms modern machine learning algorithms, providing evidence of the usefulness of mental accounting theory in understanding financial behavior.

This thesis provides three main contributions to mental accounting theory. First, this is the only current application of mental accounting to use observed delinquency rates coupled with observed financial behavior. Second, the results underscore the links between subjective time assessment, decoupling, and decision-making ineptitude. In credit card usage and repayment behavior, individuals will generally be influenced by several of these variables. Third, the results illustrate the importance of prior decisions as opposed to current finances, where the latter is the main explanatory data when assessing delinquency likelihood in the current literature.

The thesis also provides a significant contribution to the behavior scoring literature, implementing theoretical concepts as opposed to using a data-driven approach to model construction. The proposed model allows for a varying segment structure with random lag lengths, showing how delinquent cardholders are significantly more dependent on previous financial states than non-delinquent cardholders. Improving the precision of the structure of dynamic effects greatly enhances estimation and prediction, as well as giving more nuanced insights into the behavioral differences between cardholders who become delinquent compared to cardholders who do not encounter payment difficulties.

6 — Bibliography

Agarwal, S., Bubna, A. and Lipscomb, M. (2013), ‘Timing to the statement: Understanding fluctuations in consumer credit use’. Accessed March 17, 2015.

URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2187604

Agarwal, S., Liu, C. and Souleles, N. S. (2007), ‘The reaction of consumer spending and debt to tax rebates - evidence from consumer credit data’, *Journal of Political Economy* **115**(6), 986–1019.

Agarwal, S. and Mazumder, B. (2013), ‘Cognitive abilities and household financial decision making’, *American Economic Journal: Applied Economics* **5**(1), 193–207.

Ali, S. N. (2011), ‘Learning self-control’, *The Quarterly Journal of Economics* **126**, 857–892.

Allenby, G. M. and Rossi, P. E. (1999), ‘Marketing models of consumer heterogeneity’, *Journal of Econometrics* **89**, 57–78.

Amar, M., Ariely, D., Ayal, S., Cryder, C. E. and Rick, S. I. (2011), ‘Winning the battle but losing the war: The psychology of debt management’, *Journal of Marketing Research* **48**(Special Issue), 38–50.

Ameriks, J., Caplin, A. and Leahy, J. (2003), ‘Wealth accumulation and the propensity to plan’, *Quarterly Journal of Economics* **118**(3), 1007–1047.

- Ando, A. and Modigliani, F. (1963), ‘The” life cycle” hypothesis of saving: Aggregate implications and tests’, *The American economic review* **53**(1), 55–84.
- Antonides, G., de Groot, I. M. and van Raaij, W. F. (2011), ‘Mental budgeting and the management of household finance’, *Journal of Economic Psychology* **32**, 546–555.
- Ariely, D. (1998), ‘Combining experiences over time: the effects of duration, intensity changes and on-line measurements on retrospective pain evaluations’, *Journal of Behavioral Decision Making* **11**(1), 19–45.
- Arkes, H. R. and Blumer, C. (1985), ‘The psychology of sunk cost’, *Organizational Behavior and Human Decision Processes* **35**, 124–140.
- Arora, A., Candel, A., Lanford, J., LeDell, E., and Parmar, V. (2015), *Deep Learning with H2O*.
URL: <http://h2o.ai/resources>
- Atlas, S. A., Johnson, E. J. and Payne, J. W. (2017), ‘Time preferences and mortgage choice’, *Journal of Marketing Research* **54**(3), 415–429.
- Ausubel, L. M. (1991), ‘The failure of competition in the credit card market’, *American Economic Review* **81**(1), 50–81.
- Barberis, N. C. (2013), ‘Thirty years of prospect theory in economics: A review and assessment’, *Journal of Economic Perspectives* **27**(1), 173–196.
- Bartels, D. M. and Urminsky, O. (2015), ‘To know and to care: How awareness and valuation of the future jointly shape consumer spending’, *Journal of Consumer Research* **41**(6), 1469–1485.
- Bénabou, R. and Tirole, J. (2004), ‘Willpower and personal rules’, *Journal of Political Economy* **112**(4), 848–886.

- Berger, P. D., Bolton, R. N., Bowman, D., Briggs, E., Kumar, V., Parasuraman, A. and Terry, C. (2002), ‘Marketing actions and the value of customer assets’, *Journal of Service Research* **5**(1), 39–54.
- Berman, J. Z., Tran, A. T., Lynch, Jr., J. G. and Zauberaman, G. (2016), ‘Expense neglect in forecasting personal finances’, *Journal of Marketing Research* **53**(4), 535–550.
- Bettman, J. L., Sault, S. J. and Schultz, E. L. (2009), ‘Fundamental and technical analysis: substitutes or complements?’, *Accounting and Finance* **49**, 21–36.
- Bradley, A. P. (1997), ‘The use of the area under the roc curve in the evaluation of machine learning algorithms’, *Pattern Recognition* **30**(7), 1145–1159.
- Breiman, L. (1996), ‘Bagging predictors’, *Machine Learning* **24**(2), 123–140.
- Breiman, L. (2001), ‘Random forests’, *Machine Learning* **45**(1), 5–32.
- Bricker, J., Dettling, L. J., Henriques, A., Hsu, J. W., Moore, K. B., Sabelhaus, J., Thompson, J. and Windle, R. A. (2014), ‘Changes in U.S. family finances from 2010 to 2013: Evidence from the survey of consumer finances’, *Federal Reserve Bulletin* **100**(4).
- Brito, D. L. and Hartley, P. R. (1995), ‘Consumer rationality and credit cards’, *Journal of Political Economy* **103**(2), 400–433.
- Brockett, P. L. and Golden, L. L. (2007), ‘Biological and psychobehavioral correlates of credit scores and automobile insurance losses: Toward and explication of why credit scoring works’, *The Journal of Risk and Insurance* **74**(1), 23–63.

- Brown, A. L., Chua, Z. E. and Camerer, C. F. (2009), ‘Learning and visceral temptation in dynamic saving experiments’, *The Quarterly Journal of Economics* pp. 197–231.
- Brown, A. L. and Lahey, J. N. (2015), ‘Small victories: Creating intrinsic motivation in task completion and debt repayment’, *Journal of Marketing Research* **52**(6), 768–783.
- Brown, S., Taylor, K. and Price, S. W. (2005), ‘Debt and distress: Evaluating the psychological cost of credit’, *Journal of Economic Psychology* **26**, 642–663.
- Calem, P. S. and Mester, L. J. (1995), ‘Consumer behavior and the stickiness of credit-card interest rates’, *American Economic Review* **85**(5), 1327–1336.
- Canner, G. B. and Elliehausen, G. (2013), ‘Consumer experiences with credit cards’, *Federal Reserve Bulletin* **99**(5).
- Carmon, Z. and Kahneman, D. (1996), ‘The experienced utility of queuing: Real-time affect and retrospective evaluations of simulated queues’. Working Paper. Accessed January 1, 2014.
URL: <https://bit.ly/2wPUOGn>
- Chatterjee, S., Heath, T. B. and Kim, J. (2009), ‘The susceptibility of mental accounting principles to evaluation mode effects’, *Journal of Behavioral Decision Making* **22**, 120–137.
- Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002), ‘Smote: Synthetic minority over-sampling technique’, *Journal of Artificial Intelligence Research* **16**, 321–357.
- Cheema, A. and Soman, D. (2006), ‘Malleable mental accounting: The effect of flexibility on the justification of attractive spending and consumption decisions’, *Journal of Consumer Psychology* **16**(1), 33–44.

- Chen, T. and Guestrin, C. (2016), XGBoost: A scalable tree boosting system, *in* ‘Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining’, KDD ’16, ACM, New York, NY, USA, pp. 785–794.
- Cheng, A. and Cryder, C. (2018), ‘Double mental discounting: When a single price promotion feels twice as nice’, *Journal of Marketing Research* **55**(2), 226–238.
- Click, C., Lanford, J., Malohlava, M., Parmar, V. and Roark, H. (2015), *Gradient Boosted Models*.
URL: <http://h2o.ai/resources>
- Cowley, E. (2008), ‘The perils of hedonic editing’, *Journal of Consumer Research* **35**, 71–84.
- Dal Palozzolo, A., Caelen, O., Waterschoot, S. and Bontempi, G. (2013), Racing for unbalanced methods selection, *in* ‘Proceedings of the 14th International Conference on Intelligent Data Engineering and Automated Learning - IDEAL 2013’, Springer Berlin Heidelberg, pp. 24–31.
- Davies, E. and Lea, S. E. (1995), ‘Student attitudes to student debt’, *Journal of Economic Psychology* **16**(4), 663–679.
- Davies, S., Easaw, J. and Ghoshray, A. (2009), ‘Mental accounting and remittances: a study of rural Malawian households’, *Journal of Economic Psychology* **30**(3), 321–334.
- DellaVigna, S. (2009), ‘Psychology and economics: Evidence from the field’, *Journal of Economic Literature* **47**(2), 315–372.
- DeVoe, S. E. and Pfeffer, J. (2007), ‘When time is money: The effect of hourly payment on the evaluation of time’, *Organizational Behavior and Human Decision Processes* **104**(1), 1–13.

- Disney, R. and Gathergood, J. (2013), ‘Financial literacy and consumer credit portfolios’, *Journal of Banking and Finance* **37**, 2246–2254.
- Dunn, L. F. and Kim, T. (2006), ‘An empirical investigation of credit card default’. Working Paper. Accessed August 1, 2015.
URL: <http://economics.sbs.ohio-state.edu/pdf/credit.pdf>
- Duxbury, D., Keasey, K., Zhang, H. and Chow, S. L. (2005), ‘Mental accounting and decision making: Evidence under reverse conditions where money is spent for time saved’, *Journal of Economic Psychology* **26**(4), 567–580.
- Federal Reserve (2018), ‘The federal reserve payments study: 2017 annual supplement’. Accessed September 6th 2018.
URL: <https://www.federalreserve.gov/paymentsystems/2017-December-The-Federal-Reserve-Payments-Study.htm>
- Feldman, N. E. (2010), ‘Mental accounting effects of income tax shifting’, *Review of Economics and Statistics* **92**(1), 70–86.
- Frederick, S. (2005), ‘Cognitive reflection and decision making’, *Journal of Economic Perspectives* **19**(4), 15–42.
- Frederick, S., Loewenstein, G. and O’Donoghue, T. (2002), ‘Time discounting and time preference: A critical review’, *Journal of Economic Literature* **40**, 351–401.
- Fredrickson, B. L. and Kahneman, D. (1993), ‘Duration neglect in retrospective evaluations of affective episodes’, *Journal of Personality and Social Psychology* **65**(1), 45–55.
- Freund, Y. and Schapire, R. E. (1996), Experiments with a new boosting algorithm, *in* ‘Machine Learning: Proceedings of the Thirteenth International Conference’, pp. 148–156.

- Friedman, J. H. (2002), ‘Stochastic gradient boosting’, *Computational Statistics & Data Analysis* **38**(4), 367–378.
- Gal, D. and McShane, B. B. (2012), ‘Can small victories help win the war? evidence from consumer debt management’, *Journal of Marketing Research* **49**, 487–501.
- Gathergood, J. (2012), ‘Self-control, financial literacy and consumer over-indebtedness’, *Journal of Economic Psychology* **33**, 590–602.
- Gelman, A. (2008), ‘Scaling regression inputs by dividing by two standard deviations’, *Statistics in Medicine* **27**, 2865–2873.
- Geman, S. and Geman, D. (1990), ‘Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images’, *IEEE transactions on Pattern Analysis and Machine Intelligence* **6**(6), 721–741.
- Gensler, S., Leeflang, P. and Skiera, B. (2013), ‘Comparing methods to separate treatment from self-selection effects in an online banking setting’, *Journal of Business Research* **66**, 1272–1278.
- Gourville, J. T. and Soman, D. (1998), ‘Payment depreciation: The behavioral effects of temporally separating payments from consumption’, *Journal of Consumer Research* **25**, 160–174.
- Guadagni, P. M. and Little, J. D. (1983), ‘A logit model of brand choice calibrated on scanner data’, *Marketing Science* **2**(3), 203–238.
- Gupta, S. and Chintagunta, P. K. (1994), ‘On using demographic variables to determine segment membership in logit mixture models’, *Journal of Marketing Research* **31**(1), 128–136.
- Ha, S. H. and Krishnan, R. (2012), ‘Predicting repayment of the credit card debt’, *Computers & Operations Research* **39**, 765–773.

- Hall, R. E. (1978), ‘Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence’, *Journal of Political Economy* **86**(6), 971–987.
- Hamilton, R. and Khan, M. (2001), ‘Revolving credit card holders: Who are they and how can they be identified’, *The Services Industries Journal* **21**(3), 37–48.
- Hamori, S., Kawai, M., Kume, T., Murakami, Y. and Watanabe, C. (2018), ‘Ensemble learning or deep learning? application to default risk analysis’, *Journal of Risk and Financial Management* **11**(1), 12.
- Hand, D. and Anagnostopoulos, C. (2013), ‘When is the area under the receiver operating characteristic curve an appropriate measure of classifier performance?’, *Pattern Recognition Letters* **34**(5), 492 – 495.
- Hand, D. and Henley, W. (1997), ‘Statistical classification methods in consumer credit scoring: A review’, *Journal of the Royal Statistical Society* **160**(3), 523–541.
- Hastings, J. S. and Shapiro, J. M. (2013), ‘Fungibility and consumer choice: Evidence from commodity price shocks’, *The Quarterly Journal of Economics* **128**(4), 1449–1498.
- Heath, T. B., Chatterjee, S. and France, K. R. (1995), ‘Mental accounting and changes in price: The frame dependence of reference dependence’, *Journal of Consumer Research* **22**(1), 90–97.
- Hershfield, H. E., Goldstein, D. G., Sharpe, W. F., Fox, J., Yeykelis, L., Carstensen, L. L. and Bailenson, J. N. (2011), ‘Increasing saving behavior through age-progressed renderings of the future self’, *Journal of Marketing Research* **48**(Special Issue), 23–27.

- Hilgert, M. A., Hogarth, J. M. and Beverly, S. G. (2003), 'Household financial management: The connection between knowledge and behavior', *Federal Reserve Bulletin* **89**(7), 309–322.
- Hilovich, C. and Gilovich, T. (2014), 'Gift cards and mental accounting: Green-lighting hedonic spending', *Journal of Behavioral Decision Making* **27**(4), 386–393.
- Hirst, D. E., Joyce, E. J. and Schaedewald, M. S. (1994), 'Mental accounting and outcome contiguity in consumer-borrowing decisions', *Organizational Behavior and Human Decision Processes* **58**, 136–152.
- Hoelzl, E., Kamleitner, B. and Kirchler, E. (2011), 'Loan repayment plans as sequences of instalments', *Journal of Economic Psychology* **32**, 621–631.
- Hoelzl, E., Pollai, M. and Kamleitner, B. (2009), 'Experience, prediction and recollection of loan burden', *Journal of Economic Psychology* **30**, 446–454.
- Howlett, E., Kees, J. and Kemp, E. (2008), 'The role of self-regulation, future orientation and financial knowledge in long-term financial decisions', *Journal of Consumer Affairs* **42**(2), 223–242.
- Joireman, J., Balliet, D., Sprott, D., Spangenberg, E. and Schultz, J. (2008), 'Consideration of future consequences, ego-depletion, and self-control: Support for distinguishing between cfc-immediate and cfc-future sub-scales', *Personality and Individual Differences* **45**, 15–21.
- Joireman, J., Kees, J. and Sprott, D. (2010), 'Concern with immediate consequences magnifies the impact of compulsive buying tendencies on college students' credit card debt', *Journal of Consumer Affairs* **44**(1), 155–178.
- Joireman, J., Sprott, D. E. and Spangenberg, E. R. (2005), 'Fiscal responsibility and the consideration of future consequences', *Personality and Individual Differences* **39**, 1159–1168.

- Jolls, C., Stunstein, C. R. and Thaler, R. (1998), ‘A behavioral approach to law and economics’, *Stanford Law Review* **50**(5), 1471–1550.
- Kahneman, D. (2003a), ‘Maps of bounded rationality: Psychology for behavioral economics’, *American Economic Review* **93**(5), 1449–1475.
- Kahneman, D. (2003b), ‘A perspective on judgment and choice: Mapping bounded rationality’, *American psychologist* **58**(9), 697–720.
- Kahneman, D. and Tversky, A. (1979), ‘Prospect theory: An analysis of decision under risk’, *Econometrica* **47**(2), 263–291.
- Kahneman, D. and Tversky, A. (1984), ‘Choices, values and frames’, *American Psychologist* **39**(4), 341–350.
- Kamakura, W. A. and Russell, G. J. (1989), ‘A probabilistic choice model for market segmentation and elasticity structure’, *Journal of Marketing Research* **26**(4), 379–390.
- Kamleitner, B., Hoelzl, E. and Kirchler, E. (2010), ‘Experiencing costs and benefits of a loan transaction: The role of cost-benefit association’, *Journal of Economic Psychology* **31**, 1047–1056.
- Kamleitner, B. and Kirchler, E. (2006), ‘Personal loan users’ mental integration of payment and consumption’, *Marketing Letters* **17**, 281–294.
- Kamleitner, B. and Kirchler, E. (2007), ‘Consumer credit use: a process model and literature review’, *Revue europeenne de psychologie appliquee* **57**, 267–283.
- Khandani, A. E., Kim, A. J. and Lo, A. W. (2010), ‘Consumer credit-risk models via machine-learning algorithms’, *Journal of Banking & Finance* **34**, 2767–2787.
- Kivetz, R. (1999), ‘Advances in research on mental accounting and reason-based choice’, *Marketing Letters* **10**(3), 249–266.

- Klapper, L., Lusardi, A. and Panos, G. A. (2013), ‘Financial literacy and its consequences: Evidence from Russia during the financial crisis’, *Journal of Banking and Finance* **37**, 3904–3923.
- Koklu, M. and Sabanci, K. (2016), ‘Estimation of credit card customers payment status by using kNN and MLP’, *International Journal of Intelligent Systems and Applications in Engineering* **4**, 249–251.
- Koyck, L. (1954), *Distributed Lags and Investment Analysis*, North-Holland, Amsterdam.
- Krichler, E., Hoelzl, E. and Kamleitner, B. (2008), ‘Spending and credit use in the private household’, *The Journal of Socio-Economics* **37**, 519–532.
- Kruschke, J. K. (2011), *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*, Academic Press, Burlington, MA.
- Kuhn, M. (2008), ‘Building predictive models in r using the caret package’, *Journal of Statistical Software, Articles* **28**(5), 1–26.
- Laibson, D. (1997), ‘Golden eggs and hyperbolic discounting’, *The Quarterly Journal of Economics* **112**(2), 443–478.
- Langer, T., Sarin, R. and Weber, M. (2005), ‘The retrospective evaluation of payment sequences: duration neglect and peak-and-end effects’, *Journal of Economic Behavior and Organization* **58**, 157–175.
- Latham, G. P. and Seijts, G. H. (1999), ‘The effects of proximal and distal goals on performance on a moderately complex task’, *Journal of Organizational Behavior* **20**, 421–499.
- Lea, S. E., Webley, P. and Walker, C. M. (1995), ‘Psychological factors in consumer debt: Money management, economic socialization, and credit use’, *Journal of Economic Psychology* **16**(4), 681–701.

- Lee, L., Lee, M. P., Bertini, M., Zauberan, G. and Ariely, D. (2015), ‘Money, time and the stability of consumer preferences’, *Journal of Marketing Research* **52**, 184–199.
- Lessmann, S., Baesens, B., Seow, H.-V. and Thomas, L. C. (2015), ‘Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research’, *European Journal of Operations Research* **247**, 124–136.
- Lim, S. S. (2006), ‘Do investors integrate losses and segregate gains? mental accounting and investor trading decisions’, *The Journal of Business* **79**(5), 2539–2573.
- Louzada, F., Ara, A. and Fernandes, G. B. (2016), ‘Classification methods applied to credit scoring: Systematic review and overall comparison’, *Surveys in Operations Research and Management Science* **21**(2), 117–134.
- Lusardi, A. and Mitchell, O. S. (2014), ‘The economic importance of financial literacy: Theory and evidence’, *Journal of Economic Literature* **52**(1), 5–44.
- Malkoc, S. A. and Zauberan, G. (2006), ‘Deferring versus expediting consumption: The effect of outcome concreteness on sensitivity to time horizon’, *Journal of Marketing Research* **43**, 618–627.
- Manchanda, P., Ansari, A. and Gupta, S. (1999), ‘The “shopping basket”: A model for multicategory purchase incidence decisions’, *Marketing Science* **18**(2), 95–114.
- McClure, S. M., Laibson, D. I., Loewenstein, G. and Cohen, J. D. (2004), ‘Separate neural systems value immediate and delayed monetary rewards’, *Science* **306**, 503–507.

- McHugh, S., Ranyard, R. and Lewis, A. (2011), ‘Understanding and knowledge of credit cost and duration: Effects on credit judgements and decisions’, *Journal of Economic Psychology* **32**, 609–620.
- Meier, S. and Sprenger, C. (2010), ‘Present-biased preferences and credit card borrowing’, *American Economic Journal: Applied Economics* **2**(1), 193–210.
- Milkman, K. L. and Beshears, J. (2009), ‘Mental accounting and small windfalls: Evidence from an online grocer’, *Journal of Economic Behavior & Organization* **71**(2), 384–394.
- Min, I. and Kim, J.-H. (2003), ‘Modeling credit card borrowing: A comparison of type I and type II Tobit approaches’, *Southern Economic Journal* **70**(1), 128–143.
- Muraven, M., Collins, R. L. and Nienhaus, K. (2002), ‘Self-control and alcohol restraint: An initial application of the self-control strength model’, *Psychology of Addictive Behaviors* **16**(2), 113–120.
- Neema, S. and Soibam, B. (2017), ‘The comparison of machine learning methods to achieve most cost-effective prediction for credit card default’, *Journal of Management Science and Business Intelligence* **2**(2), 36–41.
- Norges Bank (2018), Kunderetta betalingsformidling 2017, Norges Bank Memo 2, Norges Bank.
- Norvilitis, J. M., Szablicki, P. B. and Wilson, S. D. (2003), ‘Factors influencing levels of credit-card debt in college students’, *Journal of Applied Social Psychology* **33**(5), 935–947.
- Novemsky, N. and Kahneman, D. (2005), ‘The boundaries of loss aversion’, *Journal of Marketing Research* **42**(2), 119–128.

- O'Donoghue, T. and Rabin, M. (1999), 'Doing it now or later', *The American Economic Review* **89**(1), 103–124.
- Okada, E. M. (2001), 'Trade-ins, mental accounting, and product replacement decisions', *Journal of Consumer Research* **27**(4), 433–446.
- Okada, E. M. and Hoch, S. J. (2004), 'Spending time versus spending money', *Journal of Consumer Research* **31**(2), 313–323.
- Oreski, S., Oreski, D. and Oreski, G. (2012), 'Hybrid system with genetic algorithm and artificial neural network and its application to retail credit risk assessment', *Expert Systems with Applications* **39**, 12605–12617.
- Overton, A. A. and MacFadyen, A. J. (1998), 'Time discounting and the estimation of loan duration', *Journal of Economic Psychology* **19**, 607–618.
- Pasha, D. M., Fatima, M., Dogar, A. M. and Shahzad, F. (2017), 'Performance comparison of data mining algorithms for the predictive accuracy of credit card defaulters', *International Journal of Computer Science and Network Security* **17**(3), 178–183.
- Perry, V. G. (2008), 'Giving credit where credit is due: The psychology of credit ratings', *The Journal of Behavioral Finance* **9**, 15–21.
- Perry, V. G. and Morris, M. D. (2005), 'Who is in control? the role of self-perception, knowledge, and income explaining consumer financial behavior', *The Journal of Consumer Affairs* **39**(2), 299–313.
- Pirog, III, S. F. and Roberts, J. A. (2007), 'Personality and credit card misuse among college students: The mediating role of impulsiveness', *Journal of Marketing Theory and Practice* **15**(1), 65–77.

- Platt, J. C. (1999), Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, *in* 'ADVANCES IN LARGE MARGIN CLASSIFIERS', MIT Press, pp. 61–74.
- Prelec, D. and Loewenstein, G. (1998), 'Red and the black: Mental accounting of savings and debt', *Marketing Science* **17**(1), 4–28.
- Prelec, D. and Simester, D. (2001), 'Always leave home without it: A further investigation of the credit-card effect on willingness to buy', *Marketing Letters* **12**(1), 5–12.
- Pundir, S. and Seshadri, R. (2012), 'A novel concept of partial Lorenz curve and partial Gini index', *International Journal of Engineering Science and Innovative Technology* **1**(2), 296–301.
- Raab, G., Elger, C. E., Neuner, M. and Weber, B. (2011), 'A neurological study of compulsive buying behavior', *Journal of Consumer Policy* **34**(4), 401–413.
- Raftery, A. E. and Lewis, S. M. (1992), 'One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo', *Statistical Science* **7**(4), 493–497.
- Raghubir, P. and Srivastava, J. (2008), 'Monopoly money: The effect of payment decoupling and form on spending behavior', *Journal of Experimental Psychology* **14**(3), 213–225.
- Rajagopal, P. and Rha, J.-Y. (2009), 'The mental accounting of time', *Journal of Economic Psychology* **30**(5), 772–781.
- Ranyard, R., Hinkley, L., Williamson, J. and McHugh, S. (2006), 'The role of mental accounting in consumer credit decision process', *Journal of Economic Psychology* **27**, 571–588.

- Redelmeier, D. A. and Kahneman, D. (1996), ‘Patients’ memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures’, *Pain* **66**(1), 3–8.
- Rick, S. I., Cryder, C. E. and Loewenstein, G. (2008), ‘Tightwads and spendthrifts’, *Journal of Consumer Research* **34**, 767–782.
- Rossi, P. E., Allenby, G. M. and McCulloch, R. (2005), *Bayesian Statistics and Marketing*, Wiley, West Sussex, England.
- Rust, R. T., Lemon, K. N. and Zeithaml, V. A. (2004), ‘Return on marketing: Using customer equity to focus marketing strategy’, *Journal of Marketing* **68**, 109–127.
- Sayman, S. and Öncüler, A. (2009), ‘An investigation of time inconsistency’, *Management Science* **55**(3), 470–482.
- Seiler, M. J., Seiler, V. L. and Lane, M. A. (2012), ‘Mental accounting and false reference points in real estate investment decision making’, *Journal of Behavioral Finance* **13**(1), 17–26.
- Shafir, E. and Thaler, R. H. (2006), ‘Invest now, drink later, spend never: On the mental accounting of delayed consumption’, *Journal of Economic Psychology* **27**, 694–712.
- Shefrin, H. M. and Thaler, R. H. (1988), ‘The behavioral life-cycle hypothesis’, *Economic Inquiry* **26**(4), 609–643.
- Shefrin, H. and Nicols, C. M. (2014), ‘Credit card behavior, financial styles, and heuristics’, *Journal of Business Research* **67**(8), 1679–1687.
- Soman, D. (2001a), ‘Effects of payment mechanism on spending behavior: The role of rehearsal and immediacy of payments’, *Journal of Consumer Research* **27**(4), 460–474.

- Soman, D. (2001*b*), ‘The mental accounting of sunk time costs: Why time is not like money’, *Journal of Behavioral Decision Making* **14**, 169–185.
- Soman, D. and Gourville, J. T. (2001), ‘Transaction decoupling: How price bundling affects the decision to consume’, *Journal of Marketing Research* **38**, 30–44.
- Soster, R. L., Monga, A. and Bearden, W. O. (2010), ‘Tracking costs of time and money: How accounting periods affect mental accounting’, *Journal of Consumer Research* **37**(4), 712–721.
- Sriram, S., Chintagunta, P. K. and Agarwal, M. K. (2010), ‘Investigating consumer purchase behavior in related technology product categories’, *Marketing Science* **29**(2), 291–314.
- Tan, A. K., Yen, S. T. and Loke, Y. J. (2011), ‘Credit card holders, convenience users and revolvers: A Tobit model with binary selection and ordinal treatment’, *Journal of Applied Economics* **14**(2), 225–255.
- Thaler, R. (1981), ‘Some empirical evidence on dynamic inconsistency’, *Economics Letters* **8**, 201–207.
- Thaler, R. (2015), ‘Misbehaving: How economics became behavioural’.
- Thaler, R. H. (1980), ‘Toward a positive theory of consumer choice’, *Journal of Economic Behavior and Organization* pp. 39–60.
- Thaler, R. H. (1985), ‘Mental accounting and consumer choice’, *Marketing Science* **4**(3), 199–214.
- Thaler, R. H. (1999), ‘Mental accounting matters’, *Journal of Behavioral Decision Making* **12**, 183–206.
- Thaler, R. H. (2016), ‘Behavioral economics: Past, present and future’, *American Economic Review* **106**(7), 1577–1600.

- Thaler, R. H. (2018), ‘From cashews to nudges: The evolution of behavioral economics’, *American Economic Review* **108**(6), 1265–1287.
- Thaler, R. H. and Johnson, E. T. (1990), ‘Gambling with the house money and trying to break even: The effects of prior outcomes’, *Management Science* **36**(6), 643–660.
- Thaler, R. H. and Shefrin, H. (1981), ‘An economic theory of self-control’, *Journal of Political Economy* **89**(2), 392–406.
- Tjur, T. (2009), ‘Coefficients of determination in logistic regression models—a new proposal: The coefficient of discrimination’, *The American Statistician* **63**(4), 366–372.
- Train, K. E. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York, NY.
- Tversky, A. and Kahneman, D. (1981), ‘The framing of decisions and the psychology of choice’, *Science* **211**(4481), 453–458.
- Tversky, A. and Kahneman, D. (1992), ‘Advances in prospect theory: Cumulative representation of utility’, *Journal of Risk and Uncertainty* **5**, 297–323.
- Ωšušteršič et al.
- Šušteršič, M., Mramor, D. and Zupan, J. (2009), ‘Consumer credit scoring models with limited data’, *Expert Systems with Applications* **36**, 4736–4744.
- Wang, L., Lu, W. and Malhotra, N. K. (2011), ‘Demographics, attitude, personality and credit card features correlate with credit card debt: A view from china’, *Journal of Economic Psychology* **32**(1), 179–193.
- Webley, P. and Plaisier, Z. (1998), ‘Mental accounting in childhood’, *Children’s Social and Economics Education* **3**(2), 55–64.

- Wu, C.-S. P. and Guszczka, J. (2003), Does credit score really explain insurance losses? multivariate analysis from a data mining point of view, *in* ‘Proceedings of the Casualty Actuarial Society’, Casualty Actuarial Society, pp. 113–138.
- Yang, S. and Allenby, G. M. (2000), ‘A model for observation, structural and household heterogeneity in panel data’, *Marketing Letters* **11**(2), 137–149.
- Yang, S., Markoczy, L. and Qi, M. (2007), ‘Unrealistic optimism in consumer credit card adoption’, *Journal of Economic Psychology* **28**, 170–185.
- Yeh, I.-C. and Lien, C. (2009), ‘The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients’, *Expert Systems with Applications* **36**(2), 2473–2480.
- Yu, L., Wang, S. and Lai, K. K. (2008), ‘Credit risk assessment with multistage neural network ensemble learning approach’, *Expert Systems with Applications* **34**, 1434–1444.
- Zauberman, G., Kim, B. K., Malkoc, S. A. and Bettman, J. R. (2009), ‘Discounting time and time discounting: Subjective time perception and intertemporal preferences’, *Journal of Marketing Research* **46**, 543–556.
- Zeelenberg, M. and van Dijk, E. (1997), ‘A reverse sunk cost effect in risky decision making: Sometimes we have too much invested to gamble’, *Journal of Economic Psychology* **18**, 677–691.
- Zhao, Y., Zhao, Y. and Song, I. (2009), ‘Predicting new customers’ risk type in the credit card market’, *Journal of Marketing Research* **46**(4), 506–517.

Appendices

A — Predictive Performance – Resampled Training Data

Table A.1 summarizes predictive performance for variations of training and testing data using the manually resampled training dataset. Data is resampled similarly to the congruent sampled data, where the delinquent months are removed. The month preceding delinquency is indicated as a positive instance, and oversampled manually. Negative instances still outnumber positives, though to a lesser degree than for the congruent sampled training data. Results are generally poorer overall for all algorithms, which is usually the case for manual sampling schemes. Note that this process mimics the routine employed when estimating the proposed model, which is the reason for its conclusion. The proposed model still shows superior performance, suggesting that the results are robust to different sampling schemes.

Table A.1: Comparison of predictive performance, manually resampled training data

	Default-only Test Sample, Unbalanced Training Data								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.677	0.265	0.329	0.679	0.177	0.277	0.607	0.270	0.190
DNN ensemble	0.632	0.236	0.274	0.623	0.259	0.172	0.513	0.251	0.081
GBM	0.530	0.095	0.077	0.526	0.236	0.089	0.557	0.114	-0.053
GBM ensemble	0.496	0.091	0.018	0.489	0.101	-0.027	0.497	0.111	0.040
DNN-GBM ensemble	0.499	<i>0.091</i>	0.025	0.493	0.100	-0.018	0.502	<i>0.111</i>	0.058
XGBoost	0.525	0.092	0.106	0.592	<i>0.100</i>	0.042	0.500	0.112	0.039
Proposed model w/ <i>k</i> NN	0.567	0.160	0.203	<i>0.705</i>	0.177	<i>0.297</i>	<i>0.622</i>	0.216	<i>0.329</i>
	Default-only Test Sample, Balanced Training Data								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.564	0.232	0.137	0.588	0.248	0.131	0.549	0.298	-0.017
DNN ensemble	0.530	0.213	0.082	0.601	0.207	<i>0.204</i>	0.510	0.245	0.104
GBM	0.529	0.187	0.151	0.592	0.199	0.129	0.504	0.225	0.073
GBM ensemble	0.512	0.175	0.122	0.593	0.187	0.158	0.503	0.216	0.071
DNN-GBM ensemble	0.514	<i>0.160</i>	0.101	0.587	0.174	0.147	0.514	0.204	0.101
XGBoost	0.540	0.166	<i>0.157</i>	0.584	<i>0.173</i>	0.070	0.529	<i>0.178</i>	0.140
Proposed model w/ <i>k</i> NN	<i>0.596</i>	0.212	0.153	<i>0.655</i>	0.216	0.155	<i>0.609</i>	0.217	<i>0.323</i>
	Balanced Test Sample								
	1 month			2 months			3 months		
	AUC	BS	PG	AUC	BS	PG	AUC	BS	PG
DNN	0.537	0.195	0.089	0.575	0.202	0.132	0.545	0.217	-0.117
DNN ensemble	0.518	0.163	0.045	0.582	0.168	0.188	0.511	0.187	-0.002
GBM	0.585	0.101	0.096	0.650	0.102	0.226	0.568	0.106	0.067
GBM ensemble	0.543	0.093	0.076	0.646	0.094	0.287	0.563	0.099	0.165
DNN-GBM ensemble	0.528	0.079	0.014	0.643	0.090	0.250	0.539	0.096	0.072
XGBoost	0.591	0.086	0.123	0.629	0.087	0.203	0.583	0.093	0.105
Proposed model w/ <i>k</i> NN	<i>0.651</i>	0.092	<i>0.223</i>	0.724	0.084	0.373	0.715	0.092	0.373

Classification metrics at 1, 2, and 3 months prior to observed delinquency. Training data for balanced test sample is chosen based on classification performance for default-only test sample. Bold font indicates best global classifier. Italic font indicates sample-specific best classifier. Algorithms as specified in section 3.3.2, classification metrics calculated as outlined in section 3.3.3: AUC = Area under the ROC curve, BS = Brier score, PG = partial Gini index. A perfect classifier would have AUC=1, PG=1, and BS=0.

B — Simulation Descriptive Statistics

Figure B.1 contains the history-plots of the two segments estimated for the proposed model, presented in table 4.4. Using a constricted random walk, these parameters enter into the calculation of the lagged independent variables. As previously explained, the model selects lagged variables from the set of two calculated lagged variables. For each iteration, the set of lagged variables that better fits an individual delinquency history given the model parameter estimates are assigned, which prompts the individual to belong to either segment 1 or segment 2. The observed trending and lack of convergence for segment 2 is due to the low impact of the segment 2 lag weights on delinquency prediction.

The B_k parameter MCMC iteration histories are drawn in figure B.2. Presented here with a moderate chain length,¹ the parameter estimates show some trending (which is solved when running a longer chain). As explained in chapter 3, the B_k parameters are estimated using a Gibbs sampler. The history plots reflect the low parameter standard deviations shown in table 4.4; the sampler shows a tendency to vary around the posterior median.² Figure B.3 shows the posterior densities for the B_k parameters.

¹The main estimation featured a burn-in of 400,000, while the total chain length was 800,000. Only half of the iterations were kept; a thinning interval of 2 was used.

²This is apparent when viewing the y-axes in the plots. The history plots display the welcomed “fuzzy caterpillar” appearance, while they vary around the posterior medians.

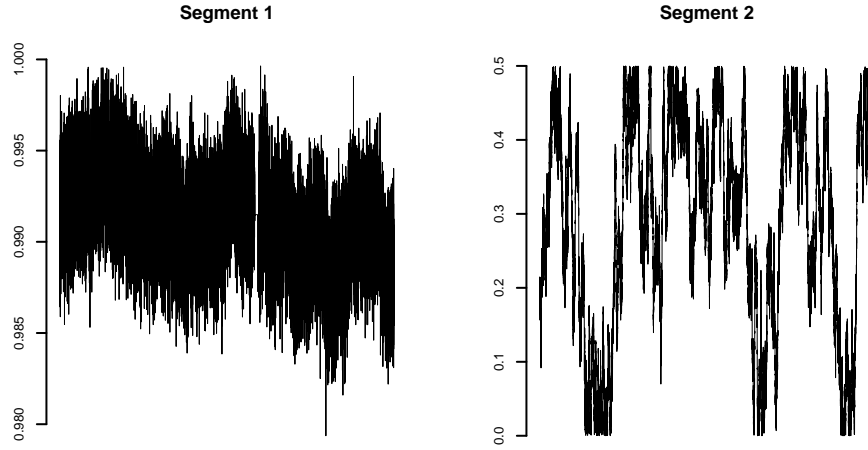


Figure B.1: *History plots for the lag weight parameters for the two segments estimated using debit data.*

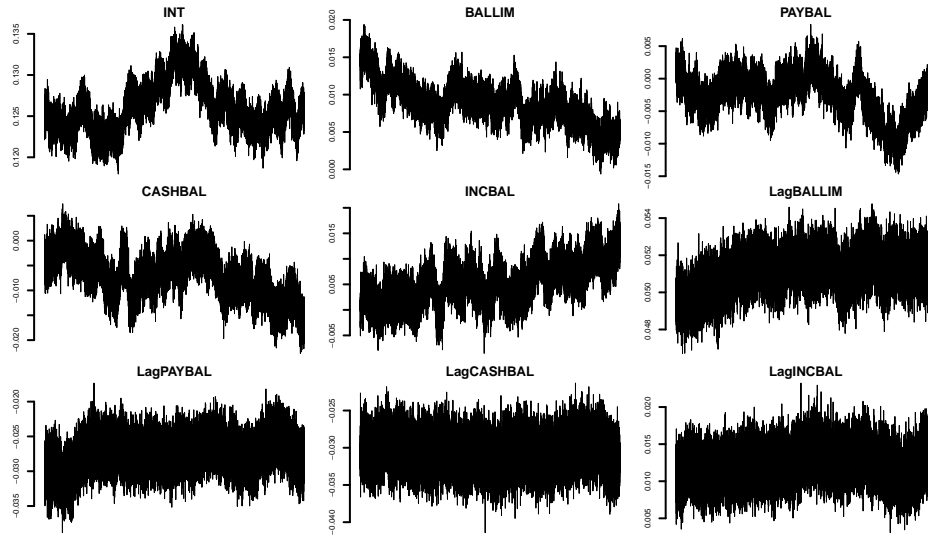


Figure B.2: *History plots for the aggregate B_k parameters using debit and credit data.*

Estimated Posterior Density

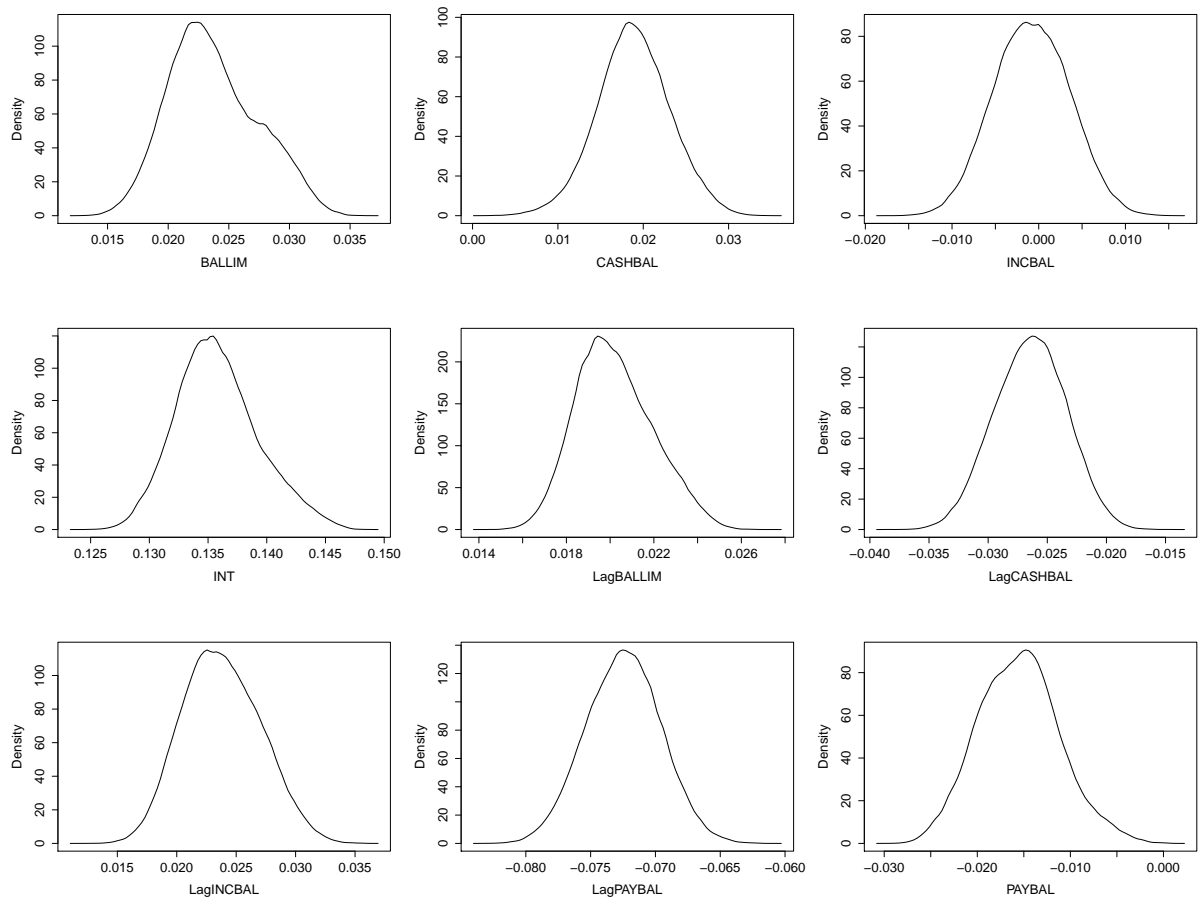


Figure B.3: *Posterior densities for the aggregate B_k parameters.*

C — Convergence Diagnostics and Figures

Table C.1 shows the Raftery and Lewis (1992) convergence diagnostics, which are typically used to demonstrate convergence by number of iterations. Convergence is achieved if the number of iterations in the simulation are higher than what the diagnostic deems necessary. As illustrated, the diagnostic suggests that the parameter estimates have converged. Figure C.1 displays the running means of the sampler. Again, some trending is displayed, which is mitigated by using a sufficient number of iterations.

	Thin	Burn-in	Total	Lower Bound	Dependence Factor
INT	25	225	232775	3746	62.1
BALLIM	29	232	283098	3746	75.6
PAYBAL	20	120	123620	3746	33.0
CASHBAL	24	144	145272	3746	38.8
INCBAL	43	473	557710	3746	148.9
lagBALLIM	9	27	33516	3746	8.9
lagPAYBAL	11	33	47597	3746	12.7
lagCASHBAL	6	18	23808	3746	6.4
lagINCBAL	5	15	19785	3746	5.3

Table C.1: *Raftery and Lewis (1992) test of convergence. If the value presented in the “lower bound” column is higher than the iterations used, convergence has not been achieved. This test was conducted on a run of 800,000 iterations.*

Sampler Running Mean

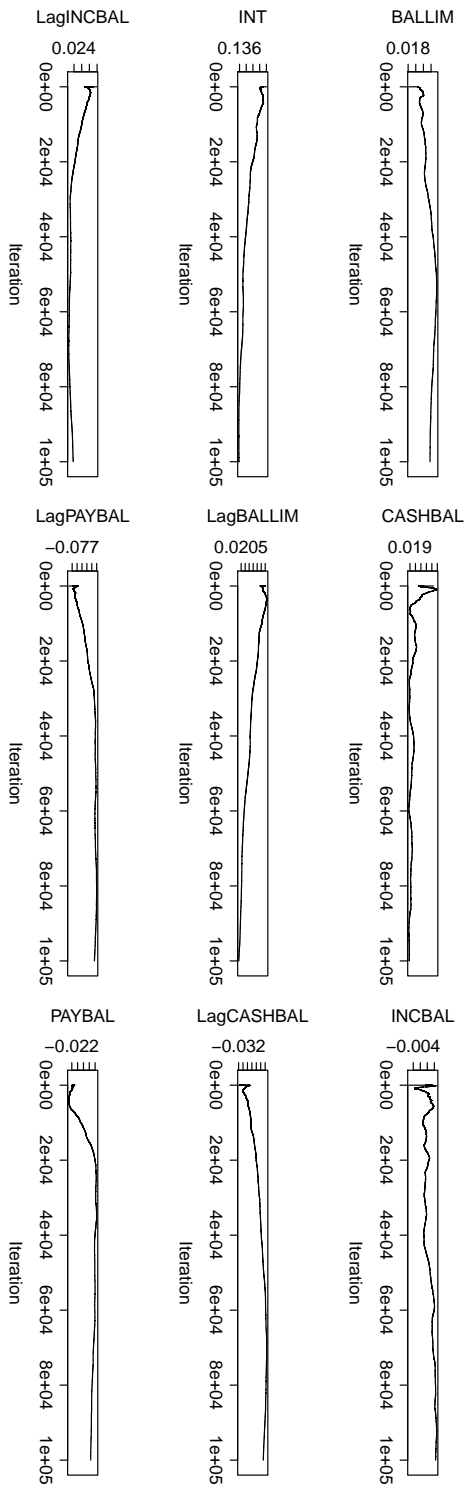


Figure C.1: Running mean of aggregate B_k estimates.