Linking Real-Time Information to Actions: Collectability Scores for Delinquent Credit-Card Accounts

Naveed Chehrazi* Thomas A. Weber†

September 2009

WISE - Extended Abstract

Abstract

In August 2009, the Federal Reserve Bank reports the volume of consumer credit-card debt in the United States to be in excess of $900 billion. According to the 2009 Nilson Report this number is projected to grow by 20% in 2010. Developing an optimal strategy for collecting such an enormous debt is a crucial operational problem that, to the best of our knowledge, has not been successfully studied in either academia or industry. We provide a new approach to the estimation of the repayment probability based on account-specific and macroeconomic information. Our model generates a probability measure of the collectability of an account balance. Unlike the FICO score, which is a relative index of creditworthiness of an individual and does not have any intrinsic meaning, our collectability score corresponds to the actual chance of collecting a given percentage of an account debt over a given time horizon. Furthermore, in contrast with FICO score, our collectability score is specialized for overdue credit-card accounts placed in collection.

*Ph.D. Candidate, Department of Management Science and Engineering, Terman Engineering Center, Stanford University, Stanford, CA 94305-4026. Email: chehrazi@stanford.edu

†Assistant Professor, Department of Management Science and Engineering, Terman Engineering Center, Stanford University, Stanford, CA 94305-4026. Phone: (650) 725-6827. Email: webert@stanford.edu.

1 Introduction

The collection of the outstanding debt from defaulted credit-card accounts (for both charge and lending segments) constitutes a significant portion of credit-card companies' business. As of June 2009 the FRB G.19 consumer credit report (released in August 2009) estimates the size of outstanding revolving consumer credit in the United States at about $917 billion. This credit-card debt is projected to grow by about 20% in 2010 according to the most recent Nilson report. Consequently, developing a sound analytical framework for the collection of this debt is of high economic relevance to the underwriting banks. Even a small relative improvement in the effectiveness of the debt-collection process promises to a significant bottom-line impact. For example, an industry-wide improvement of 1% in the overall collections return (spin) would lead to a $100 million gain. In addition to the increase in the collection yield, a comprehensive mathematical model can help banks to develop a more a realistic assessment of the risk caused by delinquency of an account due to partial recovery of debt. This can be used for improving the underwriting process as well as for designing optimal settlement offers for delinquent accounts.[1]

The problem of credit scoring has been studied, to some extend, in the economics literature. This work emphasizes a description of the lender-borrower interaction in the unsecured consumer credit market to study the welfare implications of various regulatory policies. Chatterjee et al. (2008) analyze an equilibrium model, in which credit scoring is used as an index encapsulating the borrower's reputation and creditworthiness. Ausubel (1999), Cole et al. (1995), and Chatterjee et al. (2007), among others, use game-theoretic models with asymmetric information to qualitatively model the notion of credit scoring as a screening/signaling problem, in an attempt to explain the dynamics of the market of unsecured debt. Credit scoring is also studied in the statistics literature, where different classification models have been used to rank consumers based on their credit worthiness and probability of default. For example, Hand and Henley (1996), Boyes et al. (1989), and West (2000) use different statistical methods, including neural network, discriminant analysis, and decision trees to discriminate between "bad" and "good" borrowers. Yet, none of the existing strands in the literature focus on estimating the actual probability of default for participants in the unsecured consumer debt market (credit-card industry) which besides having intrinsic meaning, can be used as a fundamental index in game-theoretic models for studying market dynamics as well as analyzing regulatory policies. Such an index also provides a rigorous framework for studying important operational issues such as collection-agency compensation, early account settlement, and debt leasing. In this paper, we develop a dynamic stochastic model for estimating (the dual of) such a probability measure for delinquent credit-card accounts: the probability of recovering defaulted debt within a given time period. The method we propose can be generalized to assess the default risk of an active credit-card account before delinquency.

1.1 Applications

The main focus of this study is to estimate a dynamic probability measure for a given delinquent account to pay off its outstanding balance over a given time horizon into the future. The estimation procedure is to take into account both macro-economic data which is common across accounts and account-specific data such as FICO score and responses to various collector actions. The constructed measure will allow to construct a repayment-probability schedule for different time horizons and repayment goals.

This estimated probability measure corresponds to actual repayment probabilities and is therefore more predictive than commonly used scoring methods which often use a weighted average of account and population data at given time instances without much flexibility when varying the repayment threshold (e.g., from 90% repayment to 85% repayment) or the horizon (e.g., from 6 months to 3 months). Such a direct dynamic probability measure can be directly used to inform the decision making about a wide range of operational questions. The following are examples of applications of a probability-based dynamic collectability score.

- **Debt Leasing.** What is the economic value of a delinquent credit-card account with total outstanding balance b as a function of its age in collection?
- **Collection Strategy.** Given the age, outstanding balance, and other attributes of an account, and

[1] See also Chehrazi and Weber (2008), who, relying on actual static settlement response data from a major credit-card company, devise a nonparametric method for simultaneously estimating response curves and optimizing settlement offers.

given a history of past account and segment responses what is the best course of action for collecting a the outstanding debt?

- **Settlement Strategy.** If settlement is an option, when and at what rate should an offer be made? Should such an offer be repeated? If yes, at what rate?
- **Agency Compensation.** What is the optimal (possibly nonlinear) commission rate, payable to the outside collecting agencies, for collecting a dollar of an outstanding debt?
- **Effort Allocation.** What is the economically most efficient way for allocating collection effort among different groups of delinquent accounts?
- **Risk Hedging.** How should the creditor hedge against the risk associated with partial collectability of credit-card debt?
- **Screening and Underwriting.** Given the risk of delinquency, what are the optimal product parameters, such as credit limit, annual percentage rate, etc., which limit balance the default risk against the expected revenue, and, at the same time, assure the bank's competitiveness.

We develop a stochastic model which exploits both account-specific and macroeconomic information in estimating aforementioned probability measure.

2 Methodology

2.1 Conventional Models

The structure of the collection process differs from bank to bank, but typically consists of different levels and different treatments depending on the age and the type of an account. For example there can be task forces specialized on business or household accounts and an account can moves from one task force to another as the collectors fail to collect a minimum expected amount. A traditional way of analyzing and making decisions in this setting, is usually based on a simple regression model where an analyst runs a regression or time series analysis over the historical data of a group of accounts with similar attributes in order to determine the parameters of the model and uses them to assess the collectability of an account. Although, this method benefits from simplicity and depth of the theory of regression and data-mining, it doesn't impose any probabilistic structure on the problem and does not utilize the availability of any macroeconomic information. The structural properties of a problem usually contain valuable information that is not included in the data itself. To be more specific, for example, it may be reasonable to assume that if a delinquent account makes a partial repayment, it is more probable that it will pay off its balance in the future. This can be justified by interpreting the payment as an action that signals the seriousness of the account holder for repaying his debt. Apart from the information embedded in the action, the actual payment can also be used to infer the account holder's financial ability. This type of (game-theoretic) information cannot be exploited by any conventional econometrics models. The macroeconomic indicators also convey a valuable information about the financial ability of the account holders. Indicators such as unemployment rate, consumer price index, interest rate, S&P 500 and other stock market indices are correlated with the overall collection performance. Sudden changes in the economy has direct impact on consumer credit industry which cannot be captured by conventional regression models.

A more sophisticated approach for modelling the collection process can be based on Markov chains. In this type of model, accounts are considered to move from one state to another state according to some transition matrix. Each state is characterized by some attributes of the account holder such as his FICO score, number of delinquent and active credit cards owned by him, etc. as well as the account's repayment history. The state transition is initiated either based on an account's age or as a result of an action performed by a collector. The state transition matrix in this framework is derived based on historical data for a specific period of account's life within the collection process and is time invariant. This transition matrix can then be used for understanding the dynamics of the collection process and to evaluate its overall performance. Despite the fact that this type of model assumes some probabilistic structure about the process, it is a static model and suffers from the aforementioned deficiencies. The transition matrix derived from historical data may not be a good estimation for today's transition matrix, and it cannot be calculated for any given period of an account's life within collection. Furthermore, it cannot be used to make any prediction about the value of a defaulted account or to make any account-level operation decisions.

2.2 Model Outline

At any point of the collection process, a delinquent account's "collectability" corresponds to the probability of paying back a certain percentage of its balance. We consider *self-exciting point processes* as an appropriate framework for describing the probabilistic account behavior over time in the collection process. As mentioned in the previous section, a dynamic probability measure constructed from a doubly stochastic point process that describes an accounts behavior can be used to directly address a variety of operational questions.

In order to understand the idea for our model, consider a given delinquent credit-card account with an outstanding balance b placed in the collection process at time $t = 0$. The account is characterized (in addition to its balance) by a set of attributes, including its FICO score, credit limit, and information about the account holder's financial credentials (e.g., the number of his credit-card accounts and their respective limits, and his mortgage status). A portion of this account-specific information does not change during the collection process and can be treated as constant. Macroeconomic information correlated with the financial credentials of the account holder, such as the S&P 500, consumer confidence index, unemployment rate, interest rate, GDP, CPI, together with time-varying account-specific information will be modelled by a continuous time (Markov) stochastic process. The collector actions and account holder reactions can also be included in the account-specific information. They represent deterministic signals to which the model's sensitivity will indicate how the corresponding actions can be expected to influence the return from the collection effort.

During the collection process, the account holder will make a set of random payments Z_i at random times τ_i, $i \in \{1, 2, \ldots\}$, until he pays off his balance. Under some mild technical assumptions, which are always satisfied in the practical context of the collection problem, the set of payment times $\tau_i > 0$, $i \in \{1, 2, \ldots\}$, will form a point process which is *uniquely* defined by a nonnegative intensity process $\lambda(t)$. The intensity process $\lambda(t)$ can be interpreted as conditional arrival rate of repayment events. At any time $t \geq 0$, we are interested in the probability that the total repayments of the account is larger than or equal to $\bar{b}$ by time $T > t$, i.e.,

$$P\left\{ \sum_{i=1}^{N(T)} Z_i > \bar{b} \,\middle|\, \mathscr{F}_t \right\},$$

where $N(T)$ is a 'counting process' defined by the point process associated with the stopping times τ_i, $i \in \{1, 2, \ldots\}$, and the filtration $\mathscr{F}_t$ denotes all available information up to time t. In this model, $\lambda(t)$ is determined by an affine function of the account-specific and the macroeconomic covariates, each of which is modelled by an affine (jump-)diffusion process. It is important to note that the diffusion intensity itself is represented by a stochastic process, which in turn accounts for the intrinsic non-observability of the process $\lambda(t)$, and the error in predicting the available covariate processes. At each point in time, the trend and the level of the intensity process is affected by the available information in the form of history and the predicted trend of the covariates. In what follows we outline one possible method of incorporating the information provided by account covariates into the prediction of the repayment intensity process.

2.2.1 The Level of the Stochastic Intensity

In our model, the level of the intensity process is determined by a linear combination of the covariates. Among the available repayment covariates, the payment process (Z_i, τ_i), $i \in \{1, 2, \ldots\}$, and the collector's actions play a prominent role in determining the level of the stochastic repayment intensity. With the payment process as one of the available covariates, it is possible to implement the action signalling structure described in the previous section. For example, a recent partial debt repayment tends to increase the chances of recovering the remaining balance within a given time horizon. This relationship is modelled by introducing a jump in the intensity at the time of payment τ_i, the size of which is correlated with the actual payment Z_i. The dependency of the intensity of a point process on its own path is termed "self-exciting" in the literature and was first studied by Hawkes (1971). The effects of the collector's actions can also be explicitly incorporated, by allowing $\lambda(t)$ to have a discontinuity at the time when the action is undertaken. The level of the corresponding jump is determined endogenously and can depend on the level of other macroeconomic indicators. Figure 1 illustrates how the level of the intensity process depends on the payment history and collector's actions in a prototype setting.

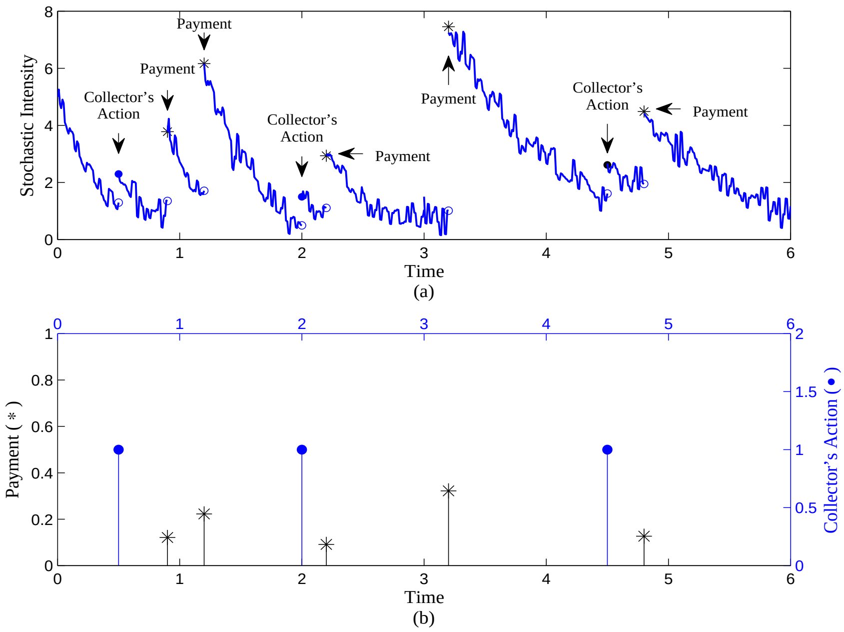


Figure 1: Dependence of the stochastic intensity of the repayment arrival times, $\lambda(t)$, on the repayment history and the collector's actions. Panel (a) shows a realization of $\lambda(t)$ as an affine function of the covariates. Panel (b) depicts the repayment history of the account as well as the timing of collection actions.

Self-exciting point processes have recently been the center of attention for researchers in credit-risk and have been in industry for the credit-rating of corporate defaultable bonds. In a typical credit-risk setup, one is interested in pricing a defaultable corporate bond which makes a fixed payment at a given time in the future, given no default. In case of default, the bond issuer will make a predetermined partial repayment at the time of default and closes the contract. This problem of pricing a defaultable bond has a number of common features with out problem of constructing a repayment collectability score, since both consist exhibit uncertain cash flows. However, the problems also have significant differences, because the credit-card accounts considered in our model have already defaulted and there are many payments which are random both in time and in the amount. Yet, our problem can in some sense be considered as the dual of certain credit-risk problems and the available theory and empirical methods developed for them can be extended and tailored to meet the specification of ours. Among many papers available on this topic, we highlight the work by Duffie et al. (2000, 2007), and Carr and Wu (2004), which are closely related to our problem.

2.2.2 The Trend of the Stochastic Intensity

Although the trend of all covariates affect the trend of the intensity process, more emphasis will be put on macroeconomic covariates and some account specific indicators such as FICO score. During the lifetime of an account in the collection process, the more the account is inactive in terms of responding to the actions undertaken by collectors and paying back the debt, the less is the chance of recovering the outstanding balance. This relation can be modelled by enforcing the intensity process to decrease as a function of time during such an still period. This trend characteristic can be made to depend on the level of macroeconomic variables and FICO specifically. In a booming economy, one can expect the intensity process to decrease more slowly when the account is inactive, exhibiting a more lasting memory for events that increase the intensity, compared to an economic downturn. Similarly, as the credit rating of an individual decrease, one

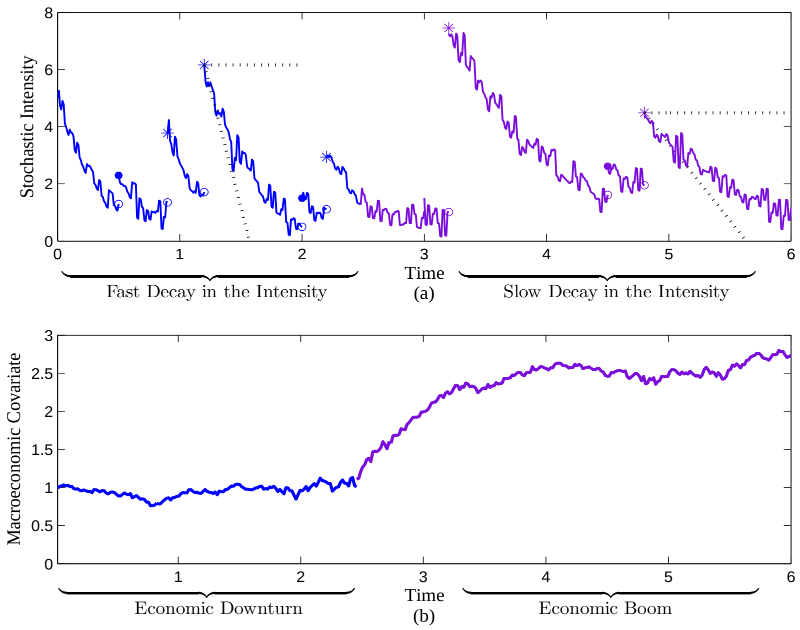


Figure 2: Dependence of the stochastic intensity of repayment arrival times, $\lambda(t)$, on the state of the economy. Panel (a) shows a realization of $\lambda(t)$ as affine function of the covariates. Panel (b) displays a sample path of the principal component of a set of macroeconomic covariates.

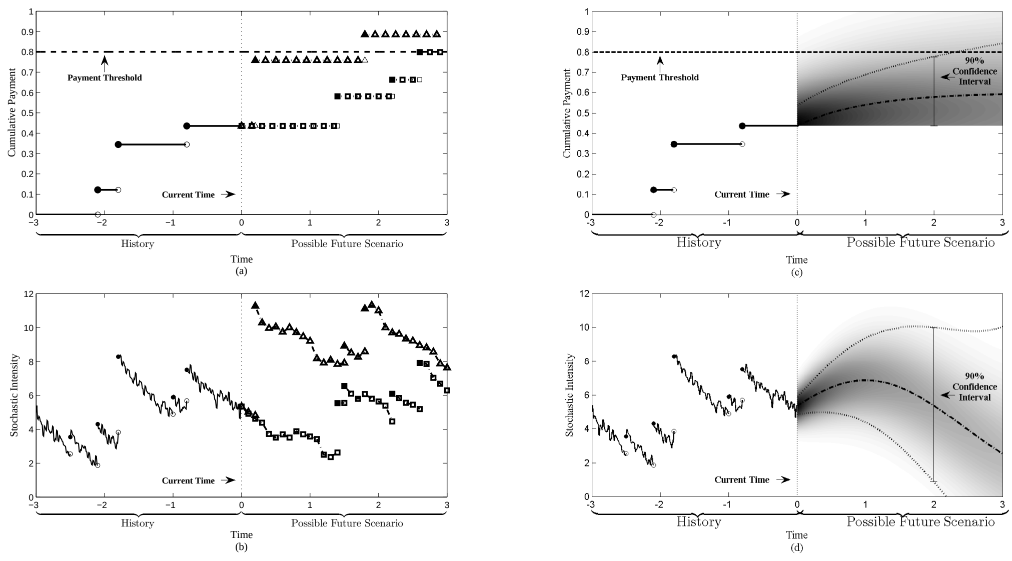
may focus on the short-term history of the account holder. Figure 2 depicts this dependency in the same prototype model as in the previous subsection.

2.2.3 Latent Variables

Latent variables, corresponding to unobserved components of the model, can also be included in our framework. These variables capture the part of the repayment pattern that is not explained by observable covariates. Latent variables are widely used in the finance literature to model and predict short-term interest rates or other economic variables of interest. Among the papers in this area, Ang and Piazzesi (2003) is closest to our topic. Finally, it is noteworthy that the constant covariates in our model affect the steady-state behavior of the intensity process. In this context, Figure 3 shows two sample paths of the stochastic intensity $\lambda(t)$ and the account holder's cumulative repayment, $\sum_{k=1}^{i} Z_k, \tau_i$, along with their history. It shows how the self-exciting characteristics of the repayment process influence the intensity, which in turn also affects the future repayment pattern.

3 Conclusion

The model in this paper opens the door for many real-time dynamic applications where an optimal action schedule needs to be developed in the presence of substantial uncertainty. The prediction method developed here combines individual (account-specific) and general (macroeconomic) information, and can therefore be interpreted as a continuous-time limit of panel data analysis, where the underlying time-series data (here cumulative payments) is monotonic, and the decision maker is interested in reaching a certain threshold. Thus, the model can also be applied to all kinds of knowledge accumulation as well as innovation processes, which exhibit the same type of monotonicity.



(a) (b)

Figure 3: Sample paths of the stochastic intensity $\lambda(t)$ and the account holder's cumulative repayment, $\sum_{k=1}^{i} Z_k, \tau_i$.

References

[1] Ang, A., Piazzesi, M. (2003) "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, Vol. 50, pp. 745–787.

[2] Ausubel, L. (1999) "Adverse Selection in the Credit Card Market," Working Paper, University of Maryland, College Park, MD.

[3] Boyes, W., Hoffman, D., Low, S. (1989) "An Econometric Analysis of the Bank Credit Scoring Problem," *Journal of Econometrics*, Vol. 40, pp. 3–14.

[4] Brémaud P. (1981) *Point Processes and Queues, Martingale Dynamics*, Springer, New York, NY.

[5] Carr, P., Wu, L. (2002) "Time-Changed Lévy Processes and Option Pricing," *Journal of Financial Economics*, Vol. 71, pp. 113–141.

[6] Chatterjee, S., Corbae, D., Ríos-Rull, J. (2007) "A Quantitative Theory of Unsecured Consumer Credit with Risk of Default," *Econometrica,* Vol. 75, No. 6, pp. 1525–1589.

[7] Chatterjee, S., Corbae, D., Ríos-Rull, J. (2008) "A Theory of Credit Scorring and Competitive Pricing of Default Risk," Working Paper, Federal Reserve Bank of Philadelphia, Philadelphia, PA.

[8] Chehrazi, N., Weber, T.A. (2008) "Monotone Approximation of Decision Problems," Working Paper, Department of Management Science and Engineering, Stanford University, Stanford, CA.

[9] Cole, H., Dow, J., English, W. (1995) "Default, Settlement, and Signalling: Lending Resumption in a Reputation Model of Sovereign Debt," *International Economic Review,* Vol. 36, No. 2, pp. 365–385.

[10] Cox, D.R., Isham, V. (1980) *Point Processes*, Chapman and Hall, Cambridge, UK.

[11] Dai, Q., Singleton, K.J. (2000) "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, Vol. 55, No. 5, pp. 1943–1978.

[12] Dellacherie, C., Meyer, P.-A. (1982) *Probabilities and Potential B: Theory of Martingales*, Elsevier Science, Amsterdam, NL.

[13] Duffie, D., Kan, R. (1996) "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, Vol. 6, pp. 379–406.

[14] Duffie, D., Pan, J., Singelton, K. (2000) "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, Vol. 68, No. 6, pp. 1343–1376.

[15] Duffie, D., Saita, L., Wang, K. (2007) "Multi-Period Corporate Default Prediction with Stochastic Covariates," *Journal of Financial Economics*, Vol. 83, pp. 635–665.

[16] Giesecke, K. (2009) "An Overview of Credit Derivatives," Working Paper, Department of Management Science and Engineering, Stanford University, Stanford, CA.

[17] Hand, D., Henley, W. (1997) "Statistical Classification Methods in Consumer Credit Scoring: A Review," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 160, No. 3, pp. 523–541.

[18] Hawkes A.G. (1971) "Expectral of Some Self-Exciting and Mutually Exciting Point Processes," *Biometrica*, Vol. 58, No. 1, pp. 83–90.

[19] Karatzas, I., Shreve, S.E. (1988) *Brownian Motion and Stochastic Calculus*, Springer, New York, NY.

[20] Protter, P.E. (2005) *Stochastic Integration and Differential Equations (Second Edition)*, Springer, New York, NY.

[21] West D. (2000) "Neural Network Credit Scoring Models," *Computers and Operations Research*, Vol. 27, pp. 1131–1152.