Risk and Risk Management in the Credit Card Industry*

Florentin Butaru[1], Qingqing Chen[1], Brian Clark[1,4],
Sanmay Das[2], Andrew W. Lo[3], Akhtar Siddique[1]

This Revision: 14 June 2015

Abstract

Using account level credit-card data from six major commercial banks from January 2009 to December 2013, we apply machine-learning techniques to combined consumer-tradeline, credit-bureau, and macroeconomic variables to predict delinquency. In addition to providing accurate measures of loss probabilities and credit risk, our models can also be used to analyze and compare risk management practices and the drivers of delinquency across the banks. We find substantial heterogeneity in risk factors, sensitivities, and predictability of delinquency across banks, implying that no single model applies to all six institutions. We measure the efficacy of a bank's risk-management process by the percentage of delinquent accounts that a bank manages effectively, and find that efficacy also varies widely across institutions. These results suggest the need for a more customized approached to the supervision and regulation of financial institutions, in which capital ratios, loss reserves, and other parameters are specified individually for each institution according to its credit-risk model exposures and forecasts.

* We thank Michael Carhill, Jayna Cummings, Misha Dobrolioubov , Dennis Glennon, Amir Khandani, Adlar Kim, Mark Levonian, David Nebhut, Til Schuerman, Michael Sullivan and seminar participants at the Consortium for Systemic Risk Analysis, the Consumer Finance Protection Bureau, the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), the Office of the Comptroller of the Currency, and the Philadelphia Fed's Risk Quantification Forum for useful comments and discussion. The views and opinions expressed in this article are those of the authors only, and do not necessarily represent the views and opinions of any institution or agency, any of their affiliates or employees, or any of the individuals acknowledged above. Research support from the MIT CSAIL Big Data program, the MIT Laboratory for Financial Engineering, and the Office of the Comptroller of the Currency is gratefully acknowledged.

[1] U.S. Department of the Treasury, Office of the Comptroller of the Currency, Enterprise Risk Analysis Division.

[2] Washington University in St. Louis, Department of Computer Science & Engineering.

[3] Massachusetts Institute of Technology, Sloan School of Management, Computer Science and Artificial Intelligence Laboratory, Electrical Engineering and Computer Science; AlphaSimplex Group, LLC.

[4] Rensselaer Polytechnic Institute (RPI), Lally School of Management.

Table of Contents

I. Introduction

The financial crisis of 2007–2009 highlighted the importance of risk management at financial institutions. Particular attention has been given, both in the popular press and the academic literature, to the risk management practices and policies at the mega-sized banks at the center of the crisis. Few dispute that risk management at these institutions—or the lack thereof—played a central role in shaping the subsequent economic downturn. Despite the recent focus, however, the risk management policies of individual institutions largely remain black boxes.

In this paper, we examine the practice of risk management and its implications of six major U.S. financial institutions using computationally intensive “machine-learning” techniques applied to an unprecedentedly large sample of account-level credit-card data. The consumer-credit market is central to understanding risk management at large institutions for two reasons. First, consumer credit in the United States has grown explosively over the past three decades, totaling $3.3 trillion at the end of 2014. From the early 1980s to the Great Recession, U.S. household debt as a percentage of disposable personal income doubled, although declining interest rates have meant that the debt service ratios have grown at a lower rate. Second, algorithmic decision-making tools, including the use of scorecards based on "hard" information, have, have become increasingly common in consumer lending (Thomas, 2000). Given the larger amount of data as well as the larger number of decisions compared to commercial credit lending, this reliance on algorithmic decision-making should not be surprising. However, the implications of these tools for risk management, for individual financial institutions and their investors, and for the economy as a whole, are still unclear.

Compared to other retail loans such as mortgages, lenders and investors have more options to actively monitor and manage credit-card accounts because they are revolving credit lines. Consequently, managing credit-card portfolios is a potential source of significant value. Better risk management could provide financial institutions with savings on the order of hundreds of millions of dollars annually. For example, lenders can cut or freeze credit lines on accounts that are likely to go into default, thereby reducing their exposure. By doing so, they can potentially avoid an increase in the balances of accounts destined to default, known in the industry as "run-up." However, by cutting these credit lines to reduce run-up, banks also run the risk of cutting the credit limits of accounts that will not default, thereby alienating customers and potentially forgoing profitable lending opportunities. More accurate forecasts of delinquencies and defaults reduce the likelihood of such false positives. Issuers and investors of securitized credit-card debt would also benefit from such forecasts and tools. And given the size of this part of the industry—$861 billion of revolving credit outstanding at the end of 2014—more accurate forecasts can also improve macroprudential policy decisions and reduce the likelihood of a systemic shock to the financial system.

Our data allow us to observe the actual risk management actions undertaken by each bank on an account level, and thus determine the possible cost savings from a given risk management strategy. For example, we can observe line decreases and realized run-ups over time, and the cross-sectional nature of our data allows us to further compare risk-management practices across institutions and examine how actively and effectively firms manage the exposure of their credit-card portfolios. We find significant heterogeneity in the credit-line management actions across our sample of six institutions.

We compare the efficacy of an institution's risk-management process using a simple measure: the ratio of the percentage of credit-line decreases on accounts that become delinquent over a forecast horizon to the percentage of line decreases on all accounts over the same period. This measures the extent to which institutions are targeting "bad" accounts and managing their exposure prior to default.[1] We find that this ratio ranges from less than one, implying that the bank was more likely to cut the lines of good accounts than those that eventually went into default, to over 13, implying the bank was highly accurate in targeting bad accounts. While these ratios vary over time, the cross-sectional ranking of the institutions remains relatively constant, suggesting that certain firms are either better at forecasting delinquent accounts or view line cuts as a beneficial risk-management tool.

Because effective implementation of the above risk-management strategies requires banks to be able to identify accounts that are likely to default, we build predictive models to classify accounts as good or bad. The dependent variable is an indicator variable equal to 1 if an account becomes 90 days past due (delinquent) over the next two, three, or four quarters. Independent variables include individual-account characteristics such as the current balance, utilization rate, and purchase volume; individual-borrower characteristics from a large credit bureau such as the number of accounts an individual has outstanding, the number of other accounts that are delinquent, and the credit score; and macroeconomic variables including home prices, income, and unemployment statistics. In all, we construct 87 distinct variables.

[1] Despite the unintentionally pejorative nature of this terminology, we adopt the industry convention in referring to accounts that default or become delinquent as "bad" and those that remain current as "good".

Using these variables, we compare three modeling techniques—logistic regression, decision trees using the C4.5 algorithm, and random forest. The models are all tested out of sample as if they were being implemented at that point in time, i.e., no future data were used as inputs in these tests. All models perform reasonably well, but the decision trees tend to perform the best in terms of classification rates. In particular, we compare the models based on well-known measures such as precision and recall, and statistics that combine them such as the F-Measure and kappa statistics.[2] We find that the decision trees and random-forest models outperform logistic regression with respect to both measures.

There is, however, a great deal of cross-sectional and temporal heterogeneity. As expected, the performance of all models declines as the forecast horizon increases. However, the performance of the models for each bank remains relatively stable over time (we test the models semi-annually starting in 2010Q4 through the end of our sample period 2013Q4). Across banks we find a great deal of heterogeneity in classification accuracy. For example, at the two-quarter forecast horizon, the mean F-Measure ranges from 63.8% at the worst performing bank to 81.6% at the best.[3] Kappa statistics show similar variability.

We also estimate the potential dollar savings from active risk management using these machine-learning models. The basic strategy is to first classify accounts as good or bad using the above models, and then cut the credit lines of the bad accounts. The cost savings depend on 1) the model accuracy and 2) how aggressively banks cut credit lines.

[2] Precision is defined as the proportion of positives identified by a technique that are truly positive. Recall is the proportion of positives that is correctly identified. The F-Measure is defined as the harmonic mean of precision and recall, and is meant to describe the balance between precision and recall. The kappa statistic measures performance relative to random classification. See Figure 2 for further details.

[3] These F-Measures represent the mean F-Measure for a given bank over time.

The potential cost of this strategy is cutting credit lines of good accounts, thereby alienating customers and losing future revenues. We follow Khandani, et al.'s (2010) methodology to estimate the value added of our models and report the cost savings for various degrees of line cuts (ranging from doing nothing to cutting the account limit to the current balance). To include the cost of alienating customers, we conservatively assume that customers incorrectly classified as bad will pay off their current balances and close their accounts. Therefore, the bank will lose out on all future revenues from such customers.

With respect to this measure, we find that our models all perform well. Assuming that cutting the lines of bad accounts would save a run-up of 30% of the current balance, we find that implementing our decision tree models would save about 55% relative to taking no action for the two-quarter-horizon forecasts. When we extend the forecast horizon, the models do not perform as well and the cost savings decline to about 25% and 22% at the three- and four-quarter horizons, respectively. These figures vary considerably across banks. The bank with the greatest cost savings had a value-added of 76%, 46%, and 35% across the forecast horizons; the bank with the smallest cost savings would only stand to gain 47%, 14%, and 9% by implementing our models across the three horizons. Of course, there are many other aspects of a bank's overall risk management program, so the quality of risk management strategy of these banks cannot be ranked solely on the basis of these results, but the results do suggest that there is substantial heterogeneity in the risk management tools and effective strategies available to banks.

The remainder of the paper is organized as follows. In Section II, we describe our dataset and discuss the security issues surrounding it and the sample-selection process used. In Section III we outline the model specifications and our approach to constructing useful variables that serve as inputs to the algorithms we employ. We also describe the machine-learning framework for creating more powerful forecast models for individual banks, and present our empirical results. We apply these results to analyze bank risk management and key risk drivers across banks in Section IV. We conclude in Section V.

II. The Data

A major U.S. financial regulator has engaged in a large-scale project to collect detailed credit-card data from several large U.S. financial institutions. As detailed below, the data contains internal account-level data from the banks merged with consumer data from a large U.S. credit bureau, comprising over 500 million records over a period of six years. It is a unique dataset that combines the detailed data available to individual banks with the benefits of cross-sectional comparisons across banks.

The underlying data contained in this dataset is confidential, and therefore has strict terms and conditions surrounding the usage and dissemination of results to ensure the privacy of the individuals and the institutions involved in the study. A third-party vendor is contracted to act as the intermediary between the reporting financial institutions, the credit bureau, and the regulatory agency and end users at the regulatory agency are not able to identify any individual consumers from the data. We are also prohibited from presenting results that would allow the identification of the banks from which the data are collected.

A. Unit of Analysis

The credit-card dataset is aggregated from two subsets we refer to as account-level and credit-bureau data. The account-level data is collected from six large U.S. financial institutions. It contains account-level variables for each individual credit-card account on the institutions' books, and is reported monthly starting January 2008. The credit-bureau data is obtained from a major credit bureau, and contains information on individual consumers reported quarterly starting the first quarter of 2009.

This process results in a merged dataset containing 186 raw data items (106 account-level items and 80 credit-bureau items). The account-level data includes items such as month-ending balance, credit limit, borrower income, borrower credit score, payment amount, account activity, delinquency, etc. The credit-bureau data includes consumer-level variables such as total credit limit, total outstanding balance on all cards, number of delinquent accounts, etc.[4]

We then augment the credit-card data with macroeconomic variables at the county and state level using data from the Bureau of Labor Statistics (BLS) and Home Price Index (HPI) data from the Federal Housing Finance Agency (FHFA). The BLS data are at the county level, taken from the State and Metro Area Employment, Earnings, and Hours (SM) series and the Local Area Unemployment (LA) series, each of which is collected under the Current Employment Statistics program. The HPI data are at the state level. The BLS data are matched using ZIP codes.

Given the confidentiality restrictions of the data, the unit of analysis in our models is the individual account. Although the data has individual account-level and credit-bureau information, we cannot link multiple accounts to a single consumer. That is, we cannot determine if two individual credit-card accounts belong to the same individual. However, the credit-bureau data does allow us to determine the total number of accounts that the

[4] The credit-bureau data for individuals is often referred to as attributes in the credit-risk literature.

owner of each of the individual accounts has outstanding. Similarly, we cannot determine unique credit-bureau records, and thus have multiple records for some individuals. For example, if individual A has five open credit cards from two financial institutions, we are not able to trace those accounts back to individual A. However, for each of the five account-level records, we would know from the credit-bureau data that the owner of each of the accounts has a total of five open credit-card accounts.

B. Sample Selection

The data collection by the financial regulator for supervisory purposes started in January 2008. For regulatory reasons, the banks from which the data have come have changed over time though the total number has stayed at eight or less. However, the collection has always covered the bulk of the credit-card market. Mergers and acquisitions have also altered the population over this period.

Our final sample consists of six financial institutions, chosen because they have reliable data spanning our sample period. Although data collection commenced in January 2008, our sample starts in 2009Q1 to coincide with the start of the credit-bureau data collection. Our sample period runs through the end of 2013.[5]

We are forced to draw a randomized subsample from the entire population of data because of the very large size of the data. For the largest banks in our dataset, we sample 2.5% of the raw data. However, as there is substantial heterogeneity in the size of the credit-card portfolios across the institutions, we sample 10%, 20%, and 40% from the

[5] We also drew samples at December 2011, and December 2012. Our results using those samples are quite similar. When we test the models, our out of time test sample extends to 2014Q2 for our measure of delinquency.

smallest three banks in our sample. The reason is simply to render the sample sizes comparable across banks so that differences in the amount of data available for the machine-learning algorithms are not driving the results.

These subsamples are selected using a simple random sampling method. Starting with the January 2008 data, each of the credit-card accounts is given an 18-digit unique identifier based on the encrypted account number. The identifiers are simple sequences starting at some constant and increasing by one for each account. The individual accounts retain their identifiers and can therefore be tracked over time. As new accounts are added to the sample in subsequent periods, they are assigned unique identifiers that increase by one for each account.[6] As accounts are charged off, sold, or closed, they simply drop out of the sample and the unique identifier is permanently retired. We therefore have a panel dataset that tracks individual accounts through time (a necessary condition for predicting delinquency) and also reflects changes in the financial institutions' portfolios over time.

Once the account-level sample is established, we merge it with the credit-bureau data. This process also requires care because the reporting frequency and historical coverage differ between the two datasets. In particular, the account-level data is reported monthly beginning in January 2008, while the credit-bureau data is reported quarterly beginning in the first quarter of 2009. We merge the data using the link file provided by the vendor at the monthly level to retain the granularity of the account-level data. Because we merge the quarterly credit-bureau data with the monthly account-level data, each credit-

[6] For example, if a bank reported 100 credit-card accounts in January 2008, the unique identifiers would be {C+1,C+2,…,C+100}. If the bank then added 20 more accounts in February 2008, the unique identifiers of these new accounts would be {C+101,C+102,…,C+120}.

bureau observation is repeated three times in the merged sample. However, we retain only the quarter-ending months for our models in this paper.

Finally, we merge the macroeconomic variables to our sample using the five-digit ZIP code associated with each account. While we do not have a long time series in our sample, there is a significant amount of cross-sectional heterogeneity that we use to identify macro trends. For example, HPI is available at the state level, and several employment and wage variables are available at the county level. Most of the macro variables are reported quarterly, which allows us to capture short-term trends.

The final merged dataset retains roughly 70% of the credit-card accounts. From here, we only retain personal credit cards. The size of the sample across all banks increases steadily over time from about 5.7 million credit-card accounts in 2009Q4 to about 6.6 million in 2013Q4.

III. Empirical Design and Models

We consider three basic types of credit-card delinquency models: C4.5 decision tree models, logistic regression, and random-forest models. In addition to running a series of "horse races" between these models, we seek a better understanding of the conditions under which each type of model may be more useful. In particular, we are interested in how the models compare over different time horizons, changing economic conditions, and across banks.

We use the open-source software package Weka to run our machine-learning models.[7] Weka offers a wide collection of open-source machine-learning algorithms for data mining. We use Weka's J48 classifier, which implements the C4.5 algorithm developed by Ross Quinlan (1993) (see, Frank, Hall, and Witten (2011)), because of its combination of speed, performance, and interpretability. This algorithm is a decision tree learner. We compare the results with those obtained using logistic regression models and random forests, also available in Weka, and include the same variables as in the decision trees. More specifically, we use a logistic regression model with a quadratic penalty function, i.e. a ridge logistic regression. This is the Weka implementation of logistic regression as per Cessie and van Houwelingen (1992). The likelihood is expressed as the following logistic function:

$$l(\beta) = \sum_i \left[Y_i \log p(X_i) + (1 - Y_i) \log\left(1 - p(X_i)\right)\right] \quad , \quad p(X_i) \stackrel{\text{def}}{=} \frac{\exp(X_i\beta)}{1 \text{ - } \exp(X_i\beta)}$$

The objective function is $l(\beta) - \lambda\|\beta\|^2$ where λ is the ridge parameter. The objective function is minimized with a quasi-Newton method.

Our third model is the random forest. Random forests learn an ensemble of decision trees, combine bootstrap aggregation with random feature selection (Breiman (2001); Breiman and Cutler (2004)). They have emerged over the last decade as perhaps the leading empirical method for many classification tasks (Caruana and Niculescu-Mezil (2006); Criminisi *et al* (2012)). While random forests often learn ensembles of a hundred or more trees, because of the size of our datasets and the computational power available,

[7] See http://www.cs.waikato.ac.nz/ml/weka/.

we benchmark the performance by learning ensembles with 20 trees to provide a reasonable tradeoff between computation time and classification accuracy.

In all, we have 87 attributes in the models composed of account-level, credit-bureau, and macroeconomic data.[8] We acknowledge that, in practice, banks tend to segment their portfolios into distinct categories when using logistic regression and estimate different models on each segment. However, for our analysis, we do not perform any such segmentation. Our rationale is that our performance metric is solely based on classification accuracy. While it may be true that segmentation results in models that are more tailored to individual segments such as prime vs. subprime borrowers, thus potentially increasing forecast accuracy, we relegate this case to future research. For our current purposes, the number of attributes should be sufficient to approach the maximal forecast accuracy using logistic regression. We also note that decision tree models are well suited to aid in the segmentation process, and thus could be used in conjunction with logistic regression, but again leave this for future research.[9]

A. Attribute Selection

Although there are few papers in the literature that have detailed account-level data to benchmark our features, we believe we have selected a set that adequately represents current industry standards, in part based on our collective experience. Glennon et al. (2008) is one of the few papers with data similar to ours. These authors use industry experience and institutional knowledge to select and develop account-level, credit-bureau,

[8] We refer to our "variables" as attributes as is common in the machine-learning literature.

[9] Another reason for not differentiating across segments is that the results might reveal the identity of the banks to knowledgeable industry insiders. The same concern arises with the size of the portfolio.

and macroeconomic attributes. We start by selecting all possible candidate attributes that can be replicated from Glennon et al. (2008, Table 3). Although we cannot replicate all of their attributes, we do have the majority of those that are shown to be significant after their selection process.

We also merge macroeconomic variables to our sample using the five-digit ZIP code associated with the account. While we do not have a long time series of macro trends in our sample, there is a significant amount of cross-sectional heterogeneity that we use to pick up macro trends.

B. Dependent Variable

Our dependent variable is delinquency status. For the purposes of this study, we define delinquency as a credit-card account greater than or equal to 90 days past due. This differs from the standard accounting rule by which banks typically charge off accounts that are 180 days or more past due. However, it is rare for an account that is 90 days past due to be recovered, and is therefore common practice within the industry to use 90 days past due as a conservative definition of default. This definition is also consistent in the literature (see, e.g., Glennon et al. (2008) and Khandani et al. (2010)). We forecast all of our models over three different time horizons—two, three, and four quarters out—to classify whether or not an account becomes delinquent within those horizons.

C. Model Timing

To predict delinquency, we estimate separate machine-learning model every six months starting with the period ending 2010Q4.[10] We estimate these models at each point in time as if we were in that time period, i.e., no future data is ever used as inputs to a model, and require a historical training period and a future testing period. For example, a model for 2010Q4 is trained on data up to and including 2010Q4, but no further. Table 2 defines the dates for the training and test samples of each of our models.

The optimal length of the training window involves a tradeoff between increasing the amount of training data available and the stationarity of the training data (hence its relevance for predicting future performance). We use a rolling window of two years as the length of the training window to balance these two considerations. In particular, we combine the data from the most recent quarter with the data from 12 months prior to form a training sample. For example, the model trained on data ending in 2010Q4 contains the monthly credit-card accounts in 2009Q4 and 2010Q4. The average training sample thus contains about two million individual records, depending on the institution and time period. In fact, these rolling windows incorporate up to 24 months of information each because of the lag structure of some of the variables (e.g., year over year change in the HPI), and an addition 12 months over which an account could become 90 days delinquent.

[10] That is, we build models for the periods ending in 2010Q4, 2011Q2, 2011Q4, 2012Q2, 2012Q4, and 2013Q2. 2013Q2 is our last model because we need an out-of-sample test period to test our forecasts; it is used only for the two-quarter models.

D. Measuring Performance

The goal of our delinquency prediction models is to classify credit-card accounts into two categories: accounts that become 90 days or more past due within the next *n* quarters ("bad" accounts), and accounts that do not ("good" accounts). Therefore, our measure of performance should reflect the accuracy with which our model classifies the accounts into these two categories.

One common way to measure performance of such binary classification models is to calculate precision and recall. In our model, precision is defined as the number of correctly predicted delinquent accounts divided by the predicted number of delinquent accounts, while recall is defined as the number of correctly predicted delinquent accounts divided by the actual number of delinquent accounts. Precision is meant to gauge the number of false positives (accounts predicted to be delinquent that stayed current) while recall gauges the number of false negatives (accounts predicted to stay current that actually went into default).

We also consider two statistics that combine precision and recall, the F-measure and the kappa statistic. The F-Measure is defined as the harmonic mean of precision and recall, and is meant to describe the balance between precision and recall. The kappa statistic measures performance relative to random classification. According to Khandani et al. (2010) and Landis and Koch (1977), a kappa statistic above 0.6 represents substantial performance. Figure 2 summarizes the definitions of these classification performance statistics measures in a so-called "confusion matrix".

In the context of credit-card portfolio risk management, however, there are account-specific costs and benefits associated with the classification decision that these performance statistics fail to capture. In the management of existing lines of credit, the primary benefit of classifying bad accounts before they become delinquent is to save the lender the run-up that is likely to occur between the current time period and the time at which the borrower goes into default. On the other hand, there are costs associated with incorrectly classifying accounts as well. For example, the bank may alienate customers and lose out on potential future business and profits on future purchases.

To account for these possible gains and losses, we use a cost-sensitive measure of performance to compute the "value added" of our classifier, as in Khandani et al. (2010), by assigning different costs to false positives and false negatives, and approximating the total savings that our models would have brought if they had been implemented. Our value-added approach is able to assign a dollar-per-account savings (or cost) of implementing any classification model. From the lender's perspective, this provides an intuitive and practical method for choosing between models. From a supervisory perspective, we can assign deadweight costs of incorrect classifications by aggregate risk levels to quantify systemic risk levels.

Following Khandani et al. (2010), our value-added function is derived from the confusion matrix. Ideally, we would like to achieve 100% true positives and true negatives, implying correct classification of all accounts, delinquent and current. However, any realistic classification will have some false positives and negatives, which will be costly.

To quantify the value-added of classification, Khandani et al. (2010) define the profit with and without a forecast as follows:

$$\Pi_{\text{no forecast}} = (TP + FN)B_C P_M - (FP + TN)B_D \quad [1]$$

$$\Pi_{\text{forecast}} = TP\, B_C P_M - FP B_D - TN B_C \quad [2]$$

$$\Delta\Pi_{\text{no forecast}} = TN(B_D - B_C) - FNB_CP_M \quad [3]$$

where B_C is the current account balance; B_D is the balance at default; P_M is the profitability margin; and *TP*, *FN*, *FP*, and *TN* are defined according to the confusion matrix. Note that Eq. [3] is broken down into a savings from lowering balances (the first term) less a cost of misclassification (the second term).

To generate a value-added for each model, the authors then compare the savings from the forecast profit ($\Delta\Pi_{\text{forecast}}$) with the benefit of perfect foresight. The savings from perfect foresight can be calculated by multiplying the total number of bad accounts (*TN* + *FP*) by the run up (B_D – B_C). The ratio of the model forecast savings (Eq. [3]) to the perfect foresight case can be written as:

$$\text{Value-Added}\left(\frac{B_D}{B_C}, r, N\right) = \frac{TN - FN[1-(1+r)^{-N}]\left[\frac{B_D}{B_C} - 1\right]^{-1}}{TN + FP} \quad [4]$$

where we substitute $[1-(1+r)^{-N}]$ for the profitability margin, *r* is the discount rate, and *N* is the discount period.

IV. Classification Results

In this section we report the results of our classification models by bank and time. There are on average about 6.1 million accounts each month in our sample. Table 1 shows the sample sizes over time. There is a significant amount of heterogeneity in terms of delinquencies across institutions and time (see Figure 1). Delinquency rates necessarily increase with the forecast horizon, since the longer horizons include the shorter ones.

Annual delinquency rates range from 1.36% to 4.36%, indicating that the institutions we are studying have very different underwriting and/or risk-management strategies.

We run individual classification models for each bank over time; separate models are estimated for each forecast horizon for each bank. Because our data ends in 2014Q2, we can only test the three- and four-quarter-horizon models on the training periods ending in 2012Q2 and 2012Q4, respectively.[11]

A. Nonstationary Environments

A fundamental concern for all prediction algorithms is generalization, i.e., whether models will continue to perform well on out-of-sample data. This is particularly important when the environment that generates the data is itself changing, and therefore the out-of-sample data is almost guaranteed to come from a different distribution than the training data. This concern is particularly relevant for financial forecasting given the non-stationarity of financial data as well as the macroeconomic and regulatory environments. And our sample period, which starts on the heels of the 2008 financial crisis and the ensuing recession, only heightens these concerns.

We address overfitting primarily by testing out of sample. Our decision tree models also allow us to control the degree of in-sample fitting by controlling what is known as the pruning parameter, which we refer to as M. This parameter acts as the stopping criterion for the decision tree algorithm. For example, when M = 2, the algorithm will continue to attempt to add additional nodes to the leaves of the tree until there are two instances

[11] For example, for the four-quarter forecast models with training data ending 2012Q2, the dependent variable is defined over the period 2012Q2 through 2013Q2, making the test date 2013Q2. We then need one year of data to test the model out-of-sample which brings us to our last month of data coverage in 2014Q2.

(accounts) or less on each leaf, and an additional node would be statistically significant. As M increases, the in-sample performance will degrade because the algorithm stops even though there may be potentially statistically significant splits remaining. However, our out-of-sample performance may actually increase for a while because the nodes blocked by increasing M are overfitting the sample. Eventually, however, even the out-of-sample performance degrades as M becomes sufficiently high.

To find a suitable value of M for our machine-learning models, we conduct overfitting tests on data from a select bank by varying the M parameter from 2 to 5,000. Within each of the 15 clusters, a value of M = 50 seems to optimize performance under a variety of assumptions in our value-added calculation, although the results are not very sensitive between 25 and 250. Similar experiments with data from other banks produced similar results. In light of these results, we use a pruning parameter of M = 50 in all of our decision tree models.

B. Model Results

In this section we show the results of the comparison of our three modeling approaches—decision trees, logistic regression, and random forests. The random-forest models are estimated with 20 random trees.[12]

To preview the results and help visualize the effectiveness of our models in terms of discriminating between good and bad accounts, we plot the model-derived risk ranking

[12] The C4.5 models produced unreliable results for the 4Q forecast horizon for bank 5 due to a low delinquency rate combined with accounts that are difficult to classify (the corresponding logistic and random-forest forecasts were the worst performing models). The random-forest models for the 4Q forecast horizon for bank 2 failed to converge in a reasonable amount of time (run-time was stopped after 24+ hours at full capacity) so those results are omitted as well. Throughout the paper, those results are indicated with N/A.

versus an account's credit score at the time of the forecast in Figure 3 for Bank 2. Accounts are rank-ordered based on a logistic regression model for a two-quarter forecast horizon. Green points represent accounts that were current at the end of the forecast horizon; blue points represent accounts 30 days past due; yellow points represent accounts 60 days past due; and red points represent accounts 90 days or more past due. We plot each account's credit-bureau score on the horizontal axis because it is a key variable used in virtually every consumer-default prediction model, and serves as a useful comparison to the machine-learning forecast.

This plot show that while credit scores discriminate between good and bad accounts to a certain degree (the red 90+ days past due accounts do tend to cluster to the left region of the horizontal axis with lower credit scores), the C4.5 decision tree model is very effective in rank-ordering accounts in terms of riskiness. In particular, the red 90+ days past due points cluster heavily at the top of the graph, implying that the machine-learning forecasts are highly effective in identifying accounts that eventually become delinquent.[13]

Table 3 shows the precision and recall for our models. We also provide the true positive and false positive rates. The results are given by bank, time, and forecast horizon for each model type. The statistics are calculated for the classification threshold that maximizes the respective model's F-Measure to provide a reasonable balance between good precision and recall.

Although selecting a modeling threshold based on the test data does introduce some look-ahead bias, we use this approach when presenting the results for two reasons. First, banks are likely to calibrate classification models using an expected delinquency rate to

[13] Analogous plots for our logistic regression and random-forest models look very similar.

select the acceptance threshold. We do not separately model delinquency rates and view the primary purpose of our classifiers as the rank-ordering of accounts. To this end, we are less concerned with forecasting the realized delinquency rates than rank-ordering accounts based on risk of delinquency. Therefore, the main role of the acceptance threshold for our purposes is for exposition and to make fair comparisons across models.

Second, the performance statistics we report–the F-measure and the Kappa statistic–are relatively insensitive to the choice of modeling threshold. Figures A1 through A3 in the Appendix show the sensitivity of these performance statistics to the choice of acceptance threshold for the C4.5, logistic regression, and random-forest models, respectively. The three plots on the left in each figure show the F-measure versus the acceptance threshold while the plots on the right show the Kappa statistic. The lines are color-coded by bank and the points represent the maximum value of the line, i.e., the acceptance thresholds used in our models.

There are a few noteworthy points here. First, for each bank, the optimal threshold remains relatively constant over time, which means that it should be easy for a bank to select a threshold based on past results and get an adequate forecast. Second, in the cases where the selected threshold varies over time, the lines are still quite flat. For example, in our C4.5 decision tree models in Figure A1, the optimal thresholds cluster by bank and the curves are very flat between 20% to 70% threshold values for the F-measure and the kappa statistics. For the random-forest models in Figure A3, the lines are not as flat, but the optimal thresholds tend to cluster tightly for each bank. In sum, it is important to remember that the goal of a bank would not be to maximize the F-measure in any case, and as long as the selected threshold is selected using any reasonable strategy, our sensitivity analysis demonstrates that it would, in all likelihood, only have a minimal effect on our main results.

Each of the models achieves a very high true positive rate which is not surprising given the low default rates. The false positive rates are reasonable, between 11% and 38% for the two-quarter horizon models. However, as the forecast horizon increases, the models become less accurate and the false positive rates increase for each bank.

Table 4 presents the F-Measure and kappa statistics by bank and time. As mentioned above, the F-measure and kappa statistics show that the C4.5 and random-forest models outperform the logistic regression models. The performance of the models declines as the forecast horizon increases. Figures 4 and 5 present the F-measures and kappa statistic graphically for the six banks. The C4.5 and random-forest models tend to consistently out-perform the logistic regression, regardless of the forecast horizon, for each statistic.

Table 5 presents the value-added for each of the models, which represents the potential gain from employing a given model versus passive risk management. Under this metric, the C4.5 and random-forest models outperform the logistic regression models. We plot the comparisons in Figure 6 for the two quarter forecast horizon models.[14] The results are similar in that the C4.5 and random-forest models outperform logistic regression. All the value-added results assume a run-up of 30% and profitability margin of about 13.5%.

For the two quarter forecast horizon, the C4.5 models produce an average per bank cost savings of between 45.2% and 75.5%. The random forests yield similar values,

[14] We also have produced similar figures for the three and four quarter forecast horizons but omit them to conserve space.

between 47.0% and 74.4%. The logistic regressions fare much worse based on the bank average values because Banks 1 and 2 show two periods of negative value added—meaning that the models did such a poor job of classifying accounts that the bank would have been better off not managing accounts at all. Even omitting these negative instances, the logistic models tend to underperform the others.

There is substantial heterogeneity in value-added across banks as well. Figure 7 plots the value added for all six banks for each model type. All models are based on a two-quarter forecast horizon. Bank 3 is always at the top of the plots meaning that our models are performing the best. Bank 4 tends to be the lowest (although still has a positive value-added) and the other four banks cluster in between.

Moving to three- and four-quarter forecast horizons, the model performance declines and as a result the value-added declines. However, the C4.5 trees and random forests remain positive and continue to outperform logistic regression. Although the relative performance degrades somewhat, our machine-learning models still provide positive value at the longest forecast horizons.

Figure 8 presents the value-added versus the assumed run-up. The value-added for each model increases with run-up. With the exception of a 10% run-up for Bank 5, all the C4.5 and random-forest models generate positive value-added for any run-up of at least 10%. The logistic models however, need to have a run-up of at least 20% for Bank 1 to break even and never do so for Bank 2.

C. Risk Management Across Institutions

In this section, we examine risk management practices across institutions. First, we compare the credit-line management behavior across institutions. Second, we examine how well individual institutions target bad accounts. In credit cards, cutting lines is a very common tool used by banks to manage their risks and one we can analyze given our dataset.

As of each test date, we take the accounts which were predicted to default over a given horizon for a given bank, and analyze whether the bank cut its credit line or not. We use the predicted values from our models to simulate the banks' real problems and avoid any look-ahead bias. In Table 6 and Figure 9 we compute the mean of the ratio of the percent of lines cut for defaulted accounts to the percent of lines cut on all accounts. A ratio greater than 1 implies that the bank is effectively targeting accounts that turn out to be bad and cutting their credit lines at a disproportionately greater rate than they are cutting all accounts, a sign of effective risk management practices. Similarly, a ratio less than one implies the opposite.[15] We report the ratio for each quarter between the model prediction and the end of the forecast horizon because cutting lines earlier is better if indeed they turn out to become delinquent.

The results show a significant amount of heterogeneity across banks. For example, Figure 9 shows that three banks (2, 3, and 5) are very effective at cutting lines of accounts predicted to become delinquent—they are between 4.8 and 13.2 times more likely to target accounts predicted to default than the general portfolio. In contrast, Banks 4 and 6 underperform, rarely cut lines of accounts predicted to default. Bank 1 tends to cut the same number of good and bad accounts. There is no clear pattern to banks' targeting of bad accounts across the forecast horizon.

[15] We plot the natural logarithm of this ratio in Figure 9 so values above zero are interpreted as effective risk management.

Of course, these results are not conclusive because banks have other risk management strategies in addition to cutting lines, and our efficacy measure relies on the accuracy of our models. However, these empirical results show that, at a minimum, risk management policies differ significantly across major credit-card issuing financial institutions.

D. Attribute Analysis

A common criticism of machine-learning algorithms is that they are essentially black boxes, with results that are difficult to interpret. For example, given the chosen pruning and confidence limits of our decision tree models, the estimated decision trees tend to have about 100 leaves. The attributes selected vary across institutions and time, hence it is very difficult to compare the trees because of their complexity. Therefore, the first goal of our attribute analysis is to develop a method for interpreting the results of our machine-learning algorithms. The single decision-tree models learned using C4.5 are particularly intuitive.

We propose a relatively straightforward approach for combining the results of the decision tree output that captures the results by generating an index based on three main criteria. We start by constructing the following three metrics for each attribute in each decision tree:

1. *Log of the number of instances classified*: This is meant to capture the importance of the attribute. If attributes appear multiple times in a single model, we sum all the instances classified. This statistic is computed for each tree.
2. *The minimum leaf number*: The minimum leaf number is the highest node on the tree where the attribute sits, and roughly represents the statistical significance of

the attribute. The logic of the C4.5 classifier is that, in general, the higher up on the tree the attribute is (i.e., the lower the leaf number), the more important is it. Therefore, the attributes will be sorted in reverse order; that is, the variable with the lowest mean minimum leaf number would be ranked first. This statistic is computed for each tree.

3. *Indicator variable equal to 1 if the attribute appears in the tree and 0 otherwise*: We combine the results of multiple models over time to derive a bank-specific attribute ranking based on the number of times attributes are selected in a given model. For example, we run six separate C4.5 models for each bank using a two quarter forecast horizon. This ranking criterion is the number of times (between zero and six) that a given attribute is selected to a model. This statistic is meant to capture the stability of an attribute over time.

We combine the above statistics into a single ranking measure by standardizing each to have a mean of 0 and standard deviation of 1 and summing them by attribute. Attributes that do not appear in a model are assigned a score equal to the minimum of the standardized distribution. We then combine the scores for all unique bank-forecast horizon combinations and rank the attributes. This leaves us with 18 individual scores for each attribute used to rank them by importance. The most important attributes should have higher scores and appear near the top of the list and be raked lower (i.e., attribute 1 is the most important).

In all, 78 of the 87 attributes are selected in at least one model. Table 7 shows the mean attribute rankings across all models, by forecast horizon, and by bank. More important attributes are ranked lower. The table is sorted based on the mean ranking for each attribute across all 18 bank-forecast horizon pairs. Columns 2-4 show the mean ranking by forecast horizon and columns 5-10 show the mean ranking by bank.

It is reassuring that the top ranking variables—days past due, behavioral score, credit score, actual payment over minimum payment, one month change in utilization, etc.—are intuitive. For example, accounts that start out delinquent (less than 90 days) are most likely to become 90 days past due, regardless of the forecast horizon or bank.

Looking across forecast horizons, we do not see much variation. In fact, the pairwise Spearman rank correlations between the attribute rankings (for all 78 attributes that appear in at least one model) are between 89.8% and 94.3%.

However, there is a substantial amount of heterogeneity across banks, as suggested by the pairwise rank correlations between banks which range from 46.5% to 80.3%. This suggests that the key risk factors affecting delinquency vary across banks. For example, the change in one-month utilization (i.e., the percentage change in the drawdown of the credit line) has an average ranking between 2.0 and 4.0 for Banks 1, 2, and 5 but ranks between 10.3 and 15.7 for Banks 3, 4, and 6. For risk managers, this is a key attribute because managing drawdown and preventing run-up prior to default is central to managing credit-card risk. Large variation across banks in other attributes such as whether an account has entered into a workout program, the total fees, and whether an account is frozen further suggest that banks have different risk management strategies.

Overall, the results in Table 7 support the validity of our models and variable ranking criteria since the most widely used attributes in the industry tend to appear near the top of our rankings. However, looking across institutions, the results suggest that banks face different exposures, likely due to differences in underwriting practices and/or risk management strategies.

There is also substantial heterogeneity across banks in how macroeconomic variables affect their customers. Macroeconomic variables are more predictive for Banks 2 and 6 at a two-quarter forecast horizon, while for Bank 6, macroeconomic variables are captured as important factors at the one-year forecast horizon. The macroeconomic variables are only in the most important 20 attributes for Bank 2 and 6 in a two-quarter forecast horizon and for Bank 6 at the one-year forecast horizon. Although they are not the most important attributes, their ranking score is still relatively high and shows that the macroenvironment has a significant impact on consumer credit risk.

As mentioned above, we had also drawn the data three other times before. Using the data as of 2012Q4 (i.e. with 12 quarters of data from 2009Q1 to 2012Q4), our results showed greater macroeconomic sensitivity. The different results are consistent with intuition since the macroeconomic environment with a vantage point of 2012Q4 was quite different from the macroeconomic environment as of 2014Q2. These results emphasize the dynamic nature of machine-learning models, a particularly important feature for estimating industry relations in transition.

V. Conclusion

In this study, we employ a unique, very large dataset consisting of anonymized information from six large banks collected by a financial regulator to build and test decision-tree, regularized logistic regression, and random-forest models for predicting credit-card delinquency. The algorithms have access to combined consumer tradeline, credit bureau, and macroeconomic data from January 2009 to December 2013. We find that decision trees and random forests outperform logistic regression in both out-of-sample and out-of-time forecasts of credit-card delinquencies. The advantage of decision-trees and random forests over logistic regression is most significant at short time horizons. The success of these models implies that there may be a considerable amount of “money left on the table” by the credit-card issuers.

We also analyze and compare risk-management practices across the banks and compare drivers of delinquency across institutions. We find that there is substantial heterogeneity across banks in terms of risk factors and sensitivities to those factors. Therefore, no single model is likely to capture the delinquency tendencies across all institutions. The results also suggest that portfolio characteristics alone are not sufficient to identify the drivers of delinquency, since the banks actively manage the portfolios. Even a nominally high-risk portfolio may have fewer volatile delinquencies because of successful active risk management by the bank.

The heterogeneity of credit-card risk management practices across financial institutions has systemic implications. Credit-card receivables form an important component of modern asset-backed securities. We have found that certain banks are significantly more active and effective at managing the exposure of their credit-card portfolios, while credit-card delinquency rates across banks are also quite different in their macroeconomic sensitivities. An unexpected macroeconomic shock may thus propagate itself through the greater delinquency rate of credit cards issued by specific financial institutions into the asset-backed securities market.

Our study provides an in-depth illustration of the potential benefits that big data and machine-learning techniques can bring to consumers, risk managers, shareholders, and regulators, all of whom have a stake in avoiding unexpected losses and reducing the cost of consumer credit. Moreover, when aggregated across a number of financial institutions, the predictive analytics of machine-learning models provide a practical means for measuring systemic risk in one of the most important and vulnerable sectors of the economy. We plan to explore this application in ongoing and future research.

References

Breiman, L. , 2001. Random Forests, *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324

Breiman, L. and A. Cutler, 2004. Random Forests (Manual) URL:http://www.stat.berkeley.edu/users/breiman/RandomForests/cc_papers.htm

Caruana, R. and Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms. In *Proc. Int. Conf. Mach. Learn. (ICML)*

Cessie, S., van Houwelingen, J.C. , 1992. Ridge Estimators in Logistic Regression. *Applied Statistics*. 41(1):191-201.

A. Criminisi, J. Shotton, and E. Konukoglu. , 2012. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Found. Trends Comput. Graphics and Vision*, 7(2–3):81–227.

Frank, E., Hall, M. A., and Witten, I. H., 2011, *Data Mining: Practical Machine Learning Tools and Techniques,* Morgan Kaufmann, Burlington, MA.

Glennon, D., Kiefer, N., Larson, E., and Choi, H. , 2008. "Development and Validation of Credit-Scoring Models," *Journal of Credit Risk*, 4.

Khandani, A. E., Kim, A. J., Lo, A. W. , 2010. Consumer credit-risk models via machine-learning algorithms, *Journal of Banking & Finance*, 34(11), 2767–2787

Landis, J.R.; Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* **33** (1): 159–174. doi:10.2307/2529310

Quinlan, J. R. , 1993. *C4.5: programs for machine learning*, Morgan Kaufman, San Francisco, CA.

Thomas, L.C. , 2000. A survey of credit and behavioral scoring – forecasting financial risk of lending to consumers, *International Journal of Forecasting* 16(2) 163-167.

Table 1 - Sample Description

This table shows the total number of accounts over time. The six banks' data are combined to show the aggregate each quarter.

Date	Number of Accounts (1,000's)
2009Q4	5,696
2010Q2	5,677
2010Q4	5,787
2011Q2	5,960
2011Q4	5,306
2012Q2	6,300
2012Q4	6,580
2013Q2	6,643
2013Q4	6,604

Table 2 - Model Timing

This table shows the model timing. The first two columns represent the start and end dates of the training data. The test period columns show the quarter in which the models are tested. All models are meant to simulate a bank's actual forecasting problem "as if" they were at the test period start date.

Training Period	**Test Period Start**		
Start - End	**2Q Forecast**	**3Q Forecast**	**4Q Forecast**
2009Q4 - 2010Q4	2011Q2	2011Q3	2011Q4
2010Q2 - 2011Q2	2011Q4	2012Q1	2012Q2
2010Q4 - 2011Q4	2012Q2	2012Q3	2012Q4
2011Q2 - 2012Q2	2012Q4	2013Q1	2013Q2
2011Q4 - 2012Q4	2013Q2	2013Q3	N/A
2012Q2 - 2013Q2	2013Q4	N/A	N/A

Table 3 - Precision, Recall, True Positive Rate, and False Positive Rates by Bank

This table shows the precision, recall, true positive rate, and false positive rates by bank, time, and forecast horizon for each model type. The statistics are defined in Figure 2. The acceptance threshold is defined as the threshold which maximizes the F-Measure.

Panel A: 2 Quarter Forecast Horizon

		C4.5 Decision Trees				Logistic Regression				Random Forests			
Bank	Test Date	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate
1	201106	71.3%	63.0%	99.9%	37.0%	17.9%	59.1%	99.0%	40.9%	68.8%	67.8%	99.9%	32.2%
1	201112	62.8%	70.3%	99.8%	29.7%	26.0%	70.2%	98.8%	29.8%	65.0%	68.3%	99.8%	31.7%
1	201206	65.5%	67.8%	99.8%	32.2%	62.7%	60.0%	99.8%	40.0%	64.2%	69.1%	99.8%	30.9%
1	201212	68.0%	65.3%	99.8%	34.7%	62.6%	62.1%	99.8%	37.9%	66.2%	67.3%	99.8%	32.7%
1	201306	68.2%	59.9%	99.9%	40.1%	58.6%	59.3%	99.8%	40.7%	58.3%	70.1%	99.7%	29.9%
1	201312	67.1%	65.6%	99.8%	34.4%	60.6%	64.5%	99.8%	35.5%	64.5%	69.4%	99.8%	30.6%
	Average:	**67.2%**	**65.3%**	**99.8%**	**34.7%**	**48.1%**	**62.5%**	**99.5%**	**37.5%**	**64.5%**	**68.7%**	**99.8%**	**31.3%**
2	201106	63.7%	73.0%	99.4%	27.0%	64.2%	71.5%	99.4%	28.5%	65.9%	71.1%	99.4%	28.9%
2	201112	60.5%	75.9%	99.2%	24.1%	61.9%	71.3%	99.3%	28.7%	60.5%	74.2%	99.2%	25.8%
2	201206	64.8%	63.5%	99.4%	36.5%	3.1%	91.8%	53.9%	8.2%	63.4%	71.2%	99.3%	28.8%
2	201212	65.7%	70.7%	99.4%	29.3%	10.0%	67.7%	90.4%	32.3%	62.0%	73.9%	99.3%	26.1%
2	201306	66.5%	66.8%	99.5%	33.2%	63.6%	68.6%	99.4%	31.4%	61.7%	72.3%	99.3%	27.7%
2	201312	63.2%	73.0%	99.3%	27.0%	62.7%	71.2%	99.3%	28.8%	60.8%	72.6%	99.2%	27.4%
	Average:	**64.1%**	**70.5%**	**99.4%**	**29.5%**	**44.3%**	**73.7%**	**90.3%**	**26.3%**	**62.4%**	**72.5%**	**99.3%**	**27.5%**
3	201106	79.9%	88.8%	99.9%	11.2%	75.7%	81.2%	99.8%	18.8%	80.0%	87.7%	99.9%	12.3%
3	201112	69.2%	92.6%	99.7%	7.4%	72.5%	82.4%	99.8%	17.6%	80.5%	85.6%	99.9%	14.4%
3	201206	81.1%	84.9%	99.9%	15.1%	73.6%	81.7%	99.9%	18.3%	83.9%	79.0%	99.9%	21.0%
3	201212	79.5%	85.4%	99.9%	14.6%	72.4%	79.3%	99.9%	20.7%	79.0%	85.5%	99.9%	14.5%
3	201306	71.6%	90.2%	99.9%	9.8%	70.8%	80.3%	99.9%	19.7%	70.6%	90.8%	99.9%	9.2%
3	201312	74.8%	88.6%	99.9%	11.4%	70.7%	84.2%	99.9%	15.8%	70.8%	90.3%	99.9%	9.7%
	Average:	**76.0%**	**88.4%**	**99.9%**	**11.6%**	**72.6%**	**81.5%**	**99.9%**	**18.5%**	**77.5%**	**86.5%**	**99.9%**	**13.5%**

(Table 3 - Panel A, cont.)

Bank	Test Date	C4.5 Decision Trees Precision	C4.5 Decision Trees Recall	C4.5 Decision Trees True Positive Rate	C4.5 Decision Trees False Positive Rate	Logistic Regression Precision	Logistic Regression Recall	Logistic Regression True Positive Rate	Logistic Regression False Positive Rate	Random Forests Precision	Random Forests Recall	Random Forests True Positive Rate	Random Forests False Positive Rate
4	201106	59.4%	64.9%	99.7%	35.1%	57.2%	62.3%	99.7%	37.7%	58.7%	67.2%	99.7%	32.8%
4	201112	61.2%	70.0%	99.8%	30.0%	53.1%	67.1%	99.7%	32.9%	62.4%	67.3%	99.8%	32.7%
4	201206	67.4%	59.0%	99.9%	41.0%	57.6%	59.3%	99.8%	40.7%	59.0%	64.6%	99.8%	35.4%
4	201212	68.6%	60.5%	99.9%	39.5%	59.0%	62.1%	99.8%	37.9%	64.0%	62.1%	99.8%	37.9%
4	201306	62.3%	65.1%	99.8%	34.9%	61.5%	61.3%	99.8%	38.7%	61.3%	66.9%	99.8%	33.1%
4	201312	68.9%	60.7%	99.9%	39.3%	57.5%	67.1%	99.8%	32.9%	64.6%	65.6%	99.9%	34.4%
	Average:	**64.6%**	**63.4%**	**99.8%**	**36.6%**	**57.7%**	**63.2%**	**99.8%**	**36.8%**	**61.7%**	**65.6%**	**99.8%**	**34.4%**
5	201106	69.6%	72.8%	99.8%	27.2%	64.5%	71.8%	99.8%	28.2%	67.2%	76.0%	99.8%	24.0%
5	201112	66.1%	72.8%	99.8%	27.2%	65.7%	69.0%	99.8%	31.0%	64.1%	76.4%	99.8%	23.6%
5	201206	70.7%	64.4%	99.9%	35.6%	66.3%	62.2%	99.8%	37.8%	65.6%	72.5%	99.8%	27.5%
5	201212	66.2%	75.4%	99.8%	24.6%	63.5%	72.7%	99.8%	27.3%	66.1%	74.5%	99.8%	25.5%
5	201306	68.4%	71.0%	99.8%	29.0%	68.0%	68.8%	99.8%	31.2%	66.9%	75.4%	99.8%	24.6%
5	201312	63.3%	77.5%	99.7%	22.5%	66.6%	70.4%	99.8%	29.6%	64.3%	75.2%	99.8%	24.8%
	Average:	**67.4%**	**72.3%**	**99.8%**	**27.7%**	**65.7%**	**69.1%**	**99.8%**	**30.9%**	**65.7%**	**75.0%**	**99.8%**	**25.0%**
6	201106	69.7%	66.5%	99.9%	33.5%	64.6%	66.4%	99.8%	33.6%	69.9%	65.9%	99.9%	34.1%
6	201112	64.0%	71.1%	99.8%	28.9%	66.0%	66.9%	99.8%	33.1%	64.5%	70.6%	99.8%	29.4%
6	201206	74.7%	67.6%	99.8%	32.4%	69.9%	71.2%	99.8%	28.8%	70.9%	71.4%	99.8%	28.6%
6	201212	42.8%	90.4%	99.1%	9.6%	67.9%	70.2%	99.8%	29.8%	66.4%	72.9%	99.7%	27.1%
6	201306	36.2%	96.2%	98.7%	3.8%	70.6%	69.5%	99.8%	30.5%	71.7%	70.2%	99.8%	29.8%
6	201312	62.8%	72.5%	99.7%	27.5%	60.9%	72.3%	99.7%	27.7%	61.8%	71.7%	99.7%	28.3%
	Average:	**58.4%**	**77.4%**	**99.5%**	**22.6%**	**66.7%**	**69.4%**	**99.8%**	**30.6%**	**67.5%**	**70.4%**	**99.8%**	**29.6%**

Panel B: 3 Quarter Forecast Horizon													
		C4.5 Decision Trees				Logistic Regression				Random Forests			
Bank	Test Date	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate
1	201109	60.3%	45.8%	99.7%	54.2%	55.8%	42.0%	99.7%	58.0%	56.0%	50.0%	99.7%	50.0%
1	201203	59.0%	44.5%	99.7%	55.5%	54.5%	39.3%	99.7%	60.7%	56.3%	46.1%	99.7%	53.9%
1	201209	53.8%	47.6%	99.6%	52.4%	52.4%	40.1%	99.6%	59.9%	53.4%	47.4%	99.6%	52.6%
1	201303	55.6%	43.3%	99.7%	56.7%	52.0%	37.7%	99.7%	62.3%	49.5%	45.4%	99.6%	54.6%
1	201309	36.9%	54.0%	99.1%	46.0%	54.4%	35.4%	99.7%	64.6%	55.7%	44.1%	99.6%	55.9%
	Average:	**53.1%**	**47.0%**	**99.6%**	**53.0%**	**53.8%**	**38.9%**	**99.7%**	**61.1%**	**54.2%**	**46.6%**	**99.6%**	**53.4%**
2	201109	52.3%	51.5%	98.5%	48.5%	54.7%	45.9%	98.8%	54.1%	55.7%	48.1%	98.8%	51.9%
2	201203	55.2%	42.5%	98.9%	57.5%	46.8%	48.8%	98.3%	51.2%	48.9%	47.6%	98.5%	52.4%
2	201209	47.6%	56.0%	98.1%	44.0%	5.0%	80.4%	52.2%	19.6%	50.3%	52.0%	98.4%	48.0%
2	201303	51.1%	45.2%	98.9%	54.8%	9.3%	49.6%	87.8%	50.4%	N/A	N/A	N/A	N/A
2	201309	50.8%	50.8%	98.4%	49.2%	48.3%	50.9%	98.2%	49.1%	N/A	N/A	N/A	N/A
	Average:	**51.4%**	**49.2%**	**98.6%**	**50.8%**	**32.8%**	**55.1%**	**87.0%**	**44.9%**	**51.7%**	**49.3%**	**98.6%**	**50.7%**
3	201109	70.1%	56.4%	99.7%	43.6%	64.7%	51.8%	99.6%	48.2%	66.8%	57.8%	99.6%	42.2%
3	201203	70.6%	55.4%	99.8%	44.6%	65.2%	52.9%	99.7%	47.1%	71.2%	55.3%	99.8%	44.7%
3	201209	67.4%	56.8%	99.7%	43.2%	66.3%	53.1%	99.7%	46.9%	70.8%	55.8%	99.8%	44.2%
3	201303	66.7%	60.3%	99.8%	39.7%	64.8%	55.1%	99.8%	44.9%	69.4%	58.1%	99.8%	41.9%
3	201309	72.8%	60.8%	99.8%	39.2%	64.1%	58.9%	99.7%	41.1%	65.7%	63.6%	99.7%	36.4%
	Average:	**69.5%**	**58.0%**	**99.8%**	**42.0%**	**65.0%**	**54.4%**	**99.7%**	**45.6%**	**68.8%**	**58.1%**	**99.7%**	**41.9%**

(Table 3 - Panel B, cont.)

		C4.5 Decision Trees				Logistic Regression				Random Forests			
Bank	Test Date	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate
4	201109	46.1%	48.7%	99.4%	51.3%	46.7%	43.2%	99.5%	56.8%	52.0%	44.3%	99.5%	55.7%
4	201203	25.2%	56.2%	98.5%	43.8%	46.0%	41.0%	99.6%	59.0%	52.9%	42.5%	99.7%	57.5%
4	201209	53.4%	39.6%	99.7%	60.4%	43.8%	43.8%	99.5%	56.2%	47.3%	44.2%	99.5%	55.8%
4	201303	51.3%	38.9%	99.7%	61.1%	48.5%	37.2%	99.7%	62.8%	45.4%	43.4%	99.6%	56.6%
4	201309	46.3%	46.8%	99.5%	53.2%	44.7%	47.4%	99.5%	52.6%	54.4%	43.5%	99.7%	56.5%
	Average:	**44.5%**	**46.0%**	**99.4%**	**54.0%**	**46.0%**	**42.5%**	**99.5%**	**57.5%**	**50.4%**	**43.6%**	**99.6%**	**56.4%**
5	201109	30.6%	43.8%	99.2%	56.2%	30.2%	34.6%	99.3%	65.4%	40.0%	36.5%	99.5%	63.5%
5	201203	39.9%	31.2%	99.6%	68.8%	28.8%	32.4%	99.4%	67.6%	36.1%	37.1%	99.5%	62.9%
5	201209	40.4%	33.7%	99.6%	66.3%	22.9%	46.6%	98.7%	53.4%	39.3%	35.8%	99.5%	64.2%
5	201303	41.0%	31.2%	99.7%	68.8%	27.1%	37.4%	99.2%	62.6%	38.9%	34.5%	99.6%	65.5%
5	201309	42.1%	34.6%	99.6%	65.4%	32.6%	31.4%	99.4%	68.6%	42.2%	36.1%	99.6%	63.9%
	Average:	**38.8%**	**34.9%**	**99.5%**	**65.1%**	**28.3%**	**36.5%**	**99.2%**	**63.5%**	**39.3%**	**36.0%**	**99.5%**	**64.0%**
6	201109	48.0%	46.0%	99.4%	54.0%	48.3%	39.9%	99.5%	60.1%	56.0%	42.5%	99.6%	57.5%
6	201203	52.9%	43.3%	99.5%	56.7%	47.8%	42.0%	99.4%	58.0%	53.5%	45.3%	99.5%	54.7%
6	201209	42.9%	55.9%	98.9%	44.1%	52.1%	48.4%	99.4%	51.6%	58.2%	51.0%	99.5%	49.0%
6	201303	58.3%	42.8%	99.6%	57.2%	54.2%	43.2%	99.5%	56.8%	59.3%	44.4%	99.6%	55.6%
6	201309	47.7%	51.1%	99.2%	48.9%	48.6%	50.5%	99.3%	49.5%	54.1%	49.1%	99.4%	50.9%
	Average:	**50.0%**	**47.8%**	**99.3%**	**52.2%**	**50.2%**	**44.8%**	**99.4%**	**55.2%**	**56.2%**	**46.4%**	**99.5%**	**53.6%**

Panel C: 4 Quarter Forecast Horizon													
		C4.5 Decision Trees				Logistic Regression				Random Forests			
Bank	Test Date	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate
1	201112	52.5%	38.9%	99.5%	61.1%	26.6%	38.2%	98.5%	61.8%	48.5%	42.1%	99.4%	57.9%
1	201206	54.5%	36.5%	99.6%	63.5%	44.5%	35.5%	99.4%	64.5%	50.3%	39.2%	99.4%	60.8%
1	201212	49.2%	39.0%	99.5%	61.0%	45.0%	34.8%	99.5%	65.2%	48.9%	40.4%	99.5%	59.6%
1	201306	53.8%	34.4%	99.6%	65.6%	47.5%	29.1%	99.6%	70.9%	48.9%	35.3%	99.5%	64.7%
	Average:	**52.5%**	**37.2%**	**99.5%**	**62.8%**	**40.9%**	**34.4%**	**99.2%**	**65.6%**	**49.2%**	**39.3%**	**99.4%**	**60.7%**
2	201112	47.3%	43.1%	98.0%	56.9%	42.1%	47.7%	97.2%	52.3%	N/A	N/A	N/A	N/A
2	201206	53.6%	40.8%	98.5%	59.2%	6.6%	86.9%	46.2%	13.1%	N/A	N/A	N/A	N/A
2	201212	47.1%	43.6%	98.2%	56.4%	5.6%	84.3%	48.4%	15.7%	N/A	N/A	N/A	N/A
2	201306	51.0%	39.6%	98.6%	60.4%	12.9%	51.6%	86.9%	48.4%	N/A	N/A	N/A	N/A
	Average:	**49.8%**	**41.7%**	**98.3%**	**58.3%**	**16.8%**	**67.6%**	**69.7%**	**32.4%**	**N/A**	**N/A**	**N/A**	**N/A**
3	201112	63.1%	47.8%	99.6%	52.2%	62.0%	43.9%	99.6%	56.1%	64.2%	47.6%	99.6%	52.4%
3	201206	63.5%	41.9%	99.7%	58.1%	58.2%	41.3%	99.6%	58.7%	68.5%	40.0%	99.7%	60.0%
3	201212	57.9%	44.3%	99.6%	55.7%	49.6%	40.8%	99.5%	59.2%	57.3%	46.0%	99.6%	54.0%
3	201306	60.8%	43.9%	99.7%	56.1%	53.9%	44.3%	99.6%	55.7%	63.5%	42.7%	99.7%	57.3%
	Average:	**61.3%**	**44.5%**	**99.6%**	**55.5%**	**55.9%**	**42.6%**	**99.6%**	**57.4%**	**63.4%**	**44.1%**	**99.7%**	**55.9%**

(Table 3 - Panel C, cont.)

		C4.5 Decision Trees				Logistic Regression				Random Forests			
Bank	Test Date	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate	Precision	Recall	True Positive Rate	False Positive Rate
4	201112	38.8%	38.8%	99.2%	61.2%	37.0%	38.8%	99.1%	61.2%	44.3%	36.0%	99.4%	64.0%
4	201206	38.2%	37.8%	99.2%	62.2%	39.3%	33.6%	99.3%	66.4%	44.4%	33.4%	99.4%	66.6%
4	201212	42.9%	36.8%	99.4%	63.2%	40.2%	36.2%	99.3%	63.8%	40.6%	37.4%	99.3%	62.6%
4	201306	26.3%	43.4%	98.4%	56.6%	42.5%	34.7%	99.4%	65.3%	45.3%	36.4%	99.4%	63.6%
	Average:	**36.6%**	**39.2%**	**99.0%**	**60.8%**	**39.7%**	**35.8%**	**99.3%**	**64.2%**	**43.6%**	**35.8%**	**99.4%**	**64.2%**
5	201112	N/A	N/A	N/A	N/A	9.1%	31.2%	97.7%	68.8%	9.5%	24.7%	98.3%	75.3%
5	201206	N/A	N/A	N/A	N/A	8.9%	9.8%	99.2%	90.2%	11.8%	16.6%	99.0%	83.4%
5	201212	N/A	N/A	N/A	N/A	9.7%	25.8%	98.2%	74.2%	10.8%	22.0%	98.6%	78.0%
5	201306	N/A	N/A	N/A	N/A	8.9%	33.9%	97.0%	66.1%	10.9%	24.0%	98.3%	76.0%
	Average:	**N/A**	**N/A**	**N/A**	**N/A**	**9.1%**	**25.2%**	**98.0%**	**74.8%**	**10.8%**	**21.8%**	**98.5%**	**78.2%**
6	201112	49.9%	36.0%	99.4%	64.0%	48.1%	30.7%	99.4%	69.3%	47.0%	36.8%	99.3%	63.2%
6	201206	55.7%	37.6%	99.4%	62.4%	45.0%	38.8%	99.0%	61.2%	52.3%	40.9%	99.2%	59.1%
6	201212	38.9%	46.0%	98.6%	54.0%	54.0%	37.5%	99.4%	62.5%	49.5%	45.1%	99.1%	54.9%
6	201306	52.9%	40.9%	99.3%	59.1%	54.0%	40.8%	99.3%	59.2%	52.2%	44.2%	99.2%	55.8%
	Average:	**49.4%**	**40.1%**	**99.2%**	**59.9%**	**50.3%**	**36.9%**	**99.3%**	**63.1%**	**50.3%**	**41.8%**	**99.2%**	**58.2%**

Table 4 - F-Measure and Kappa Statistics by Bank and Time

This table shows the F-Measure and Kappa statistics by bank, time, and forecast horizon for each model type. The statistics are defined in Figure 2. The statistics are based on the acceptance threshold that maximizes the respective statistic for a given bank-time-model.

Panel A: 2 Quarter Forecast Horizon								
		F-Measure				Kappa Statistic		
Bank	Time	C4.5 Tree	Logistic	Random Forest		C4.5 Tree	Logistic	Random Forest
1	201012	66.9%	27.5%	68.3%		69.8%	49.9%	70.0%
1	201106	66.3%	37.9%	66.6%		68.1%	49.9%	68.4%
1	201112	66.6%	61.3%	66.5%		68.7%	65.0%	68.8%
1	201206	66.7%	62.3%	66.7%		68.3%	64.4%	68.7%
1	201212	63.8%	58.9%	63.6%		67.3%	61.9%	66.2%
1	201306	66.4%	62.5%	66.9%		68.4%	65.7%	68.7%
	Average:	**66.1%**	**51.7%**	**66.4%**		**68.4%**	**59.5%**	**68.5%**
2	201012	68.0%	67.7%	68.4%		69.2%	68.7%	69.1%
2	201106	67.3%	66.3%	66.7%		68.0%	66.9%	67.2%
2	201112	64.2%	6.0%	67.1%		67.9%	49.6%	67.9%
2	201206	68.1%	17.4%	67.4%		68.1%	49.6%	67.5%
2	201212	66.6%	66.0%	66.6%		67.3%	66.2%	66.8%
2	201306	67.8%	66.7%	66.2%		67.9%	66.9%	66.3%
	Average:	**67.0%**	**48.3%**	**67.0%**		**68.1%**	**61.3%**	**67.5%**
3	201012	84.1%	78.4%	83.7%		83.5%	78.0%	83.2%
3	201106	79.2%	77.1%	83.0%		75.6%	76.2%	82.4%
3	201112	82.9%	77.5%	81.4%		82.6%	77.0%	81.9%
3	201206	82.3%	75.7%	82.1%		81.8%	75.3%	81.5%
3	201212	79.8%	75.3%	79.4%		77.8%	74.6%	77.6%
3	201306	81.1%	76.9%	79.4%		79.3%	75.6%	77.4%
	Average:	**81.6%**	**76.8%**	**81.5%**		**80.1%**	**76.1%**	**80.7%**
4	201012	62.1%	59.6%	62.7%		65.3%	62.9%	65.1%
4	201106	65.3%	59.3%	64.7%		66.4%	63.2%	66.6%
4	201112	62.9%	58.4%	61.7%		66.3%	63.5%	65.2%
4	201206	64.3%	60.5%	63.0%		66.9%	62.8%	66.0%
4	201212	63.6%	61.4%	64.0%		67.7%	63.4%	66.0%
4	201306	64.6%	62.0%	65.1%		67.6%	64.2%	66.3%
	Average:	**63.8%**	**60.2%**	**63.5%**		**66.7%**	**63.3%**	**65.9%**
5	201012	71.2%	67.9%	71.3%		72.0%	68.0%	71.6%
5	201106	69.3%	67.3%	69.8%		70.0%	67.7%	70.6%
5	201112	67.4%	64.2%	68.9%		69.6%	67.2%	70.3%
5	201206	70.5%	67.8%	70.0%		70.2%	67.6%	70.0%
5	201212	69.7%	68.4%	70.9%		70.4%	69.3%	70.7%
5	201306	69.7%	68.4%	69.3%		69.9%	68.7%	70.0%
	Average:	**69.6%**	**67.3%**	**70.0%**		**70.4%**	**68.1%**	**70.5%**
6	201012	68.0%	65.5%	67.8%		68.7%	67.8%	69.7%
6	201106	67.4%	66.5%	67.4%		67.8%	68.3%	68.2%
6	201112	71.0%	70.5%	71.1%		72.3%	72.1%	72.0%
6	201206	58.1%	69.0%	69.5%		34.7%	69.6%	69.9%
6	201212	52.6%	70.0%	70.9%		13.0%	71.1%	72.2%
6	201306	67.3%	66.1%	66.4%		66.4%	64.4%	66.3%
	Average:	**64.1%**	**67.9%**	**68.9%**		**53.8%**	**68.9%**	**69.7%**

(Table 4, cont.)

		Panel B: 3 Quarter Forecast Horizon						
		F-Measure				Kappa Statistic		
Bank	Time	C4.5 Tree	Logistic	Random Forest		C4.5 Tree	Logistic	Random Forest
1	201012	52.0%	47.9%	52.8%		60.9%	59.1%	61.1%
1	201106	50.8%	45.7%	50.7%		59.8%	57.9%	60.0%
1	201112	50.5%	45.5%	50.2%		59.2%	57.9%	59.9%
1	201206	48.7%	43.7%	47.4%		58.6%	56.1%	58.0%
1	201212	43.8%	42.9%	49.2%		30.5%	56.2%	59.1%
	Average:	**49.2%**	**45.1%**	**50.1%**		**53.8%**	**57.5%**	**59.6%**
2	201012	51.9%	49.9%	51.6%		59.2%	58.7%	59.0%
2	201106	48.1%	47.8%	48.3%		57.8%	56.9%	57.0%
2	201112	51.5%	9.3%	51.1%		58.9%	49.2%	58.2%
2	201206	48.0%	15.6%	N/A		57.3%	49.4%	N/A
2	201212	50.8%	49.6%	N/A		56.9%	56.9%	N/A
	Average:	**50.0%**	**34.4%**	**50.3%**		**58.0%**	**54.2%**	**58.1%**
3	201012	62.5%	57.5%	61.9%		67.4%	64.2%	67.1%
3	201106	62.1%	58.4%	62.2%		67.6%	64.4%	67.7%
3	201112	61.7%	59.0%	62.4%		67.3%	65.0%	67.7%
3	201206	63.4%	59.5%	63.2%		68.1%	62.4%	68.0%
3	201212	66.3%	61.4%	64.6%		69.9%	65.6%	65.5%
	Average:	**63.2%**	**59.2%**	**62.9%**		**68.1%**	**64.3%**	**67.2%**
4	201012	47.3%	44.9%	47.9%		57.3%	55.6%	57.3%
4	201106	34.8%	43.4%	47.1%		-5.9%	55.8%	57.9%
4	201112	45.4%	43.8%	45.7%		56.6%	55.7%	56.8%
4	201206	44.2%	42.1%	44.4%		56.7%	54.9%	56.2%
4	201212	46.6%	46.0%	48.3%		58.1%	56.1%	57.7%
	Average:	**43.7%**	**44.0%**	**46.7%**		**44.6%**	**55.6%**	**57.2%**
5	201012	36.0%	32.3%	38.2%		21.9%	49.8%	53.6%
5	201106	35.0%	30.5%	36.6%		53.3%	49.8%	52.9%
5	201112	36.8%	30.7%	37.5%		52.2%	49.8%	52.6%
5	201206	35.5%	31.4%	36.6%		52.4%	49.8%	52.7%
5	201212	38.0%	32.0%	38.9%		52.7%	49.8%	53.4%
	Average:	**36.2%**	**31.4%**	**37.6%**		**46.5%**	**49.8%**	**53.0%**
6	201012	47.0%	43.7%	48.3%		47.7%	57.9%	58.1%
6	201106	47.6%	44.7%	49.1%		52.1%	58.1%	59.5%
6	201112	48.6%	50.1%	54.3%		40.4%	60.2%	61.0%
6	201206	49.4%	48.1%	50.8%		59.2%	58.8%	59.4%
6	201212	49.3%	49.6%	51.4%		47.2%	57.5%	57.5%
	Average:	**48.4%**	**47.2%**	**50.8%**		**49.3%**	**58.5%**	**59.1%**

(Table 4, cont.)

Panel C: 4 Quarter Forecast Horizon							
		F-Measure			Kappa Statistic		
Bank	Time	C4.5 Tree	Logistic	Random Forest	C4.5 Tree	Logistic	Random Forest
1	201012	44.7%	31.4%	45.1%	57.1%	49.6%	57.3%
1	201106	43.7%	39.5%	44.1%	56.8%	55.8%	56.9%
1	201112	43.5%	39.2%	44.3%	56.9%	55.4%	57.4%
1	201206	41.9%	36.1%	41.0%	56.2%	54.0%	56.2%
	Average:	**43.5%**	**36.5%**	**43.6%**	**56.7%**	**53.7%**	**56.9%**
2	201012	45.1%	44.7%	N/A	55.2%	55.0%	N/A
2	201106	46.3%	12.3%	N/A	53.6%	48.9%	N/A
2	201112	45.3%	10.6%	N/A	55.5%	49.1%	N/A
2	201206	44.6%	20.7%	N/A	55.5%	49.1%	N/A
	Average:	**45.3%**	**22.1%**	**N/A**	**54.9%**	**50.5%**	**N/A**
3	201012	54.4%	51.4%	54.7%	62.7%	61.0%	63.4%
3	201106	50.5%	48.3%	50.5%	61.7%	60.0%	62.1%
3	201112	50.2%	44.8%	51.0%	60.2%	56.7%	60.6%
3	201206	51.0%	48.6%	51.1%	61.7%	59.9%	61.7%
	Average:	**51.5%**	**48.3%**	**51.8%**	**61.6%**	**59.4%**	**62.0%**
4	201012	38.8%	37.8%	39.7%	54.8%	53.9%	55.2%
4	201106	38.0%	36.2%	38.1%	53.9%	52.9%	54.3%
4	201112	39.6%	38.1%	38.9%	54.4%	53.5%	54.4%
4	201206	32.7%	38.2%	40.4%	10.2%	53.7%	54.7%
	Average:	**37.3%**	**37.6%**	**39.3%**	**43.3%**	**53.5%**	**54.6%**
5	201012	N/A	14.1%	13.8%	N/A	49.8%	49.8%
5	201106	N/A	9.3%	13.8%	N/A	49.8%	49.8%
5	201112	N/A	14.1%	14.5%	N/A	49.7%	49.8%
5	201206	N/A	14.1%	15.0%	N/A	49.8%	49.8%
	Average:	**N/A**	**12.9%**	**14.3%**	**N/A**	**49.8%**	**49.8%**
6	201012	41.8%	37.5%	41.3%	55.8%	55.1%	55.7%
6	201106	44.9%	41.7%	45.9%	53.4%	56.9%	57.4%
6	201112	42.1%	44.3%	47.2%	36.3%	56.9%	57.0%
6	201206	46.1%	46.5%	47.9%	51.9%	58.0%	57.6%
	Average:	**43.7%**	**42.5%**	**45.6%**	**49.4%**	**56.7%**	**56.9%**

Table 5 - Value Added by Bank and Time

This table shows the value added results by bank, time, and forecast horizon for each model type. The statistics are based on the acceptance threshold that maximizes the respective statistic for a given bank-time-model. Value added is defined in Eq. (4). Each value assed assumes a margin of 5% (r = 5%), a run-up of 30% ($(B_d-B_r)/B_d$), and a discount horizon of three years (N = 3). The numbers represent the percentage cost savings of implementing each model versus passive risk management. The profit margin is used to estimate the opportunity cost of a false negative so that mis-classifying more profitable accounts is more costly.

		Value Added - 2Q Forecast			Value Added - 3Q Forecast			Value Added - 4Q Forecast		
Bank	Time	C4.5 Tree	Logistic	Random Forest	C4.5 Tree	Logistic	Random Forest	C4.5 Tree	Logistic	Random Forest
1	201106	51.5%	-63.9%	53.8%	32.1%	26.9%	32.2%	22.9%	-9.6%	21.8%
1	201112	51.4%	-20.5%	51.6%	30.5%	24.4%	29.9%	22.7%	15.4%	21.6%
1	201206	51.6%	43.8%	51.6%	29.0%	23.6%	28.6%	20.7%	15.5%	21.3%
1	201212	51.4%	45.2%	51.7%	27.6%	21.9%	24.4%	21.0%	14.5%	18.6%
1	201306	47.2%	40.2%	47.3%	12.1%	21.9%	28.2%	N/A	N/A	N/A
1	201312	51.0%	45.4%	52.1%	N/A	N/A	N/A	N/A	N/A	N/A
	Average:	**50.7%**	**15.1%**	**51.4%**	**26.3%**	**23.7%**	**28.6%**	**21.8%**	**8.9%**	**20.8%**
2	201106	54.1%	53.4%	54.4%	30.2%	28.6%	30.8%	21.3%	17.9%	N/A
2	201112	53.4%	51.4%	52.3%	26.9%	23.7%	25.1%	24.7%	-471.2%	N/A
2	201206	47.9%	-1201%	52.5%	28.0%	-618.7%	28.7%	21.3%	-555.0%	N/A
2	201212	53.9%	-209.7%	53.3%	25.6%	-171.0%	N/A	22.3%	-106.4%	N/A
2	201306	51.5%	50.7%	51.9%	28.5%	26.2%	N/A	N/A	N/A	N/A
2	201312	53.8%	51.9%	51.3%	N/A	N/A	N/A	N/A	N/A	N/A
	Average:	**52.4%**	**-200.5%**	**52.6%**	**27.8%**	**-142.2%**	**28.2%**	**22.4%**	**-278.6%**	**N/A**
3	201106	78.7%	69.4%	77.7%	45.5%	39.0%	44.7%	35.1%	31.7%	35.5%
3	201112	73.9%	68.2%	76.2%	44.9%	40.1%	45.1%	31.0%	27.8%	31.7%
3	201206	75.9%	68.4%	72.1%	44.3%	40.9%	45.3%	29.7%	22.0%	30.4%
3	201212	75.4%	65.6%	75.1%	46.7%	41.5%	46.4%	31.1%	27.1%	31.6%
3	201306	74.0%	65.3%	73.6%	50.5%	44.0%	48.5%	N/A	N/A	N/A
3	201312	75.0%	68.4%	73.4%	N/A	N/A	N/A	N/A	N/A	N/A
	Average:	**75.5%**	**67.5%**	**74.7%**	**46.4%**	**41.1%**	**46.0%**	**31.7%**	**27.1%**	**32.3%**
4	201106	44.8%	41.1%	45.7%	22.8%	20.8%	25.8%	11.0%	8.7%	15.4%
4	201112	49.8%	40.2%	48.9%	-19.6%	19.2%	25.3%	10.1%	10.0%	14.4%
4	201206	46.0%	39.5%	44.2%	23.9%	18.3%	21.8%	14.6%	11.8%	12.5%
4	201212	47.9%	42.5%	46.2%	22.1%	19.3%	19.7%	-11.9%	13.4%	16.4%
4	201306	47.2%	43.9%	47.7%	22.2%	20.8%	26.9%	N/A	N/A	N/A
4	201312	48.3%	44.6%	49.3%	N/A	N/A	N/A	N/A	N/A	N/A
	Average:	**47.3%**	**42.0%**	**47.0%**	**14.3%**	**19.7%**	**23.9%**	**6.0%**	**11.0%**	**14.7%**
5	201106	58.4%	53.8%	59.1%	-1.3%	-1.6%	11.6%	N/A	-110.0%	-81.4%
5	201112	55.9%	52.6%	57.0%	9.9%	-3.9%	7.2%	N/A	-36.1%	-40.0%
5	201206	52.3%	47.8%	55.3%	11.2%	-24.5%	10.8%	N/A	-83.2%	-60.3%
5	201212	57.9%	53.7%	57.1%	10.9%	-8.2%	9.9%	N/A	-124.3%	-64.9%
5	201306	56.1%	54.1%	58.4%	13.0%	1.9%	13.7%	N/A	N/A	N/A
5	201312	57.1%	54.4%	56.2%	N/A	N/A	N/A	N/A	N/A	N/A
	Average:	**56.3%**	**52.7%**	**57.2%**	**8.7%**	**-7.3%**	**10.6%**	**N/A**	**-88.4%**	**-61.7%**
6	201106	53.3%	49.9%	53.0%	23.4%	20.5%	27.3%	19.6%	15.7%	18.0%
6	201112	52.9%	51.3%	52.9%	25.8%	21.2%	27.4%	24.0%	17.3%	23.9%
6	201206	57.2%	57.3%	58.1%	22.2%	28.2%	34.3%	13.2%	23.0%	24.3%
6	201212	35.6%	55.1%	56.1%	28.9%	26.6%	30.6%	24.4%	25.0%	25.8%
6	201306	19.2%	56.3%	57.6%	25.7%	26.3%	30.1%	N/A	N/A	N/A
6	201312	53.1%	51.2%	51.6%	N/A	N/A	N/A	N/A	N/A	N/A
	Average:	**45.2%**	**53.5%**	**54.9%**	**25.2%**	**24.6%**	**30.0%**	**20.3%**	**20.2%**	**23.0%**

Table 6 – Credit Line Cuts

This table describes how banks manage credit lines. The numbers in the table represent the ratio of the percentage of accounts predicted to default whose credit lines were cut divided by the total percentage of accounts whose credit lines were cut. A ratio greater than one means a bank is likely actively targeting credit-card accounts to manage risk. The models are as defined above.

Panel A: 2 Quarter Forecast Horizon

		C4.5 Decision Trees				Logistic Regression				Random Forests			
Bank	Test Date	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	201106	1.37	0.87	-	-	1.63	1.32	-	-	1.26	0.80	-	-
1	201112	0.93	0.68	-	-	0.10	0.08	-	-	0.48	0.37	-	-
1	201206	0.72	0.46	-	-	0.42	0.23	-	-	0.70	0.43	-	-
1	201212	0.95	0.59	-	-	0.20	0.14	-	-	0.59	0.43	-	-
1	201306	1.14	0.69	-	-	0.60	0.40	-	-	0.53	0.28	-	-
1	201312	1.29	0.99	-	-	0.59	0.45	-	-	0.71	0.59	-	-
	Average:	**1.07**	**0.71**	**-**	**-**	**0.59**	**0.43**	**-**	**-**	**0.71**	**0.48**	**-**	**-**
2	201106	0.86	0.68	-	-	0.60	0.37	-	-	0.90	0.65	-	-
2	201112	0.47	0.41	-	-	0.58	0.56	-	-	0.36	0.36	-	-
2	201206	7.69	26.22	-	-	1.69	1.94	-	-	7.58	28.46	-	-
2	201212	9.01	24.86	-	-	2.03	1.74	-	-	9.05	25.80	-	-
2	201306	10.40	11.39	-	-	8.13	9.77	-	-	9.74	11.03	-	-
2	201312	3.30	11.04	-	-	2.64	10.80	-	-	2.76	11.09	-	-
	Average:	**5.29**	**12.43**	**-**	**-**	**2.61**	**4.20**	**-**	**-**	**5.07**	**12.90**	**-**	**-**
3	201106	5.12	6.71	-	-	5.40	6.73	-	-	5.04	6.71	-	-
3	201112	6.23	6.98	-	-	6.82	7.04	-	-	6.20	7.02	-	-
3	201206	7.41	8.06	-	-	7.64	8.06	-	-	7.27	8.03	-	-
3	201212	7.19	7.40	-	-	7.58	7.39	-	-	7.10	7.44	-	-
3	201306	7.23	7.59	-	-	7.42	7.49	-	-	7.16	7.65	-	-
3	201312	6.98	7.48	-	-	7.36	7.48	-	-	7.15	7.76	-	-
	Average:	**6.69**	**7.37**	**-**	**-**	**7.04**	**7.37**	**-**	**-**	**6.65**	**7.43**	**-**	**-**
4	201106	0.34	0.19	-	-	0.00	0.00	-	-	0.00	0.00	-	-
4	201112	0.29	0.16	-	-	0.00	0.00	-	-	0.17	0.10	-	-
4	201206	0.00	0.00	-	-	0.00	0.00	-	-	0.00	0.00	-	-
4	201212	0.00	0.00	-	-	0.00	0.00	-	-	0.00	0.00	-	-
4	201306	0.00	0.00	-	-	0.00	0.00	-	-	0.00	0.00	-	-
4	201312	0.00	0.00	-	-	0.00	0.00	-	-	0.00	0.00	-	-
	Average:	**0.11**	**0.06**	**-**	**-**	**0.00**	**0.00**	**-**	**-**	**0.03**	**0.02**	**-**	**-**

(Table 7 - Panel A, cont.)

		C4.5 Decision Trees				Logistic Regression				Random Forests			
Bank	Test Date	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
5	201106	7.10	4.55	-	-	6.46	4.03	-	-	6.24	3.73	-	-
5	201112	7.76	5.33	-	-	6.64	4.75	-	-	6.22	4.40	-	-
5	201206	6.68	4.03	-	-	4.22	2.42	-	-	5.05	2.95	-	-
5	201212	8.42	5.99	-	-	7.72	5.23	-	-	8.83	5.95	-	-
5	201306	8.57	5.28	-	-	5.21	3.44	-	-	6.19	3.86	-	-
5	201312	4.99	3.82	-	-	4.37	3.27	-	-	4.28	3.31	-	-
	Average:	**7.25**	**4.83**	**-**	**-**	**5.77**	**3.86**	**-**	**-**	**6.13**	**4.03**	**-**	**-**
6	201106	1.12	0.34	-	-	2.60	0.80	-	-	0.88	0.27	-	-
6	201112	0.79	0.60	-	-	0.25	0.14	-	-	0.14	0.08	-	-
6	201206	0.91	0.33	-	-	0.43	0.20	-	-	0.78	0.29	-	-
6	201212	0.34	0.14	-	-	0.36	0.17	-	-	0.23	0.11	-	-
6	201306	0.00	0.12	-	-	0.12	0.07	-	-	0.14	0.08	-	-
6	201312	0.47	0.28	-	-	0.24	0.21	-	-	0.25	0.15	-	-
	Average:	**0.61**	**0.30**	**-**	**-**	**0.67**	**0.26**	**-**	**-**	**0.40**	**0.16**	**-**	**-**

		C4.5 Decision Trees				Logistic Regression				Random Forests			
Panel B: 3 Quarter Forecast Horizon													
Bank	Test Date	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	201109	0.93	0.90	0.85	-	0.66	0.56	0.51	-	0.79	0.74	0.68	-
1	201203	1.36	0.93	0.74	-	0.45	0.34	0.28	-	0.76	0.55	0.43	-
1	201209	0.99	0.78	0.76	-	0.36	0.22	0.24	-	0.56	0.46	0.48	-
1	201303	1.41	1.18	1.07	-	0.69	0.56	0.43	-	0.37	0.30	0.31	-
1	201309	2.72	1.68	1.29	-	0.46	0.54	0.49	-	0.25	0.31	0.33	-
	Average:	**1.48**	**1.10**	**0.94**	**-**	**0.52**	**0.44**	**0.39**	**-**	**0.55**	**0.47**	**0.44**	**-**
2	201109	0.65	0.55	0.55	-	0.51	0.45	0.50	-	N/A	N/A	N/A	-
2	201203	0.50	0.61	0.60	-	0.53	0.32	0.33	-	0.23	0.28	0.31	-
2	201209	8.66	27.52	19.18	-	1.67	1.93	1.96	-	6.90	28.42	19.95	-
2	201303	11.85	14.92	11.83	-	1.83	1.91	1.99	-	N/A	N/A	N/A	-
2	201309	3.58	8.63	7.23	-	1.69	4.81	4.01	-	N/A	N/A	N/A	-
	Average:	**5.05**	**10.45**	**7.88**	**-**	**1.25**	**1.89**	**1.76**	**-**	**3.56**	**14.35**	**10.13**	**-**
3	201109	5.90	7.43	5.78	-	6.76	7.48	5.78	-	5.71	7.49	5.82	-
3	201203	6.33	7.01	5.85	-	6.54	7.12	5.92	-	6.32	7.12	5.92	-
3	201209	7.12	7.45	5.89	-	7.33	7.58	5.99	-	7.07	7.56	5.98	-
3	201303	7.44	7.37	5.88	-	7.05	6.45	5.31	-	7.36	7.44	5.93	-
3	201309	5.80	6.06	4.77	-	6.61	6.76	5.32	-	7.08	7.61	5.97	-
	Average:	**6.52**	**7.06**	**5.64**	**-**	**6.86**	**7.08**	**5.66**	**-**	**6.71**	**7.44**	**5.92**	**-**
4	201109	0.49	0.39	0.33	-	0.23	0.13	0.20	-	0.00	0.00	0.00	-
4	201203	0.16	0.24	0.22	-	0.00	0.00	0.00	-	0.21	0.22	0.20	-
4	201209	0.52	0.31	0.25	-	0.00	0.00	0.00	-	0.00	0.00	0.00	-
4	201303	0.00	0.00	0.00	-	0.00	0.00	0.00	-	0.00	0.00	0.00	-
4	201309	0.00	0.00	0.00	-	0.00	0.00	0.00	-	0.00	0.00	0.00	-
	Average:	**0.23**	**0.19**	**0.16**	**-**	**0.05**	**0.03**	**0.04**	**-**	**0.04**	**0.04**	**0.04**	**-**
5	201109	11.43	7.69	6.00	-	2.52	1.70	1.20	-	5.35	3.29	2.93	-
5	201203	10.62	7.33	5.74	-	2.83	1.99	1.68	-	7.84	5.62	4.40	-
5	201209	11.57	7.82	6.21	-	1.58	0.91	0.64	-	6.70	4.35	3.55	-
5	201303	13.67	9.44	7.61	-	2.99	2.18	1.74	-	11.53	8.10	6.18	-
5	201309	11.18	7.10	6.02	-	2.09	1.52	1.22	-	6.85	4.12	3.31	-
	Average:	**11.69**	**7.88**	**6.32**	**-**	**2.40**	**1.66**	**1.30**	**-**	**7.66**	**5.10**	**4.07**	**-**
6	201109	1.18	0.92	0.76	-	1.51	1.15	0.88	-	0.41	0.36	0.27	-
6	201203	0.54	0.40	0.28	-	1.07	0.80	0.46	-	0.55	0.42	0.26	-
6	201209	0.38	0.26	0.16	-	0.16	0.10	0.08	-	0.23	0.17	0.13	-
6	201303	0.29	0.29	0.25	-	0.08	0.09	0.08	-	0.07	0.09	0.08	-
6	201309	0.82	0.46	0.41	-	0.12	0.14	0.11	-	0.00	0.00	0.00	-
	Average:	**0.64**	**0.46**	**0.37**	**-**	**0.58**	**0.46**	**0.32**	**-**	**0.25**	**0.21**	**0.15**	**-**

Panel C: 4 Quarter Forecast Horizon													
		C4.5 Decision Trees				Logistic Regression				Random Forests			
Bank	Test Date	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	201112	1.37	1.11	0.94	0.85	0.11	0.10	0.10	0.10	0.86	0.64	0.56	0.51
1	201206	1.05	0.80	0.71	0.64	0.40	0.30	0.26	0.26	0.78	0.62	0.53	0.46
1	201212	1.60	1.22	1.08	0.96	0.35	0.29	0.32	0.25	0.78	0.69	0.60	0.53
1	201306	1.30	0.88	0.87	0.78	0.67	0.50	0.55	0.51	0.27	0.21	0.25	0.25
	Average:	**1.33**	**1.00**	**0.90**	**0.81**	**0.38**	**0.30**	**0.31**	**0.28**	**0.67**	**0.54**	**0.49**	**0.44**
2	201112	0.52	0.60	0.50	0.42	0.37	0.48	0.37	0.30	N/A	N/A	N/A	N/A
2	201206	6.93	21.08	15.22	13.08	1.74	1.95	1.97	1.97	N/A	N/A	N/A	N/A
2	201212	8.88	23.94	20.26	18.64	1.75	1.95	1.91	1.87	N/A	N/A	N/A	N/A
2	201306	8.29	8.50	6.91	6.28	1.84	2.00	2.27	2.35	N/A	N/A	N/A	N/A
	Average:	**6.15**	**13.53**	**10.72**	**9.60**	**1.42**	**1.59**	**1.63**	**1.62**	**N/A**	**N/A**	**N/A**	**N/A**
3	201112	6.41	6.88	5.81	5.07	6.43	6.94	5.84	5.09	6.26	6.92	5.84	5.10
3	201206	7.28	7.76	6.15	5.13	7.35	7.75	6.14	5.12	7.25	7.93	6.26	5.21
3	201212	7.02	7.21	5.70	4.84	7.47	7.11	5.63	4.78	7.10	7.37	5.81	4.93
3	201306	7.03	7.26	5.74	4.81	7.12	7.05	5.63	4.74	N/A	N/A	N/A	N/A
	Average:	**6.93**	**7.28**	**5.85**	**4.96**	**7.09**	**7.21**	**5.81**	**4.93**	**6.87**	**7.41**	**5.97**	**5.08**
4	201112	0.48	0.34	0.23	0.22	0.27	0.15	0.11	0.10	0.25	0.14	0.10	0.09
4	201206	0.23	0.18	0.16	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	201212	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	201306	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Average:	**0.18**	**0.13**	**0.10**	**0.09**	**0.07**	**0.04**	**0.03**	**0.02**	**0.06**	**0.03**	**0.02**	**0.02**
5	201112	N/A	N/A	N/A	N/A	0.00	0.00	0.00	1.74	N/A	N/A	N/A	N/A
5	201206	N/A	N/A	N/A	N/A	0.00	4.01	5.45	4.31	N/A	N/A	N/A	N/A
5	201212	N/A	N/A	N/A	N/A	0.00	1.16	0.85	1.36	N/A	N/A	N/A	N/A
5	201306	N/A	N/A	N/A	N/A	0.00	1.19	0.87	0.75	N/A	N/A	N/A	N/A
	Average:	**N/A**	**N/A**	**N/A**	**N/A**	**0.00**	**1.59**	**1.79**	**2.04**	**N/A**	**N/A**	**N/A**	**N/A**
6	201112	1.35	0.74	0.65	0.39	0.15	0.08	0.07	0.12	0.52	0.29	0.30	0.18
6	201206	1.21	0.50	0.40	0.28	0.49	0.22	0.16	0.10	1.11	0.41	0.30	0.22
6	201212	0.85	0.34	0.34	0.31	0.19	0.10	0.12	0.10	0.53	0.24	0.21	0.21
6	201306	0.27	0.31	0.29	0.27	0.00	0.06	0.09	0.08	0.23	0.18	0.16	0.17
	Average:	**0.92**	**0.47**	**0.42**	**0.31**	**0.21**	**0.12**	**0.11**	**0.10**	**0.60**	**0.28**	**0.25**	**0.20**

Table 7 – Attribute Analysis

This table shows the mean attribute ranking across all models, by forecast horizon, and by bank. For each unique bank and forecast horizon pair, the time series of C4.5 decision tree models reported in Tables 3-6 are combined and attributes are assigned a score based on 1) the number of instances classified, 2) the minimum leaf on each tree they appear, and 3) the number of models for which they are selected. The scores are standardized and summed to generate an importance metric for each attribute for each bank-forecast horizon pair. More important attributes are ranked lower. The table is sorted based on the mean ranking for each attribute across all bank-forecast horizon pairs. Columns 2-4 show the mean ranking by forecast horizon and columns 5-10 show the mean ranking by bank. In all, 78 of the 87 attributes were selected in at least one model.

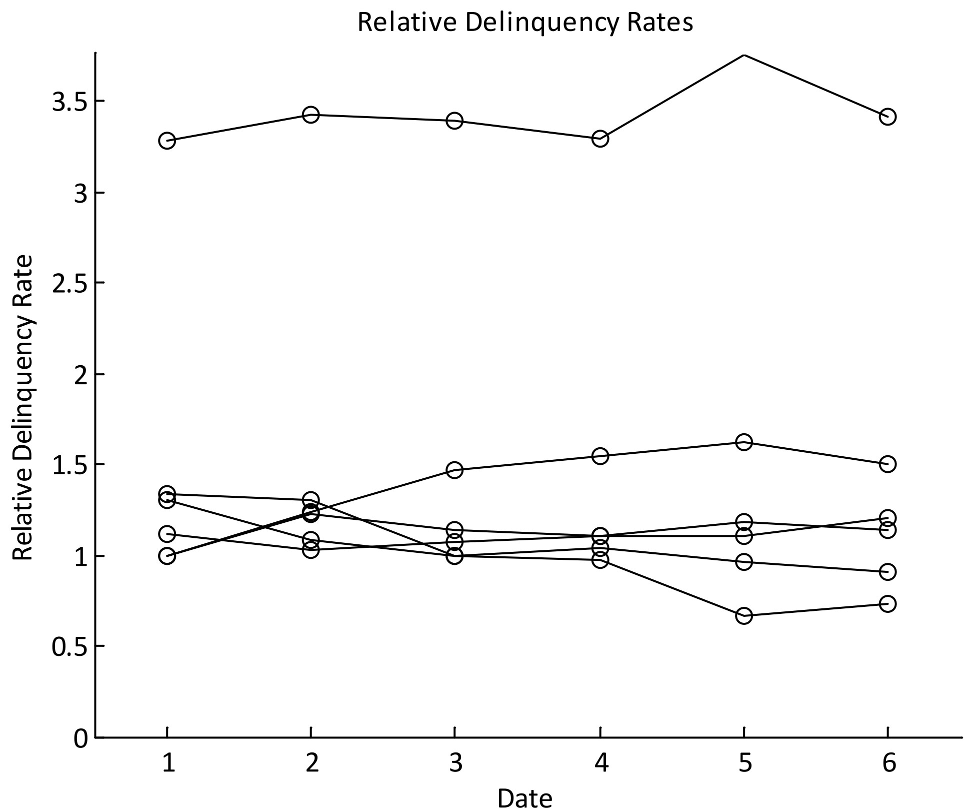
Attribute	All Models	2Q Horizon	3Q Horizon	4Q Horizon	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Days past due	1.4	1.2	1.7	1.5	1.0	1.7	1.0	1.7	2.3	1.0
Behavioral score	3.7	3.2	4.3	3.7	8.3	1.3	2.3	3.0	1.7	5.7
Refreshed credit score	6.3	7.8	6.0	5.0	5.0	8.0	7.0	9.0	4.0	4.7
Actual payment / minimum payment	6.7	5.2	6.3	8.5	9.7	11.3	3.7	5.7	5.0	4.7
1 mo. chg. in monthly utilization	7.8	5.5	7.8	10.0	4.0	3.3	15.7	11.3	2.0	10.3
Payment equal minimum payment in past 3 mo.s (0,1)	8.6	7.8	8.5	9.5	6.3	8.7	6.7	10.3	6.3	13.3
Cycle end balance	9.8	10.8	10.2	8.3	11.0	6.3	12.7	7.7	17.0	4.0
Cycle end balance	11.9	8.8	16.0	11.0	3.0	13.0	13.0	14.0	12.0	16.7
Cycle utilization	12.1	19.3	8.7	8.3	8.7	21.3	4.7	22.3	9.0	6.7
Number of accounts 30+ days past due	12.6	12.8	12.7	12.2	18.7	5.0	10.3	7.3	13.0	21.0
Total fees	15.9	16.2	12.8	18.8	15.0	21.3	8.3	14.3	9.3	27.3
Workout program flag	16.8	23.5	14.2	12.7	6.7	19.3	10.3	4.0	24.0	36.3
Total number of bank card accounts	17.8	18.5	17.5	17.3	22.0	21.7	19.0	14.3	17.7	12.0
Current credit limit	17.9	18.8	18.2	16.7	21.0	7.7	30.7	16.7	10.0	21.3
Line frozen flag (current mo.)	17.9	17.5	15.7	20.5	9.7	16.3	48.7	1.3	9.0	22.3
Monthly utilization	19.9	21.5	15.3	23.0	16.7	30.0	42.3	12.0	13.7	5.0
Number of accounts 60+ days past due	23.2	22.3	27.2	20.0	21.0	19.0	20.7	18.7	19.3	40.3
3 mo. chg. in credit score	24.4	21.8	24.2	27.2	8.7	27.3	28.3	21.7	32.3	28.0
Number of accounts in charge off status	26.3	26.0	27.7	25.2	27.3	17.0	24.0	18.3	39.0	32.0
1 mo. chg. in cycle utilization	27.0	29.3	26.7	25.0	17.7	38.3	10.7	30.3	28.3	36.7
6 mo. chg. in credit score	27.1	28.8	28.3	24.2	12.7	42.3	25.0	41.3	20.3	21.0
Total number of accounts 60+ days past due	27.9	21.5	32.3	30.0	31.7	24.3	18.0	11.3	41.3	41.0
Total balance on all 60+ days past due accounts	30.2	36.5	30.5	23.7	36.3	28.0	19.7	17.7	32.3	47.3
Total number of accounts verified	30.3	32.3	28.0	30.7	46.7	18.7	42.7	31.0	24.7	18.3
Flag if greater than 0 accounts 60 days past due	30.5	36.2	27.2	28.2	39.3	42.3	16.0	36.0	34.3	15.0
Line frozen flag (1 mo. lag)	30.9	15.5	34.5	42.7	16.3	8.0	33.3	29.0	47.3	51.3
3 mo. chg. in monthly utilization	33.4	30.2	34.8	35.2	19.0	22.7	31.7	42.7	40.0	44.3
Number of accounts 90+ days past due	33.7	43.5	29.8	27.8	34.3	25.0	33.3	31.7	36.0	42.0
6 mo. chg. in behavioral score	34.6	34.5	37.2	32.2	36.0	55.7	22.0	45.3	21.7	27.0
Account exceeded the limit in past 3 mo.s (0,1)	35.3	28.5	46.0	31.3	31.0	23.0	64.7	28.3	34.0	30.7

(Table 7, cont.)

Attribute	All Models	2Q Horizon	3Q Horizon	4Q Horizon	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
3 mo. chg. in cycle utilization	35.4	28.8	33.5	44.0	29.7	48.0	29.0	38.7	18.7	48.7
Flag if the card is securitized	36.2	35.5	36.7	36.3	24.0	13.7	30.3	28.7	71.7	48.7
Total number of accounts opened in the past year	36.4	41.7	36.0	31.5	41.0	24.0	38.7	45.0	28.3	41.3
Total number of bank card accounts 60+ days past due	37.4	38.5	32.8	41.0	47.3	25.0	23.7	25.3	40.7	62.7
Total balance of all revolving accounts / total balance	39.3	41.0	34.5	42.5	30.0	40.3	43.0	43.3	33.3	46.0
Total number of accounts	41.3	34.2	48.7	41.0	40.7	26.3	35.3	32.3	64.0	49.0
Product type	41.4	38.5	41.7	44.0	20.3	61.0	73.0	71.7	11.3	11.0
Unemployment rate	41.6	41.8	37.2	45.7	42.3	36.7	48.3	54.7	29.3	38.0
Flag if greater than 0 accounts 30 days past due	41.6	47.7	39.7	37.5	55.3	37.3	35.7	44.7	22.0	54.7
Purchase volume / credit limit	43.4	43.5	38.2	48.5	30.3	58.3	32.3	70.3	36.0	33.0
Utilizatiion of all bank card accounts	45.2	53.5	39.5	42.7	39.0	54.0	63.3	52.7	28.3	34.0
Flag if greater than 0 accounts opened in the past year	45.8	49.7	44.0	43.8	64.0	25.7	56.7	58.0	38.7	32.0
Flag if greater than 0 accounts 90 days past due	46.2	47.7	44.8	46.0	42.7	38.3	28.3	54.7	60.0	53.0
Avg. weekly hours worked (private) (12 mo. chg.)	46.2	44.8	49.2	44.5	61.0	37.0	55.7	42.7	52.3	28.3
Avg. hourly wage (private) (3 mo. chg.)	47.7	49.5	43.2	50.3	53.7	56.3	60.0	45.3	36.3	34.3
Avg. weekly hours worked (leisure) (12 mo. chg.)	47.9	49.7	43.0	51.0	53.3	40.0	57.0	60.7	54.3	22.0
Number of total nonfarm (NSA)	48.2	53.2	48.2	43.3	40.7	54.3	52.0	48.7	49.7	44.0
Avg. weekly hours worked (trade and transportation) (1	48.6	46.7	51.0	48.2	49.3	49.0	34.3	52.0	51.0	56.0
Avg. weekly hours worked (private) (3 mo. chg.)	49.8	48.2	44.2	57.0	48.7	46.7	53.0	42.3	50.3	57.7
Number of total nonfarm (NSA) (12 mo. chg.)	50.2	49.7	45.2	55.7	45.3	58.0	50.3	44.3	49.0	54.0
Avg. weekly hours worked (trade and transportation) (3	50.3	50.8	50.3	49.7	52.7	44.0	55.0	61.0	44.7	44.3
Avg. hourly wage (trade and transportation) (3 mo. chg	50.3	48.8	50.0	52.2	55.3	38.0	61.3	38.0	54.3	55.0
Total non-mortgage balance / total limit	50.6	55.0	46.3	50.3	51.7	64.7	55.7	38.7	46.0	46.7
Avg. hourly wage (private) (12 mo. chg.)	51.8	50.3	53.5	51.5	56.0	45.7	59.0	54.0	47.3	48.7
Avg. hourly wage (trade and transportation) (12 mo. ch	51.8	57.2	48.8	49.3	52.0	55.0	60.0	47.3	37.3	59.0
Avg. weekly hours worked (leisure) (3 mo. chg.)	51.9	52.5	50.5	52.7	51.3	43.3	39.7	59.7	64.7	52.7
6 mo. chg. in cycle utilization	52.1	46.7	54.7	54.8	33.0	70.3	48.0	64.7	38.3	58.0
Avg. hourly wage (leisure) (12 mo. chg.)	53.2	49.0	53.5	57.2	47.0	48.3	53.3	46.3	62.0	62.3
Avg. hourly wage (leisure) (3 mo. chg.)	53.6	52.7	52.7	55.5	58.7	60.7	62.3	37.3	66.3	36.3
Total credit limit to number of open bank cards	54.0	52.0	52.3	57.7	68.0	56.0	45.3	41.7	49.0	64.0
Number of total nonfarm (NSA) (3 mo. chg.)	54.2	51.3	55.2	56.0	62.3	45.0	54.0	57.3	40.0	66.3
Flag if total limit on all bank cards greater than zero	54.8	50.0	60.2	54.3	59.3	72.0	43.7	33.3	67.3	53.3
Unemployment rate (3 mo. chg.)	55.0	58.3	53.5	53.2	52.3	56.0	68.3	55.3	52.7	45.3
Number of total nonfarm (NSA) (3 mo. chg.)	55.9	59.2	64.2	44.3	58.0	61.7	50.0	55.0	62.0	48.7
Total private (NSA) (12 mo. chg.)	56.0	57.5	53.5	57.0	53.3	47.7	54.3	64.3	56.3	60.0
Percent chg. in credit limit (lagged 1 mo.)	56.5	57.0	52.5	60.0	66.7	74.3	10.7	68.7	58.7	60.0

(Table 7, cont.)

Attribute	All Models	2Q Horizon	3Q Horizon	4Q Horizon	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6
Unemployment rate (12 mo. chg.)	58.3	53.8	65.3	55.7	42.7	66.7	66.7	61.7	64.3	47.7
Percent chg. in credit limit current 1 mo.	58.4	60.0	59.7	55.5	71.0	74.7	11.7	65.7	68.7	58.7
6 mo. chg. in monthly utilization	58.6	48.7	65.5	61.5	46.0	59.0	50.3	72.7	62.0	61.3
Flag if total limit on all retail cards greater than zero	59.6	55.0	60.3	63.3	62.0	73.0	63.0	31.0	76.0	52.3
Total balance on all accounts / total limit	60.5	55.2	66.2	60.2	72.0	56.7	69.0	57.7	53.3	54.3
Flag if greater than 0 retail cards 60 days past due	60.9	68.0	61.8	53.0	68.7	43.3	55.3	63.7	75.7	59.0
Cash advance volume / credit limit	61.7	64.5	64.5	56.2	72.7	47.0	74.0	63.0	59.0	54.7
Total credit limit to number of open retail accounts	67.0	66.7	67.3	67.0	70.7	72.7	74.0	69.7	67.7	47.3
Line increase in current mo. flag (0,1)	67.8	68.7	69.0	65.7	74.7	75.3	57.0	64.7	67.0	68.0
Number of accounts in collection	68.0	66.0	73.3	64.7	69.3	65.3	76.7	65.3	73.7	57.7
Flag if total balance over limit on all open bank cards	68.1	65.8	67.0	71.3	74.7	76.7	70.0	63.7	66.7	56.7
Number of accounts under wage garnishment	68.7	71.2	68.2	66.7	75.7	70.0	66.3	65.3	67.7	67.0



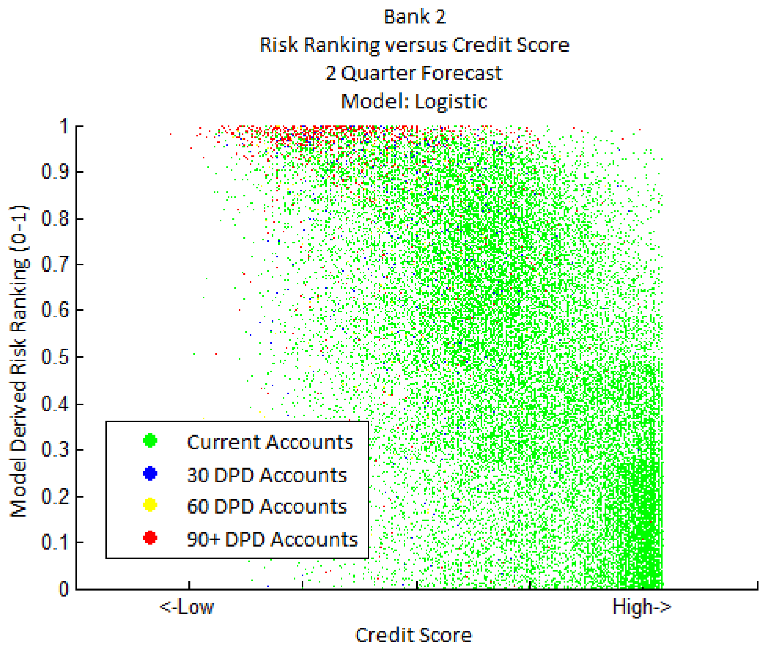
This figure shows the relative delinquency rates over time. Due to data confidentiality restrictions, we do not report the actual delinquency rates over time. Each line represents an individual bank over time. The delinquency rates are all reported relative to the bank with the lowest two quarter delinquency rate in 2010Q4.

Figure 1 – Relative delinquency rates over time

		Model Prediction	
		Good	Bad
Actual Outcome	Good	True Positive (TP)	False Negative (FN)
	Bad	False Positive (FP)	True Negative (TN)

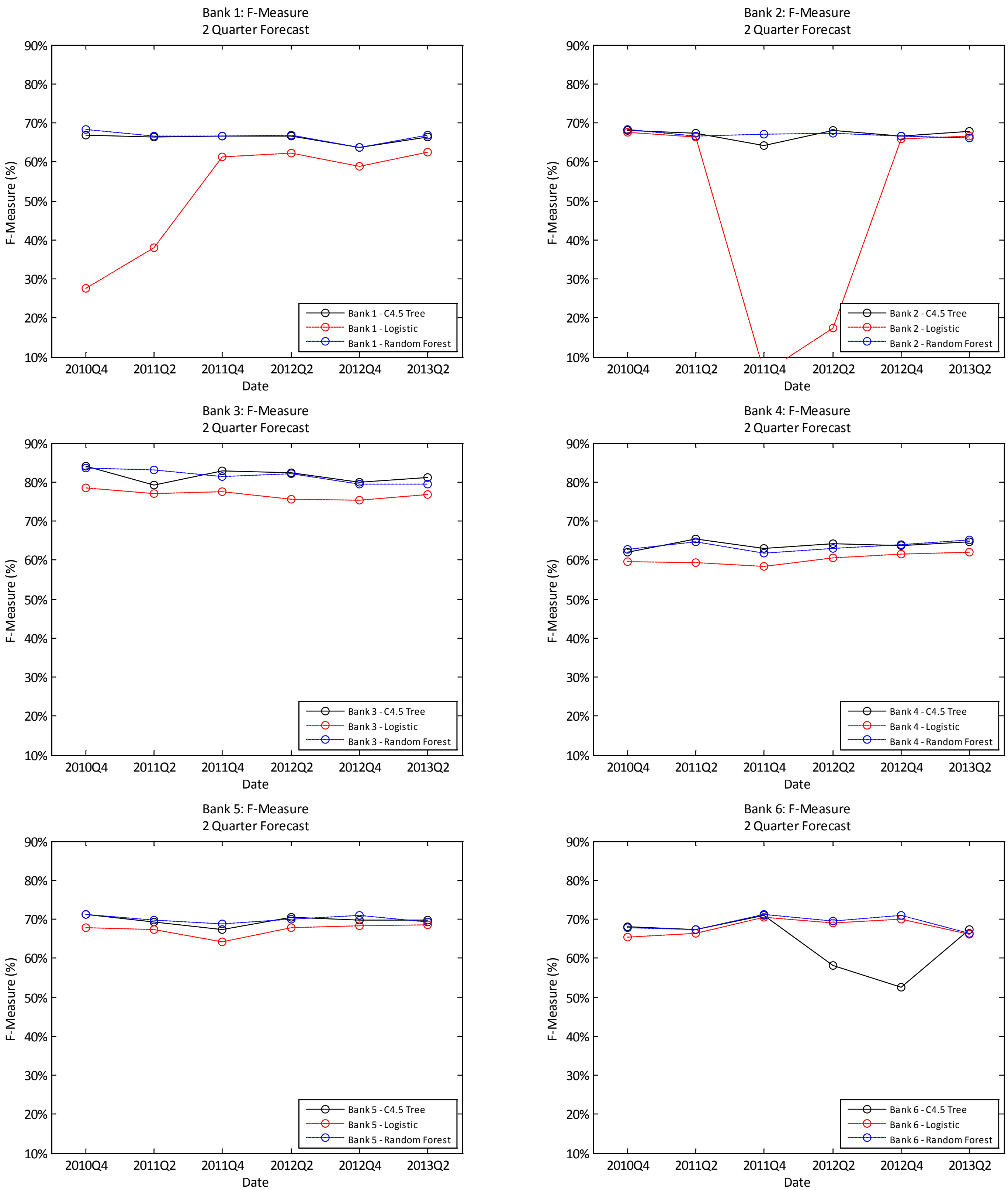
Precision = TN/(TN+FN)
Recall = TN/(TN+FP)
True Positive Rate = TP/(TP+FN)
False Positive Rate = FP/(FP+TN)
F-Measure = (2*Recall*Precision)/(Recall+Precision)
Kappa Statistic = (P_a – P_e)/(1-P_e), where P_a = (TP+TN)/N and P_e = [(TP+FN)/N]*[(TP+FN)/N]
This figure shows a sample confusion matrix and defines our performance statistics.

Figure 2 Performance Statistics



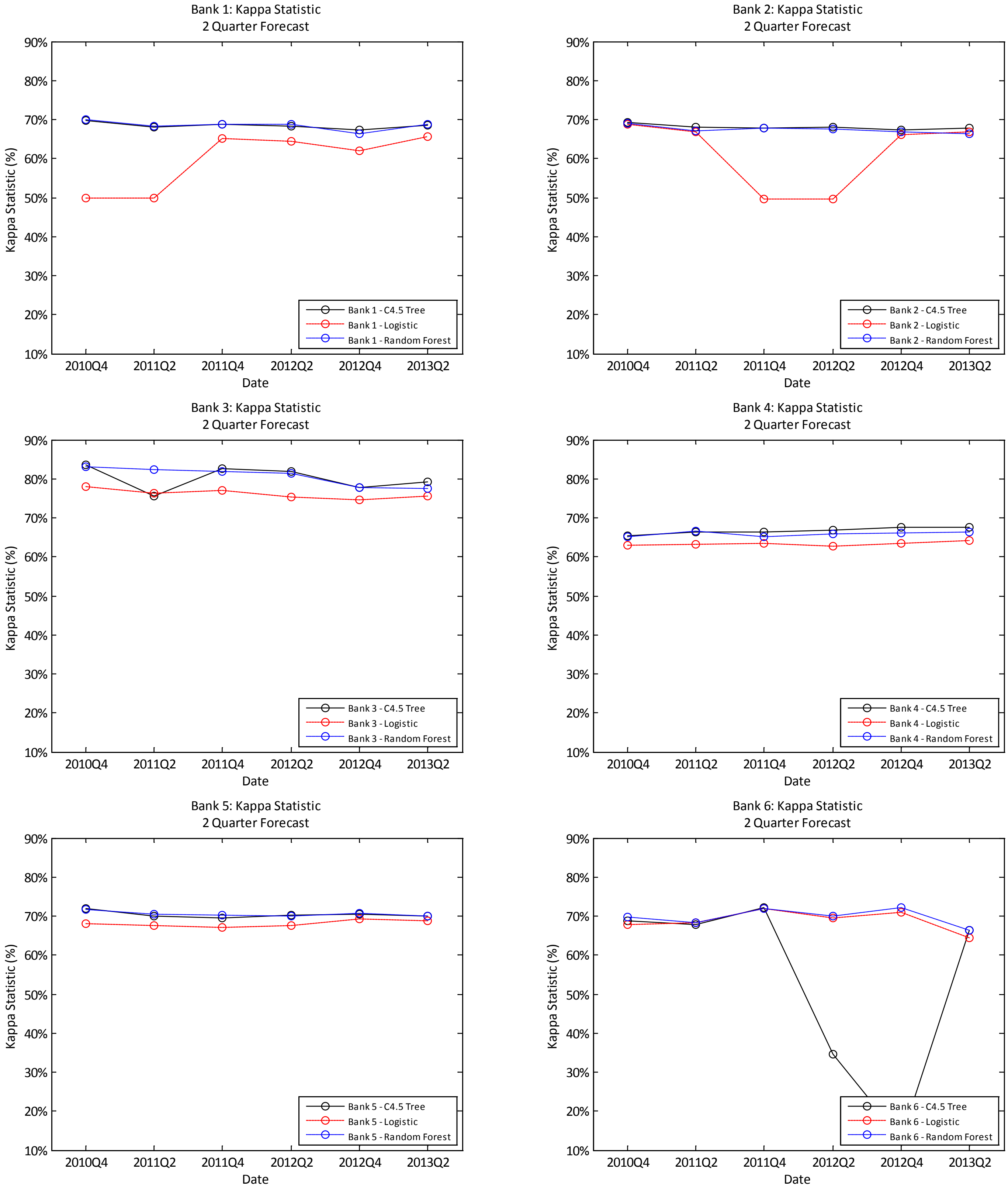
The figure plots the model-derived risk ranking versus an account's credit score at the time of the forecast for Bank 2. Accounts are rank-ordered based on a logistic regression model for a two quarter forecast horizon. Green points are accounts that were current at the end of the forecast horizon; blue points are 30 days past due; yellow points are 60 days past due; and red points are 90+ days past due.

Figure 3- Model Risk Ranking versus Credit Score



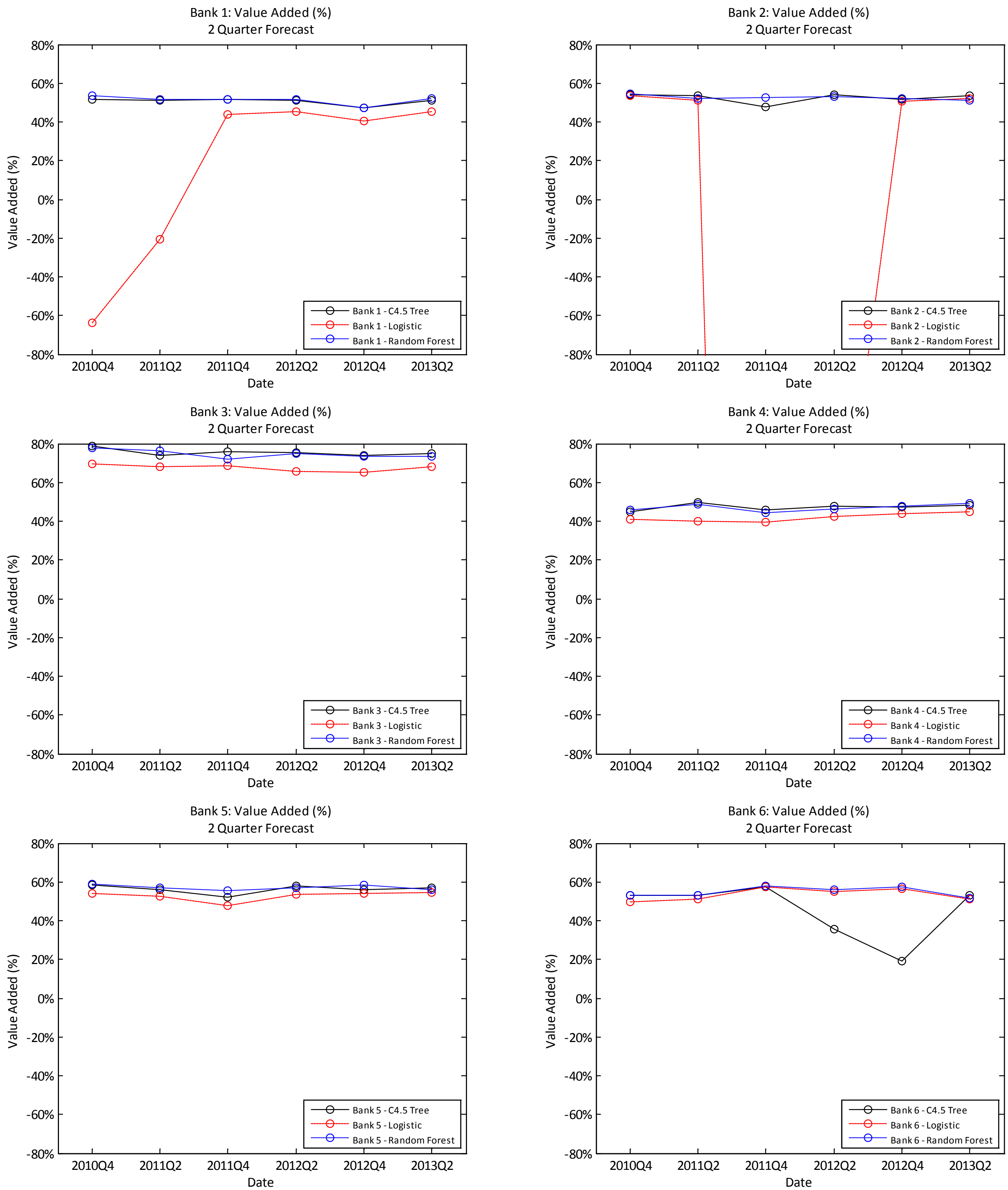
These figures plot the F-Measures for each model over time for each bank. The statistics plotted are for the two quarter horizon forecasts.

Figure 4 F-Measures for each bank and model type over time.



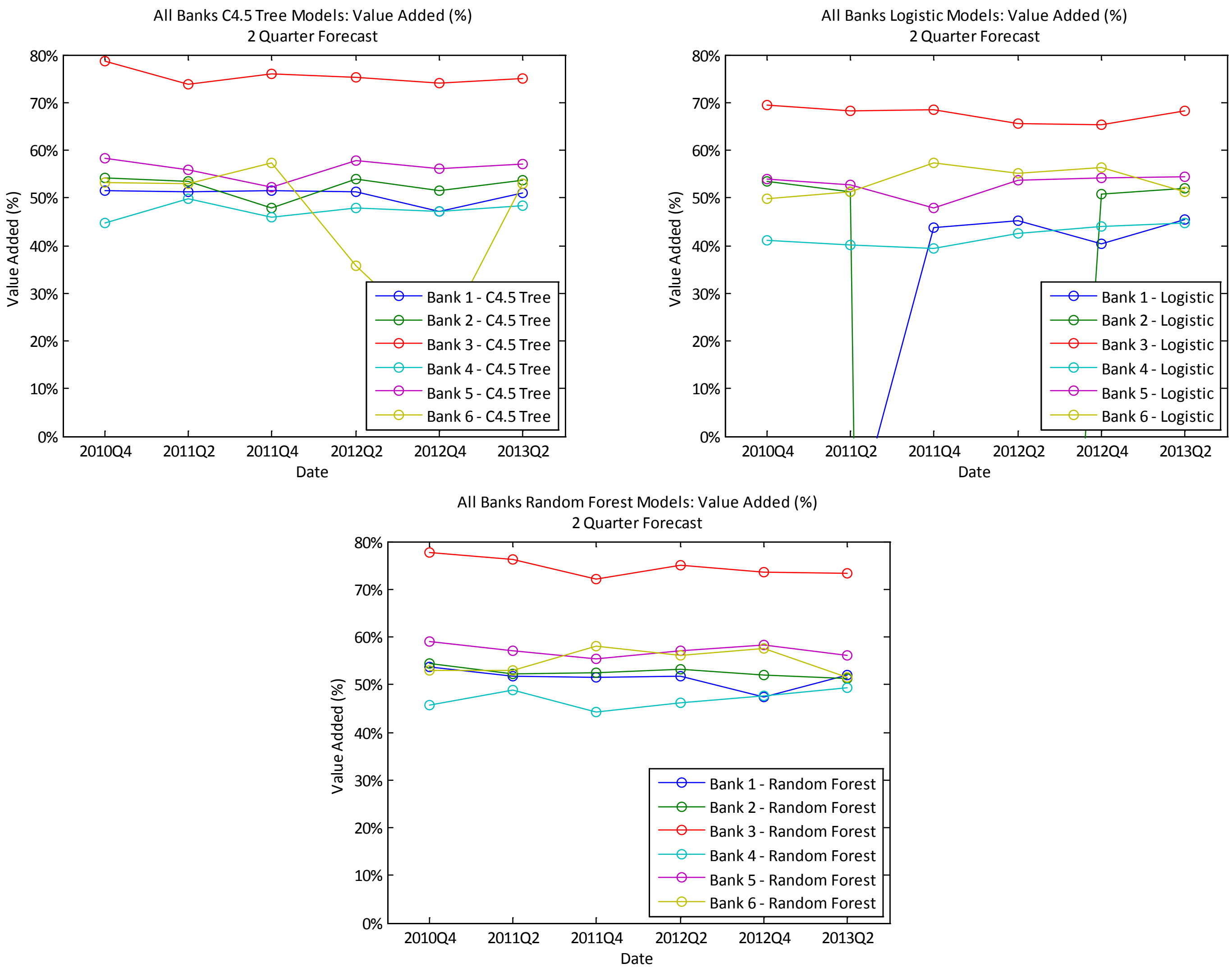
These figures plot the Kappa Statistics for each model over time for each bank. The statistics plotted are for the two quarter horizon forecasts.

Figure 5 - Kappa Statistics for each bank and model type over time.



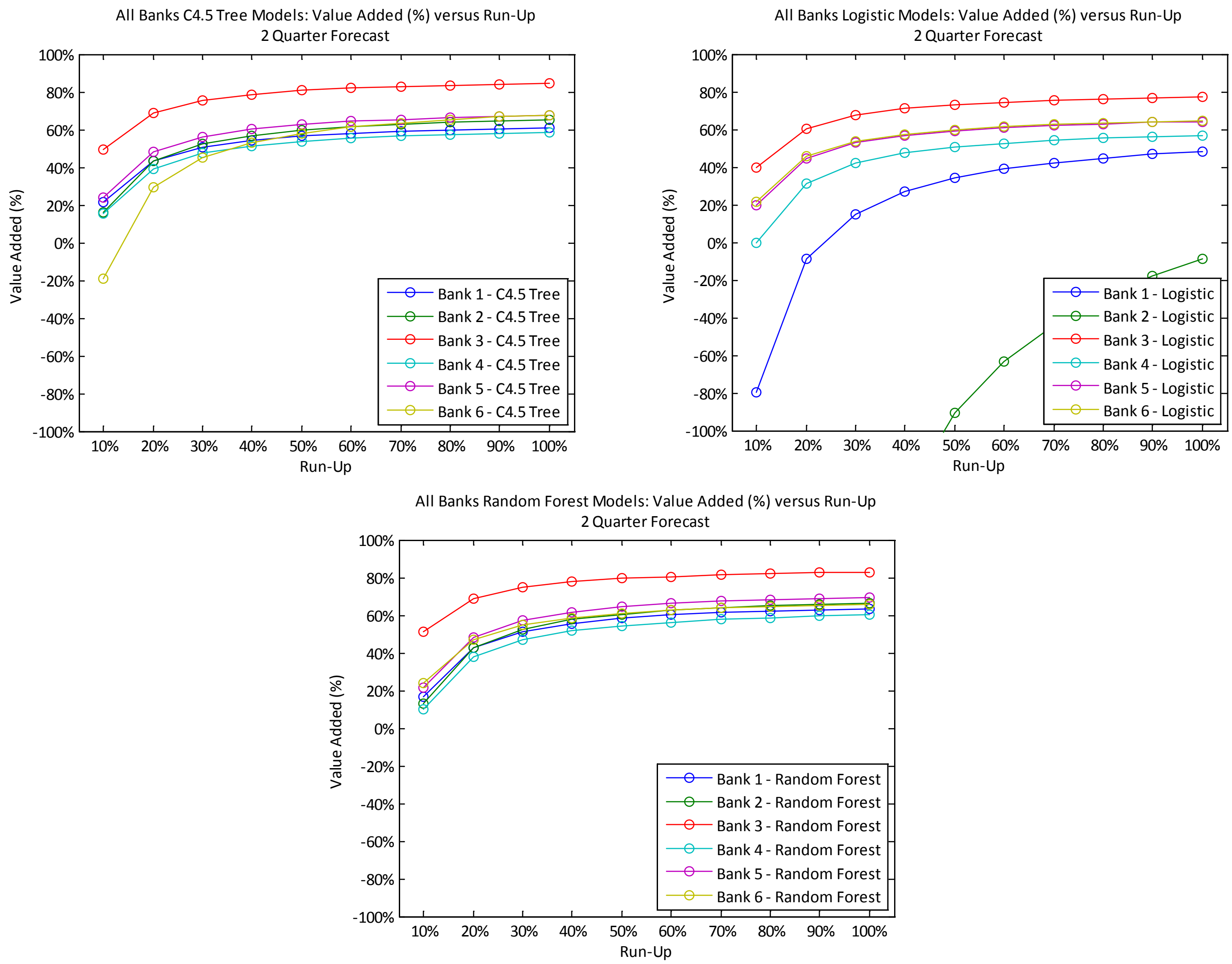
These figures plot the Value Added as defined by Eq. (4) for each model over time for each bank. The statistics plotted are for the two quarter horizon forecasts.

Figure 6- Value Added by Bank and Model



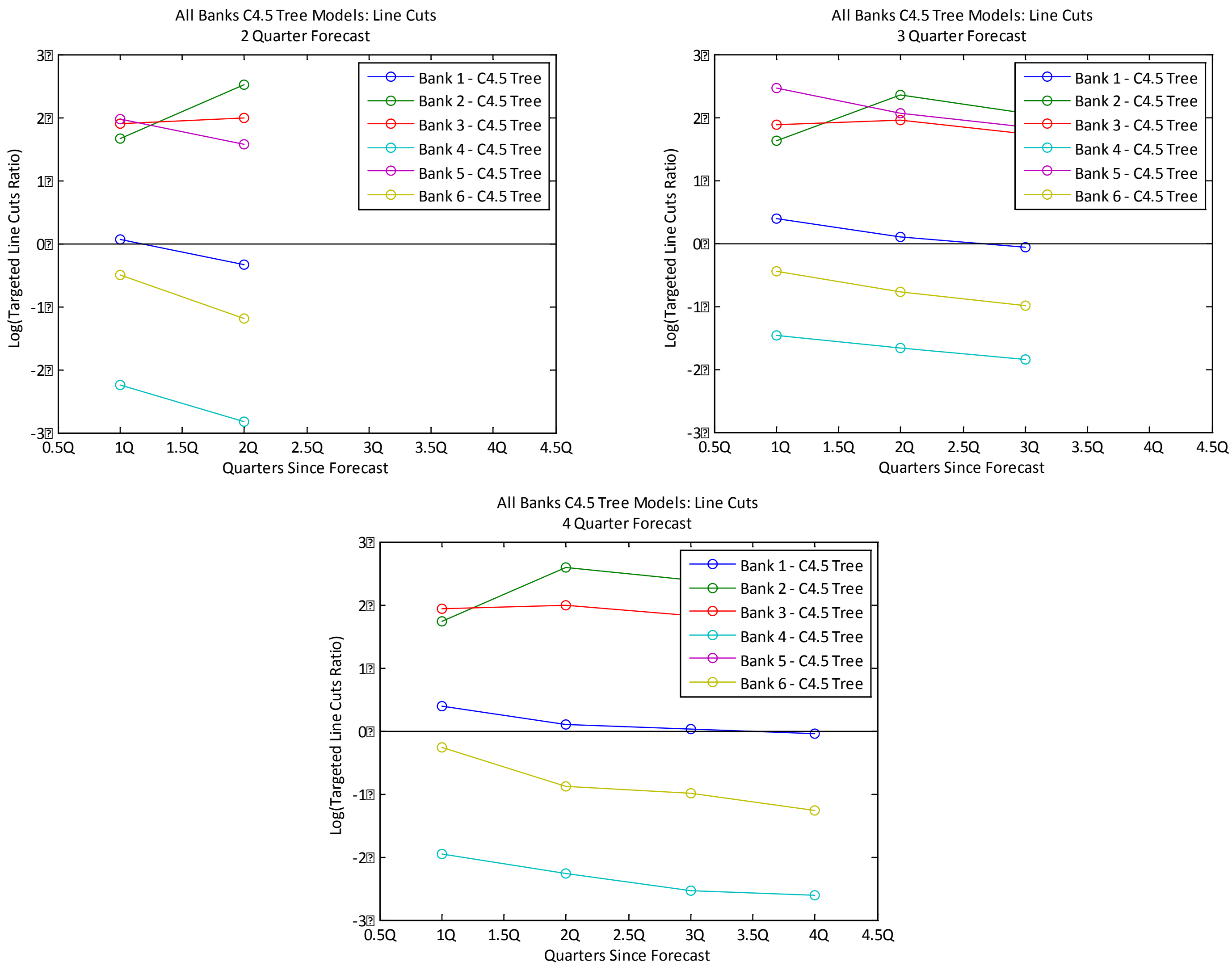
These figures plot the Value Added as defined by Eq. (4) over time. The statistics plotted are for the two quarter horizon forecasts. Clockwise from the top left, the figures show the value added for C4.5 decision trees, logistic regression, and random-forest models. Note the vertical axis is cut off at 0% and the logistic regression models for bank 1 and bank 2 are negative for the first two and third and fourth time periods, respectively.

Figure 7 - Value Added by Model Type



These figures plot the Value Added as defined by Eq. (4) versus run-up. The statistics plotted are for the two quarter horizon forecasts. Clockwise from the top left, the figures show the value added for C4.5 decision trees, logistic regression, and random-forest models. Note the vertical axis is cut off at -100% and the logistic regression models for bank 1, bank 2, and bank 3 are negative for low values of run-up.

Figure 8 - Value Added Versus Run-Up

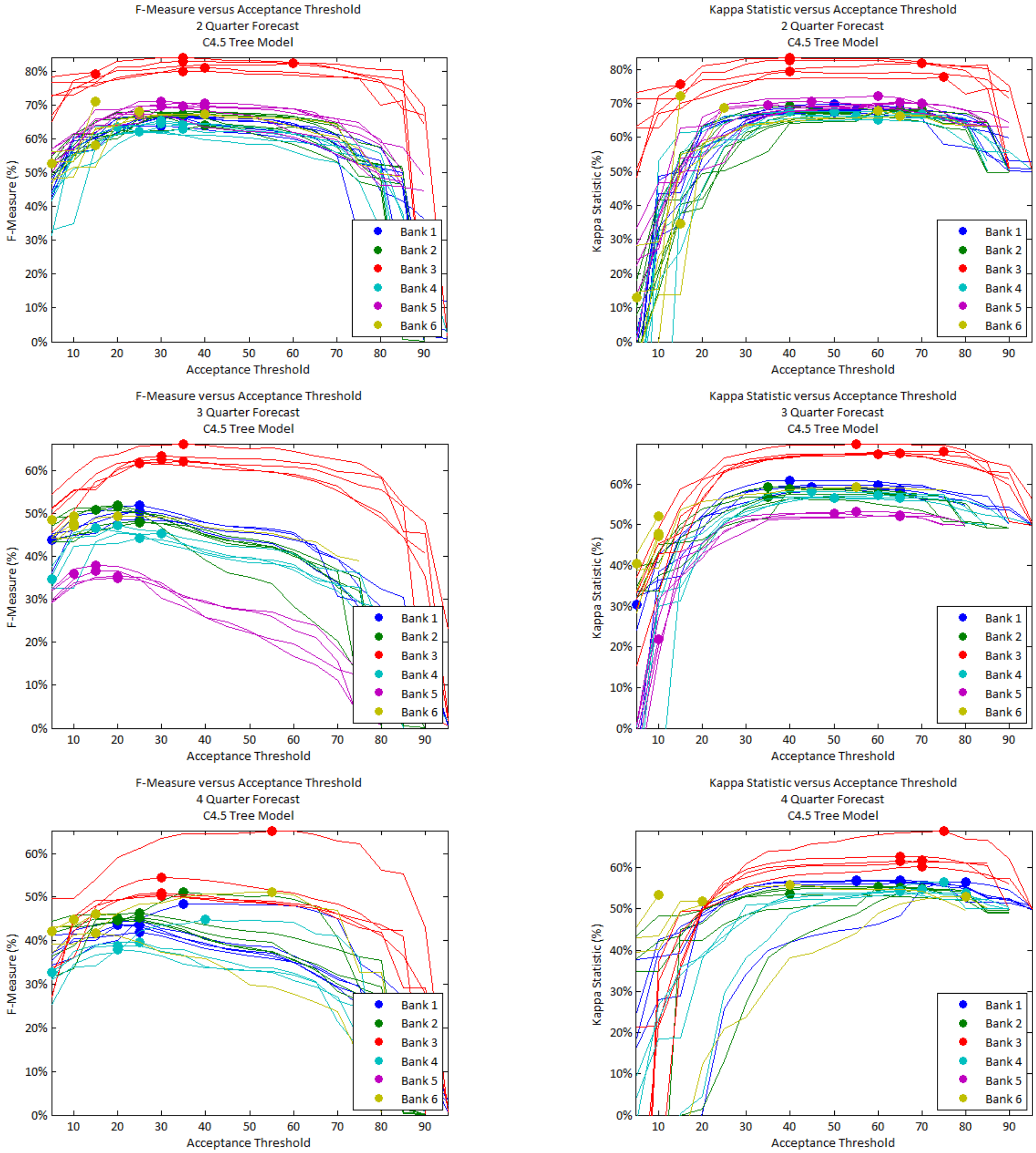


The figures show how well banks target bad accounts and cut their credit lines relative to randomly selecting lines to cut. The targeted line ratio is defined as the percentage of accounts that our models predict to become delinquent whose lines are cut relative to the total percentage of accounts whose lines are cut. A ratio of one (log of zero) means a bank is no more active in cutting credit lines of cards classified as bad than accounts classified as good. Higher ratios signal more active risk management. The ratios for each bank are plotted on a log scale. The plots show the ratios for each quarter following our forecast through the end of the forecast horizon. Clockwise from the top left, the figures show the value added for C4.5 decision trees, logistic regression, and random-forest models.

Figure 9 - Credit Line Cuts

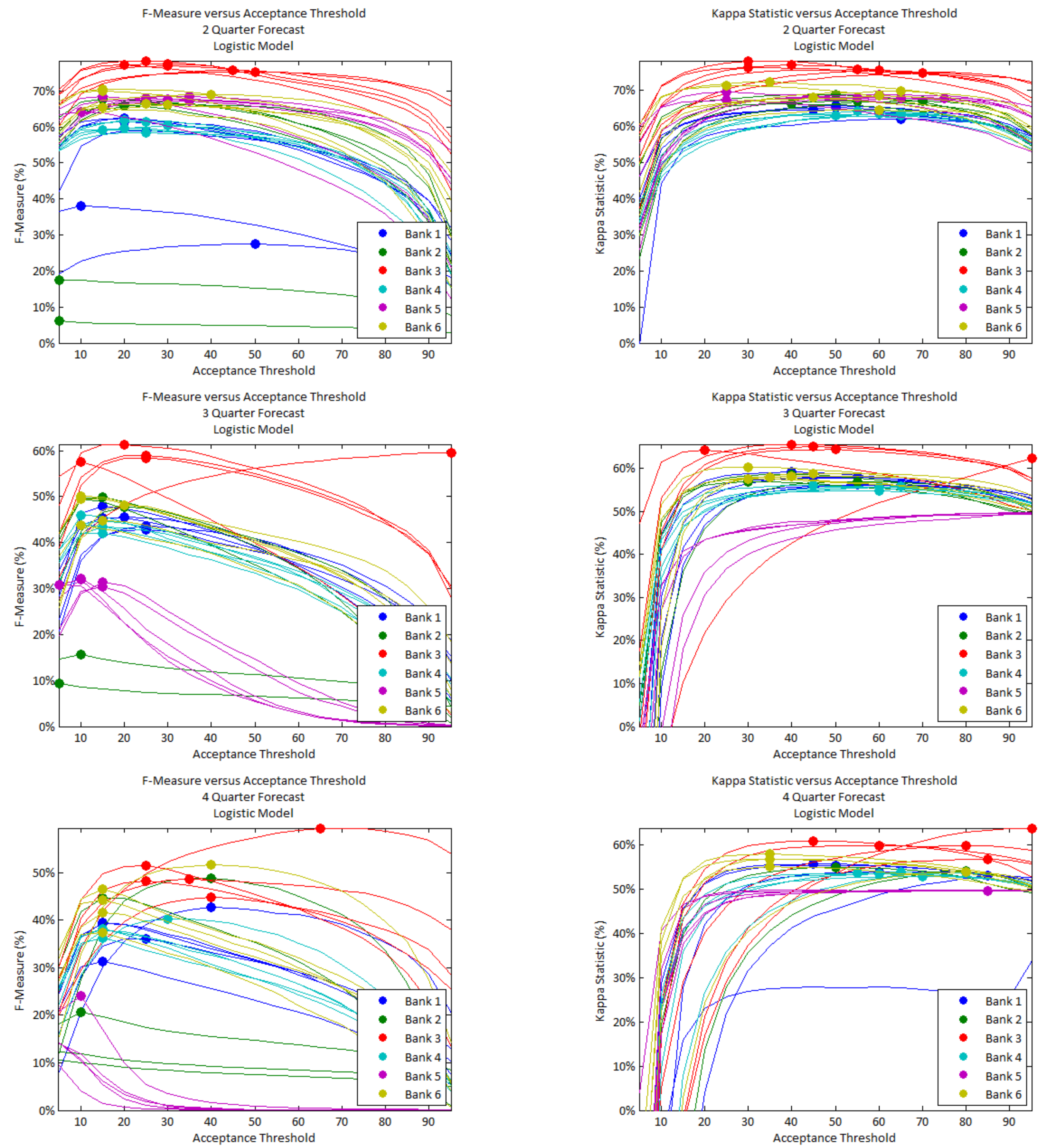
Appendix 1: Variables Descriptions for Tradeline and Attributes Data

Account Level Features:	Credit Bureau Features:	Macroeconomic Features:
Cycle end balance	Flag if greater than 0 accounts 90 days past due	Unemployment rate
Refreshed credit score	Flag if greater than 0 accounts 60 days past due	Unemployment rate (3 mo. chg.)
Behavioral score	Flag if greater than 0 accounts 30 days past due	Unemployment rate (12 mo. chg.)
Current credit limit	Flag if greater than 0 bank cards 60 days past due	Number of total nonfarm (NSA)
Line frozen flag (0,1)	Flag if greater than 0 retail cards 60 days past due	Number of total nonfarm (NSA) (3 mo. chg.)
Line decrease in current mo. flag (0,1)	Flag if total limit on all bank cards greater than zero	Number of total nonfarm (NSA) (12 mo. chg.)
Line increase in current mo. flag (0,1)	Flag if total limit on all retail cards greater than zero	Total private (NSA) (3 mo. chg.)
Actual payment / minimum payment	Flag if greater than 0 accounts opened in the past year	Total private (NSA) (12 mo. chg.)
Days past due	Total number of accounts	Avg. weekly hours worked (private) (3 mo. chg.)
Purchase volume / credit limit	Total balance on all accounts / total limit	Avg. weekly hours worked (private) (12 mo. chg.)
Cash advance volume / credit limit	Total non-mortgage balance / total limit	Avg. hourly wage (private) (3 mo. chg.)
Balance transfer volume / credit limit	Total number of accounts 60+ days past due	Avg. hourly wage (private) (12 mo. chg.)
Flag is the card is securitized	Total number of bank card accounts	Avg. weekly hours worked (trade and transportation) (3 mo. chg.)
chg. in securitization status (1 mo.)	Utilizatiion of all bank card accounts	Avg. weekly hours worked (trade and transportation) (12 mo. chg.)
Percent chg. in credit limit (lagged 1 mo.)	Number of accounts 30+ days past due	Avg. hourly wage (trade and transportation) (3 mo. chg.)
Percent chg. in credit limit current 1 mo.)	Number of accounts 60+ days past due	Avg. hourly wage (trade and transportation) (12 mo. chg.)
Total fees	Number of accounts 90+ days past due	Avg. weekly hours worked (leisure) (3 mo. chg.)
Workout program flag	Number of accounts under wage garnishment	Avg. weekly hours worked (leisure) (12 mo. chg.)
Line frozen flag (1 mo. lag)	Number of accounts in collection	Avg. hourly wage (leisure) (3 mo. chg.)
Line frozen flag (current mo.)	Number of accounts in charge off status	Avg. hourly wage (leisure) (12 mo. chg.)
Product type	Total balance on all 60+ days past due accounts	House price index
3 mo. chg. in credit score	Total number of acocunts	House price index (3 mo. chg.)
6 mo. chg. in credit score	Total credit limit to number of open bank cards	House price index (12 mo. chg.)
3 mo. chg. in behavioral score	Total credit limit to number of open retail accounts	
6 mo. chg. in behavioral score	Total number of accounts opened in the past year	
mo.ly utilization	Total balance of all revolving accounts / total balance on all accounts	
1 mo. chg. in mo.ly utilization	Flag if total balance over limit on all open bank cards = 0%	
3 mo. chg. in mo.ly utilization	Flag if total balance over limit on all open bank cards = 100%	
6 mo. chg. in mo.ly utilization	Flag if total balance over limit on all open bank cards > 100%	
Cycle utilization		
1 mo. chg. in cycle utilization		
3 mo. chg. in cycle utilization		
Account exceeded the limit in past 3 mo.s (0,1)		
Payment equal minimum payment in past 3 mo.s (0,1)		
6 mo. chg. in cycle utilization		



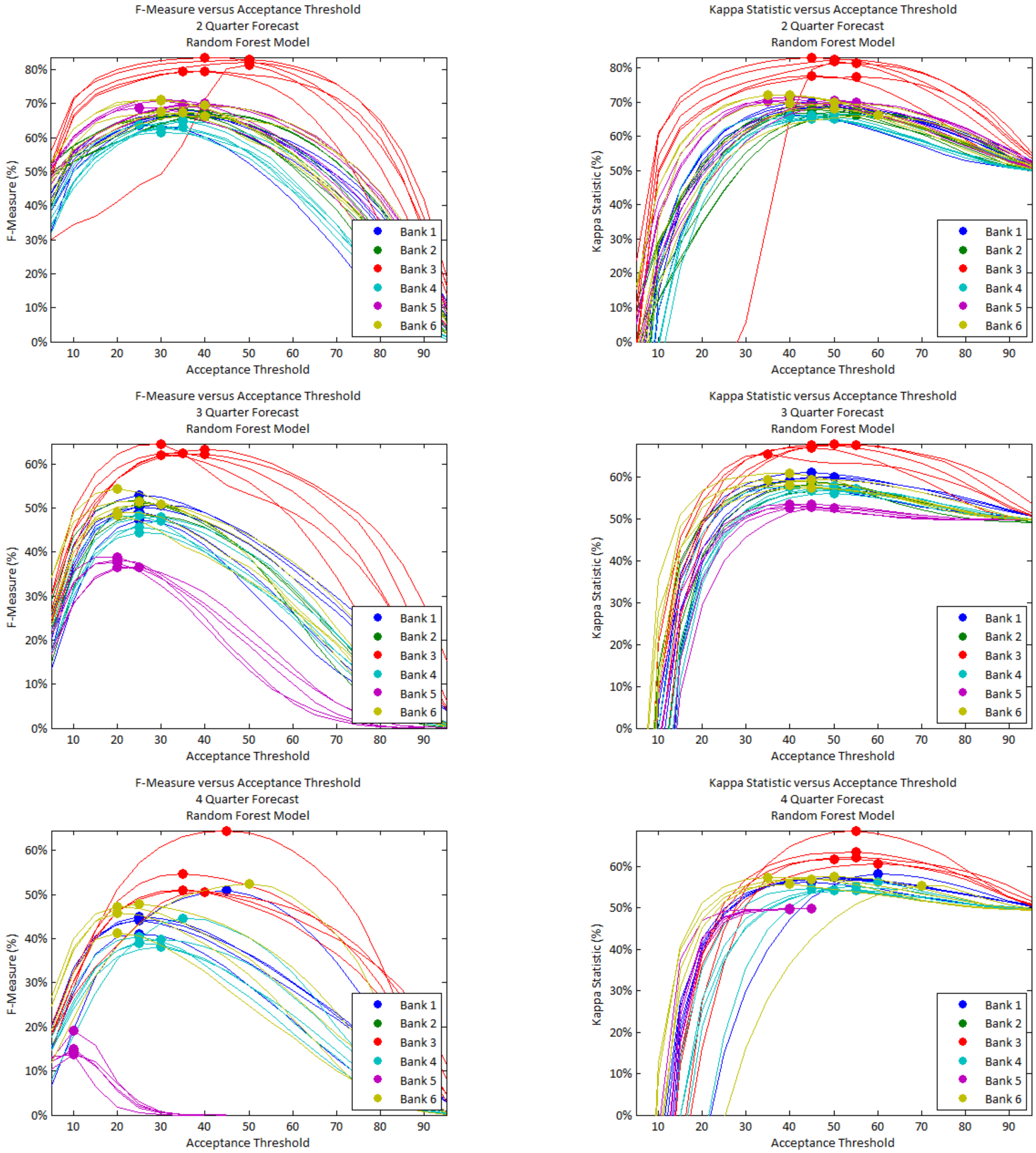
The figures on the left show the F-measure versus the acceptance threshold for each C4.5 model. The figures on the right show the Kappa statistic versus the acceptance threshold. The acceptance threshold is given as a percentage. The dots designate the acceptance threshold that maximizes the respective statistic.

Figure A1 – Sensitivity to Choice of Acceptance Threshold for C4.5 Models



The figures on the left show the F-measure versus the acceptance threshold for each logistic regression model. The figures on the right show the Kappa statistic versus the acceptance threshold. The acceptance threshold is given as a percentage. The dots designate the acceptance threshold that maximizes the respective statistic.

Figure A2 – Sensitivity to Choice of Acceptance Threshold for Logistic Regression Models



The figures on the left show the F-measure versus the acceptance threshold for each random-forest model. The figures on the right show the Kappa statistic versus the acceptance threshold. The acceptance threshold is given as a percentage. The dots designate the acceptance threshold that maximizes the respective statistic.

Figure A3 – Sensitivity to Choice of Acceptance Threshold for Random-forest Models