Expert Systems With Applications 165 (2021) 113954

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Loss rate forecasting framework based on macroeconomic changes: Application to US credit card industry

Sajjad Taghiyeh [a,*], David C. Lengacher [b], Robert B. Handfield [a]

[a] North Carolina State University, Raleigh, NC, USA
[b] Raynolds American, Winston-Salem, NC, USA

ARTICLE INFO

Keywords:
Expert system
Time series forecasting
Loss forecasting
Macroeconomic indicators
Financial industry

ABSTRACT

A major part of the balance sheets of the largest U.S. banks consists of credit card portfolios. Hence, managing the charge-off rates is a vital task for the profitability of the credit card industry. Different macroeconomic conditions affect individuals' behavior in paying down their debts. In this paper, we propose an expert system for loss forecasting in the credit card industry using macroeconomic indicators. We select the indicators based on a thorough review of the literature and experts' opinions covering all aspects of the economy, consumer, business, and government sectors. The state of the art machine learning models are used to develop the proposed expert system framework.

We develop two versions of the forecasting expert system, which utilize different approaches to select between the lags added to each indicator. Among 19 macroeconomic indicators that were used as the input, six were used in the model with optimal lags, and seven indicators were selected by the model using all lags. The features that were selected by each of these models covered all three sectors of the economy. Using the charge-off data for the top 100 US banks ranked by assets from the first quarter of 1985 to the second quarter of 2019, we achieve mean squared error values of 1.15E−03 and 1.04E−03 using the model with optimal lags and the model with all lags, respectively. The proposed expert system gives a holistic view of the economy to the practitioners in the credit card industry and helps them to see the impact of different macroeconomic conditions on their future loss.

1. Introduction

Similar to any industry, the goal in the consumer credit industry is to maximize profits by measuring and controlling risk and avoiding exposure to default (also known as charge-off), as much as possible. The term charge-off means an outstanding credit card debt, which is written off as bad debt. Consumers must issue payments by the due date, and failure to do so will result in putting the consumer's account into delinquency or default. Typically, a bad credit card debt will be marked as charged-off after six months of non-payment, and it is withdrawn as an asset from the lender's accounts. This is usually a final action since it is an indication to lenders that the consumer will never pay off their account. Thus the account is written-off as bad debt. The charge-off rate for a given bank or issuer is calculated by dividing the dollar amount of charge-offs by average outstanding balances on credit cards issued by the firm. A higher charge-off rate exhibits a higher risk to a company. Usually, strategic business analysis is incorporated by credit card issuers to develop credit policy and guidelines with legal and regulatory constraints. Credit policy helps an institution develop strategies within the planned asset quality range that are consistent with the institution's profitability goals. Accurate prediction of charge-off rates has been one of the major challenging tasks in the credit card industry. A forecasting model that over-predicts the charge-off values will lead the company to take credit tightening actions sooner which can lead to a considerable decline in the potential profit. On the other hand, a forecasting model which under-predicts the charge-off rate will delay the credit tightening actions and may result in a significant loss. Hence, it is imperative that the forecasting model predicts the charge-off rate as accurately as possible. The charge-off rate has shown a strong tie to economic conditions, and it has hit its highest level during the financial crisis, which was 10.79% according to U.S. Federal Reserve data. Increasing the charge-off rates during the 2008 financial crisis led to the question of how we can predict the charge-off rate based on macroeconomic indicators under different economic conditions.

There has been extensive research on the relationship between charge-off risk and general economic climate, resulting in a general

* Corresponding author.
E-mail addresses: staghiy@ncsu.edu (S. Taghiyeh), lengacd@rjrt.com (D.C. Lengacher), rbhandfi@ncsu.edu (R.B. Handfield).

https://doi.org/10.1016/j.eswa.2020.113954
Received 6 June 2020; Received in revised form 16 July 2020; Accepted 31 August 2020
Available online 10 September 2020
0957-4174/

belief that macroeconomic factors directly affect bad debts and charge-offs. Historical data obtained from credit bureaus along with consumer performance data are analyzed by lenders to predict the future behavior of consumers and their risk of going delinquent or charging off. These predictive models classify consumers into different segments and align the bank's strategies towards these segments accordingly. The problem is that many businesses rely only on these models to make decisions, and fail to include certain economic factors into their risk models. Sometimes, to include economic conditions, these predictive models are adjusted by several percentage points in the charge-off rate using a fraction of macroeconomic indicators. However, most of the time, only a fraction of economic aspects are reviewed for these adjustments, as they are deemed to be the most influential.

Consumers' charge-off behavior can be heavily affected as the economy goes through good times (expansion phase) and bad times (the contraction phase), and they are not explicitly modeled in prediction models developed by credit risk management, which raises the question of how charge-off rate will change in different economic conditions. During economic expansion, consumers and businesses have enough income to pay their debts by their respective due dates, and thus this phase is associated with a small number of delinquencies and charge-offs. On the other hand, in the contraction phase, the number of bad debts will increase, which eventually will lead to a significant jump in the charge-off rate. Credit card companies can be affected by economic factors, and including economic factors in the decision-making process may significantly impact their ability to make effective charge-off decisions proactively. Failing to incorporate economic factors may lead to consequences that may take years for the company to recover. Since many other factors, such as government regulations, are already reducing the profits of credit card business, there is a need for a new approach that incorporates the relationship between economic factors and charge-off.

Early credit card portfolio literature could not find conclusive evidence on the effects of macroeconomic factors on charge-offs over the business cycle. For example, personal bankruptcy and credit card delinquencies in the 1990s were investigated in Gross and Souleles (2002), and authors concluded that the relationship between charge-offs and macroeconomic factors had changed substantially over the investigation period and there was no conclusive evidence to prove a relationship between charge-off rate and macroeconomic factors. They also concluded that the unemployment rate has no significant impact on the charge-off rate. They used panel data on credit card accounts for their analysis. However, later in Agarwal and Liu (2003), the authors stated that the unemployment rate has significant predictive power for the charge-off rate. They noted that the reason behind the fact that previous empirical studies could not find a consistent relationship between economic factors and bankruptcy is that those studies were either suffering from inadequate data or the variation in the unemployment rate was not sufficient during their analysis period.

Following the Great Recession in 2008, credit card companies focused heavily on controlling credit losses. Their emphasis is mostly on the unemployment rate, as it has a strong correlation with the charge-off rate. However, in the past few years, the unemployment rate was going down while the charge-off rate was increasing, and a model using the unemployment rate as its only input may not be able to capture the uptrend in the charge-off rate. Hence, credit card companies need to focus on other economic factors that can affect charge-offs, and most importantly, they need to look at the economy as a whole. Analyzing the impact of variables from all segments of the economy will provide lenders with a holistic insight and will help them to make more effective decisions to reduce future losses.

There are limited cases in the body of literature that focus on charge-off prediction models incorporating macroeconomic variables in the United States. The slope of U.S. Treasury bond yields over time was mentioned by Estrella and Hardouvelis (1991) and Estrella and Mishkin (1998) to have a strong relationship with output growth and recessions in the United States up to eight quarters in the future. Stock prices (Estrella & Mishkin, 1998), credit market activity (Levanon et al., 2011), index of leading economic indicator (Berge & Jordà, 2011; Stock & Watson, 2002) and several interest rates, housing indices and unemployment rate measures (Ng, 2014) as leading indicators for future economic conditions. Moreover, there are different views regarding the significance of specific economic factors. For instance, industrial production was found to be a significant predictor of corporate charge-offs by Figlewski et al. (2012). However, research was done by Giesecke et al. (2011) has shown that it may not be an important factor in forecasting the charge-off rate. Stochastic and fuzzy optimization algorithms can also be used in the financial industry to improve the efficiency of the algorithms (Mokhtarimousavi et al., 2019, 2018; Rosen et al., 2016; Taghiyeh et al., 2020b; Taghiyeh & Xu, 2016).

The author was motivated to perform this study when he started working as an analyst at one of the leading credit card issuer companies in the United States. The models in production were using only the unemployment rate as their input to forecast future values of the loss rate, which had an R-squared value of about 63%. Aside from the relatively low R-squared value, the charge-off rate was going up in the past couple of years, but the unemployment rate was going down. Therefore, their model was unable to predict uptrend in the charge-off rate, and it was crucial to develop a new prediction model for the charge-off rate by incorporating macroeconomic factors from all aspects of the economy.

In this study, we sought to identify and analyze economic indicators that have a significant relationship with the charge-off rate in the credit card industry. Next, we will use machine learning techniques, namely, linear regression with Lasso, linear regression with Ridge, random forest, and gradient boosting machine to develop a loss forecasting framework using selected macroeconomic indicators. Finally, using the model selection approach introduced in Taghiyeh et al. (2020a) (MSIC algorithm), we will forecast each of the selected indicators to predict year over year changes. Nineteen macroeconomic indicators from three major economic categories will be used for this analysis. These economic categories include consumer, business, and government segments. The use of indicators from all segments gives a comprehensive view of the economic impact on charge-offs. Credit card companies have recently identified the unemployment rate and housing indices as charge-off accelerators. These two metrics will be included in our analysis to confirm or deny their assumptions. The consumer confidence index is another factor that can be seen to have an impact on charge-off rates, as consumer behavior may change payment behavior when they are optimistic or pessimistic towards the future. However, this index is very volatile and may fluctuate each month as the report comes out (Censky, 2010). Other macroeconomic indicators used in this research are new from a charge-off analysis standpoint. Charge-off data from the top 100 banks in the United States from 1985 to 2019 were used in this study to confirm if the selected macroeconomic indicators have significant predictive power for the duration of the analysis. The design of a prediction model covering all aspects of the economy will add significant value to financial institutions. In 2019, the total revolving credit in the United States was an average of $1076.2 billion (Board of Governors of the Federal Reserve System (US), 2020a). Taking into account the 3.63% charge-off rate in 2019 (Board of Governors of the Federal Reserve System (U.S.), 2020), the total amount of charge-offs in the United States was $39 billion. Even a 1% decrease in the charge-off rate, which is indeed possible using an accurate forecasting model, could lead to $390 million in savings for the US economy. Executives and managers can incorporate this information into their decision process to anticipate any future credit losses and fluctuations and modify their policies to avoid such upcoming losses.

The remainder of this paper is organized as follows. Section 2 reviews the literature on loss forecasting. Section 3 presents the details of our proposed loss forecasting framework based on macroeconomic indicators. Section 4 presents an empirical evaluation of the approach using loss data from the top 100 banks in the United States. We summarize our conclusions and discuss the practical implications of our work in Section 5.

2. Literature review

To the best of our knowledge, there is only one study that examines the relationship between economic factors and the charge-off rate in the US economy (Liu & Xu, 2003). In an empirical study by Liu and Xu (2003), the authors use step-wise regression and vector autoregression to identify economic factors which demonstrate prediction of credit card charge-offs. The goal of their research was to develop a predictive model based on these variables. The authors concluded that the unemployment rate, consumer confidence index, household debt service burden, inflation rate, personal bankruptcy filings, and stock market returns are the variables that are useful in predicting the charge-off rate. However, there are a few issues with their work that justifies the need for a more recent and thorough analysis of economic variables for developing a predictive model for charge-offs. The first issue is that their analysis is focused on the period of 1986–1998, and there have been quite a few changes in the credit card industry structure, as well as global economic conditions. Second, Liu and Xu (2003), only include seven economic variables in their analysis, and they are not covering all economic dimensions spanning government, business, and household spending.

There exist several studies on the relationship between charge-offs and economic conditions. Ausubel (1997) noted that in a generally healthy economy, in which unemployment is relatively low and gross domestic product is growing reasonably, both bankruptcy and charge-off rates increased. This statement was made against the foundational belief that the charge-off rate will increase during depressed economic periods and decrease in robust economic periods, but more recent research has shown that other economic factors may contribute to charge-offs. For instance, debt-to-disposable income ratio was found by Stavins et al. (2000) to have a strong correlation with credit card charge-offs and bankruptcy rates.

The unemployment rate, consumer price index and the number of bankruptcy filings were deemed to be highly correlated with the charge-off rate in the case of Hong Kong (Fung & Wong, 2002). The authors used a vector regression model as the basis for their analysis. Macroeconomic indicators were analyzed by Agarwal and Liu (2003) to investigate rates of credit card delinquency. The authors concluded that macroeconomic fluctuations correlated with bankruptcy and delinquency rates. They also found that the unemployment rate has a strong effect on the rate of delinquency. By applying portfolio theory to consumer lending, Desai et al. (2014) extend the work of Musto and Souleles (2006). The authors used credit scores along with charge-off and bankruptcy rates to predict charge-offs. These authors also state that the effect of fluctuations in housing prices is not homogeneous across the population, and people with low credit scores, who highly leverage their credit, are more sensitive to these changes in housing prices.

A regime-switching model was employed by Giesecke et al. (2011) to evaluate the predictive power of macroeconomic variables for charge-off rates. Changes in the gross domestic product, stock returns, and their volatility were identified as significant variables. Reduced-form Cox intensity models were fit by Figlewski et al. (2012) to analyze the relationship between a range of macroeconomic and firm-specific factors and charge-off and significant credit ratings. They found that both factor categories were significant, but macroeconomic variables were highly dependent on the inclusion of other factors. In an extension to their work, in Bellotti and Crook (2013), a discrete time survival model was proposed to predict the probability of charge-off. They claim that using macroeconomic variables along with behavioral factors could produce the best predictive fit. Borrowers' characteristic were also found in Leow and Crook (2014) to impact charge-off and recovery behavior significantly. In the study of Rubaszek and Serwa (2014), interest rate spread, and income uncertainty were found to impact the amount of household credit using both theoretical and empirical models. The relationship between the age of the borrower and the probability of charge-offs in the US was investigated by Debbaut et al. (2016). The authors concluded that the probability of charge-off is lower in younger borrowers. Using macroeconomic indicators in Turkey, Mazibaş and Tuna (2017) analyzed the reasons behind recent fluctuations in household debt. In a study performed in Korea, Kim et al. (2017) used account-level credit data to find a positive relationship between the probability of delinquency and the amount of debt.

As the literature review performed in this section suggests, the basis of this research is supported by scholars in the field. As we can see, most of the prior research shows that macroeconomic factors affect lenders and financial institutions, and by studying the effects of macroeconomic indicators, we predict a more accurate perception of future lending risks. It is essential for credit card companies to incorporate macroeconomic indicators in their risk models to predict future risks and operate effectively in both the expansion and contraction phases of the economy. This way, they can avoid any unnecessary loss in their portfolio due to a lack of perspective towards economic conditions. Several economic factors were studied in previous research studies regarding the charge-off rate. However, in this study, we will primarily cover economic indicators that encompass all segments of the economy, namely households, government, and business segments. Credit card issuers suffer from unexpected charge-offs due to lack of insight from economic conditions, and a charge-off prediction model which is based on macro-economy data will help managers to make effective business and strategic decisions. This research seeks to fill this gap and find the economic indicators with the most significant power to predict future charge-off rates and will use these indicators to build a loss forecasting model for predicting the charge-off rate using machine learning models.

3. Methodology

In this section, we will use machine learning tools to develop the loss forecasting framework. First, we will explain the important trade-off between interpretability and accuracy that is a hot topic when it comes to using machine learning models, and we will discuss the reason behind the selection of machine learning models in our proposed loss forecasting framework.

3.1. Interpretability vs. accuracy

Since machine learning models have been shown to be useful in improving forecasting accuracy, they were widely used before sensible interpretability measures were developed. A common issue with these models is that they operate as black-box algorithms to some extent. There exists a long history of discussion regarding the trade-off between interpretability and accuracy. This dilemma has led to two approaches (Krishnan, 2019):

- Accurate models but black-box: The most accurate classification models are typically the black-box models such as neural networks, random forest, gradient boosting machine or an ensemble of these models (Duda et al., 2012; Saeys et al., 2007; Scholkopf & Smola, 2001). These models are often referred to as black-box models as they are typically non-parametric or have very large number of parameters (Fernández-Delgado et al., 2014), and a common complaint is that the inner workings of these models are hard to understand. Most of these models just generate a value or probability as their output without providing a means of logically interpreting the prediction.
- White-box models but relatively weak: At the other end of the spectrum are the models which are easy to understand and provide illustrative graphs and parameters, such as linear regression (Tibshirani, 1996) and decision trees (Quinlan, 1986). However, these models are inflexible and often provide lower accuracy as compared to the black-box models.

It is true that the understandability and interpretability of a model is vital in science and industry. Scientists and researchers need to understand the model to be able to trust the results as they may influence their work or even their lives. However, the existing tension is motivating analysts and data scientists to seek to develop more complex though less transparent predictive models and machine learning algorithms. Some of the most popular variants of these algorithms which have proven to be accurate and effective include artificial neural networks, gradient boosting machines, and random forest. As mentioned earlier, these models are labeled as black-box algorithms as their inner-workings are not as obvious as traditional regression models. The same reason that makes them more accurate is the reason behind the lack of interpretability in the predictions of these models, which are the complex non-linear methods behind them. Hence, we are faced with a trade-off as the black-box models are typically more accurate in predicting non-linear phenomena, but they come at the expense of lack of interpretability.

When it comes to commercial applications, such as financial problems and credit scoring, it is useful to be able to use interpretable models which let the user analyze the functional form of the model. Considering the degree of interpretability, one may categorize the regression and classification models into the following three groups (Hall & Gill, 2018):

- High interpretability (monotonic and linear models): This class of models include the ones that incorporate linear and monotonic functions, such as traditional linear regression algorithms, which gives them the ability to be explanatory and highly interpretable. Any change in the input variable will lead to a defined rate of change in the response function. This change will be in a specified direction and by a defined magnitude, which is given by coefficients of the corresponding input variables. Due to the monotonic nature of these models, users are able to use intuition and reasoning to explain the predictions made by them. Suppose that a credit application is denied by a creditor and customer asks for the reasons behind this decision. If they have used a linear regression model as the basis for their decision, they will be able to provide an explanation as their charge-off prediction model often uses factors such as credit score, length of credit history and account balance as input and they are monotonically related to the person's ability to pay off any future debt. Most of the time, these explanations are generated automatically after a decision is made by the model.
- Medium interpretability (monotonic and nonlinear models): Despite the fact that most machine learning models use nonlinear functions, some of them provide the ability to add constraints to force monotonicity with respect to each input variable. However, they will not include any coefficients representing the change in the output variable while changing a specific independent variable. The only difference is that the added constraints will ensure that the response variable will change in only one direction when a single input variable changes. These models are able to provide a level of reasoning behind the prediction and will provide a feature importance list as part of their output. There is also another class of machine learning models, such as the multivariate adaptive regression splines (Friedman, 1991), which are linear but nonmonotonic. These models are less interpretable than their monotonic counterparts and often less accurate than the nonlinear machine learning models, and hence, they are less popular in the machine learning field.
- Low interpretability (nonmonotonic and nonlinear models): This is the case for the majority of existing machine learning models, in which the response functions are created using nonmonotonic and nonlinear functions. Since the changes in output variables are in both a positive and negative direction with a non-fixed rate for a change in any input variable, the relative importance measure is often the only interpretability measure that is provided by these models. However, due to a relatively higher accuracy associated with these models, they have attracted a lot of attentions from researchers and practitioners.

As mentioned earlier, nonmonotonic and nonlinear machine learning models can provide a list of feature importance that gives a level of interpretability to these models, and estimating the influence of each input variable to the prediction of the model is a question of interest to researchers. Being able to understand the importance of each feature will help the user understand the model and obtain a level of trust for predictions made by the model. There exists a body of research focusing on the estimation of the associated feature importance of each variable (Adebayo et al., 2018; Selvaraju et al., 2017; Zintgraf et al., 2017). Many of these estimation methods have interesting theoretical properties, such as preservation of relevance (Bach et al., 2015) or implementation invariance (Sundararajan et al., 2017).

In this research, we will employ machine learning models from both sides of the spectrum (black-box with low interpretability and white-box with high interpretability) to evaluate their accuracy and compare their performance to be able to make an educated decision about the trade-off between accuracy and interpretability. Our goal is to achieve an acceptable level of accuracy while maintaining the interpretability of the prediction as much as possible. Hence, first we will train and evaluate the results of our benchmark machine learning models, and when it comes to selecting the best performing model, our priority is to select the model with the highest interpretability if there is not a significant loss of accuracy. The machine learning models that we will use are linear regression with lasso, linear regression with ridge, gradient boosting machine and random forest. The first two belong to the class of monotonic and linear models, which are highly interpretable and provide coefficients for each input variable. The third and fourth model (gradient boosting machine and random forest), belong to the family of nonlinear and nonmonotonic machine learning models and provide a low level of interpretability. We will use the H2O library in R (Nykodym et al., 2016) to implement these machine learning models, in which the feature importance lists are generated using the specific methods tailored for each of the machine learning models.

Based on the literature review, several macroeconomic indicators that were likely to have correlations with the charge-off rate were selected. Among the selected macroeconomic indicators, 19 indicators were selected by the experts in the credit card industry to form the basis of this research. The goal is to use these indicators as independent variables in a machine learning based model to predict the charge-off rate, which is our dependent variable. In the first step, we apply different transformations (e.g., square root, exponential, ...) to normalize the selected indicators and find the transformation with the highest correlation to dependent variables. We also add lags from 1 to 4 quarters to each indicator and find the correlation of each of the lagged indicators with charge-off rate. This way, we incorporate the lagged effects of each macroeconomic indicator. The next step in data preparation is to convert all indicators and charge-off rate to year over year changes. To do so, for each indicator, we record the percentage of change comparing to the corresponding period in the last year. This way, instead of using the actual values for macroeconomic indicators to predict the charge-off rate, we build a model that uses the changes in each indicator to forecast the change in charge-off rate. The data transformation, correlation analysis and converting to year over year changes steps that followed our data preparation step are standard approaches in statistical analysis and in the machine learning field. Moreover, we have taken one additional step in our data preparation process, which is adding lags to each indicator. Using multiple lags can help the model to capture the lagged effect of each indicator on the value of charge-off rates. For example, it takes one to two quarters between the time a spike is observed in unemployment insurance claims until the time that the charge-off rate increases. However, other algorithms that do not use lagged indicators are not able to capture this lagged effect.

After we have generated our input data, we used two versions of Lasso regression (Lasso with optimal lags and Lasso with all lags) to select the features with the most significant correlation to our output data. The difference between these two feature selection methods lies in the approach we use to generate their input. In the first feature selection model (Lasso with optimal lags), for each indicator, we select the lag, which has the highest correlation with the charge-off rate. Therefore, the model has 19 independent variables corresponding to optimal lags for each of the selected macroeconomic indicators. In the second approach, which is Lasso with all lags, we include all the lags in the input data and let the model select between lags. Note that, in the second feature selection method, we let the model choose more than one lag from each indicator. In doing so, the model can capture the trends for the year-over-year changes of each macroeconomic indicator. The advantage of using Lasso for feature selection over similar feature selection methods such as variants of stepwise selection (e.g. forward and backward selection) is that Lasso both improves the forecasting accuracy and selects between the correlated covariates to reduce the chance of overfitting. However, stepwise selection methods only focus on selecting between the correlated covariates without considering the improvement in prediction accuracy. The other advantage of using Lasso for feature selection is that the basis of the algorithm is linear regression, which provides a good level of interpretability to the final output of the model.

We used the indicators selected by each of the feature selection methods as the input to train machine learning models and capture the relationship between the selected macroeconomic indicators and the charge-off rate. As mentioned earlier, the benchmark machine learning models in this study are Lasso regression, Ridge regression, gradient boosting machine (GBM), and random forest (RF). The rationale behind the selection of these four models is to include both black-box (with less interpretability) and regression-based models with high levels of interpretability. Gradient boosting machine and random forest are among black-box models which are able to capture complex nonlinear trends, which consequently may improve their forecasting accuracy. However, it comes at the cost of lower interpretability and a higher chance of overfitting. On the other hand, Lasso and Ridge regression are from the family of linear regression models, which gives them a good level of interpretability and there is a low chance of overfitting. Nonetheless, due to their linear nature, they may not be able to capture very complex trends, which may result in lower forecasting accuracy comparing to black-box models. Using these four models as benchmarks can aid in the selection between both black-box (low interpretability) and white-box (high interpretability) models. All these machine learning models are readily available to use in machine learning packages such as H2O (Nykodym et al., 2016) and Scikit-Learn (Pedregosa et al., 2011). We have used H2O library in R to implement these machine learning models.

As it is common in the machine learning field, we split the data into training and test sets to train and evaluate the performance of each machine learning model. Two sets of machine learning models need to be developed since we have two versions of input data resulted from different feature selection approaches.

The last piece of building the loss forecasting framework is to predict future values for each of the selected macroeconomic indicators and use the trained machine learning model to predict future charge-off levels. To predict each macroeconomic indicator, seven well-known forecasting models have been used, namely, naïve forecasting, moving average, simple exponential smoothing, Holt, Holt-Winters, ARIMA, and Theta. These models are selected among the models considered in the forecasting competitions, such as M3-Competition. Three variants of the MSIC algorithm proposed in Taghiyeh et al. (2020a) are used to select the best performing forecasting model for each macroeconomic indicator. The reason we selected the MSIC algorithm to forecast future values of macroeconomic indicators is that in a prior study by Taghiyeh et al. (2020a), the MSIC algorithm showed a significant improvement over traditional forecasting model selection methods, making it a good candidate to incorporate in our forecasting framework. Using the results from the forecasting model selected by the MSIC algorithm, the trained machine learning models are then used to predict the future values of the charge-off rate. The process that we followed in the forecasting step of our algorithm is a standard procedure in time series forecasting field, (with the exception of using the MSIC algorithm to select between benchmark forecasting models), thus providing us with more accurate forecasts. Fig. 1 shows the steps of the loss forecasting framework proposed in this study. The details of our proposed Loss rate forecasting framework are outlined in Appendix.

In the next section, we will apply the proposed loss forecasting model on the loss rate data from the top 100 banks in the U.S. from 1985 to 2019.

4. Numerical experiments

In this section, we will test the proposed loss forecasting framework on the Charge-off rate data from the first quarter of 1985 to the second quarter of 2019. This data is retrieved from the "Board of Governors of the Federal Reserve System (U.S.)" (Board of Governors of the Federal Reserve System (U.S.), 2020) database and is an aggregated charge-off report for the top 100 US banks ranked by assets. As we mentioned in the introduction section, this study was originally motivated while the author was working for one of the leading credit card issuers in the U.S., and a part of this work was originally developed in that company. However, due to the confidentiality issues, all the data related to the company is omitted in this research, and the equivalent publicly accessed datasets were being used as the basis for the numerical experiments.

To select the macroeconomic indicators for this study, initially, the "Principles for navigating the big debt crises" by Ray Dalio (Dalio, 2018) was reviewed, and the macroeconomic indicators which were mentioned in the book that had a significant correlation with debt, charge-off rate, and economic cycles were selected. Several additional macroeconomic indicators were also added to the list using the research articles reviewed in the literature review section. This list was provided to the experts in the leading credit card company, including a senior manager and a director from the credit risk assessment department. These experts provided their feedback on these indicators and selected 19 indicators that they believed are the ones having the most significant relationship with the charge-off rate and cover all aspects of the economy while having the smallest overlap to reduce the risk of overfitting. The list of selected macroeconomic indicators is shown in Table 1. Please refer to Table A.1 in the Appendix for the list of references corresponding to each macroeconomic indicator.

Figs. 2, 3, 4, and 5 depict the values of the indicators in each segment against charge-off rate. Since each indicator has a different unit, we used the vertical axis to show the charge-off rate, and other indicators were scaled to makes us able to compare their trend against the charge-off rate. The shadowed regions show the U.S. recession periods from 1985 to 2019. As Fig. 2 shows, "unemployment rate" and "initial unemployment insurance claims" have very similar trends to loss rate. That may be the reason that these two indicators are mostly used in the credit card industry to predict the charge-off rate. However, if we look at the values of loss rate, "unemployment rate", and "initial unemployment insurance claims" from the second quarter of 2018 to the second quarter of 2019, we see that "unemployment rate" and "initial unemployment insurance claims" are decreasing, but the loss rate has an increasing trend. Hence, there is no way to predict the loss rate in this period by solely using the "unemployment rate" and "initial unemployment insurance claims" as independent variables. Using macroeconomic indicators from all segments of the economy is

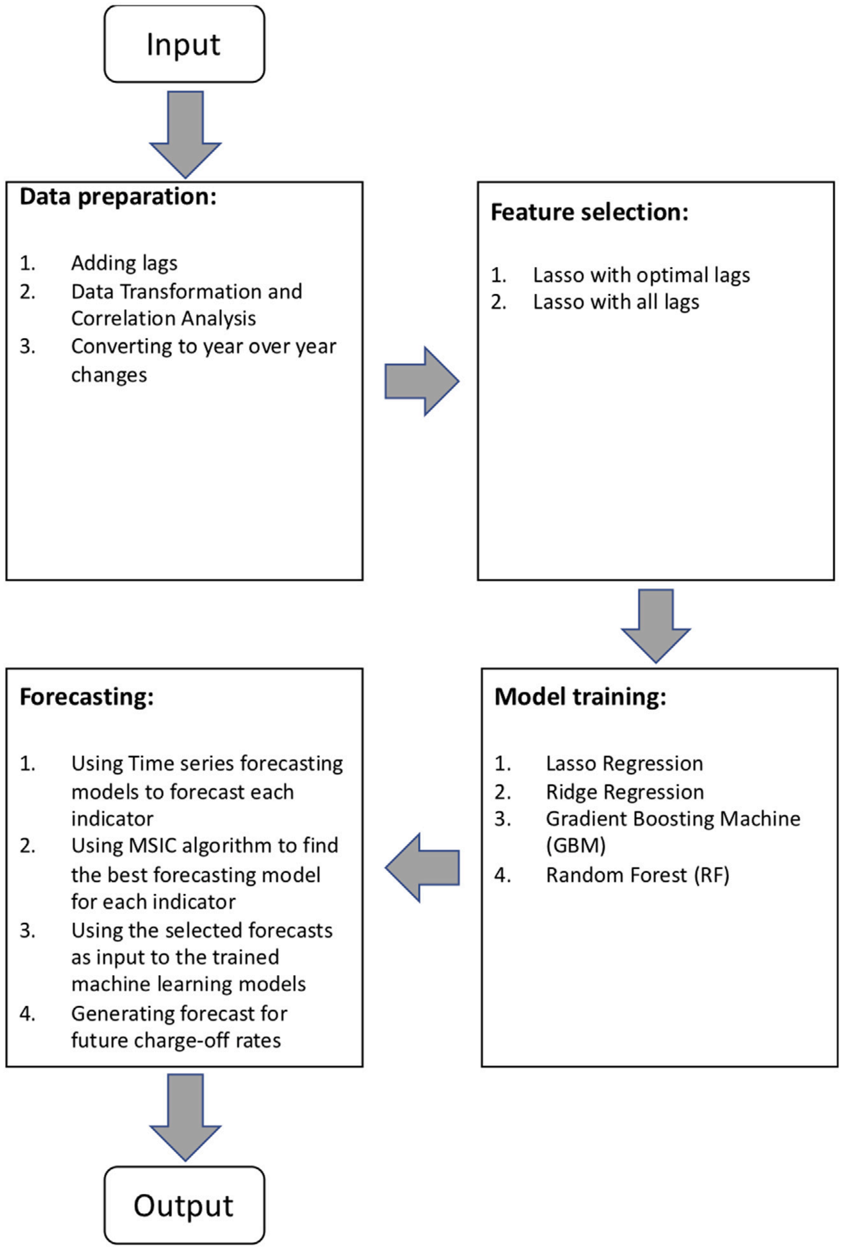

Fig. 1. Steps to develop the proposed loss forecasting framework.

Table 1
List of macroeconomic indicators used in this study for building the loss forecasting framework.

Consumer segment (Part 1)	Consumer segment (Part 2)	Business segment	Government segment
Building permits	S&P 500 index	Industrial production index	M1
Housing starts	Dow Jones industrial average	ISM manufacturing new orders	M2
Initial unemployment insurance claims	Total credit utilization	ISM Purchasing Mangers Index (PMI)	Yield (10 years minus 3 month)
Unemployment rate	Revolving credit utilization	Weekly hours worked by manufacturing workers	Yield (10 years minus federal fund rate)
Consumer Confidence Index (CCI)	Non revolving credit utilization		
University of michigan sentiment index			

one of the advantages of our proposed loss forecasting model, which will make it able to capture the uptrend even when the "unemployment rate" and "initial unemployment insurance claims" are decreasing.

Another interesting fact is that "M1" and "M2" in Fig. 5 have significantly different trends before and after the great recession in 2008. As can be seen, the government started printing money in the great recession to add stimulation to the economy and overcome the recession. However, if we look at the trends, they printed money with a significantly higher rate after the great recession, which may be a negative factor for the economy and could play an important role in our loss forecasting framework when we train the model. Moreover, "building permits" and "housing starts" have a very similar trend in Fig. 2, and to avoid overfitting, only one of them needs to be selected for building a prediction model. The same is true for "CCI" and "U.M. consumer sentiment index" in Fig. 2, "Dow Jones industrial average" and "S&P 500 index" in Fig. 3, and "yield (10 year minus 3 month" and

Table 2
Results of correlation analysis and their statistical significance for economic indicators using different lags.

Indicators	Lag	Correlation	P-Values	Significance at $\alpha = 0.1$
Building permits	1	−0.35	0.0212	Yes
CCI	2	−0.49	0.0466	Yes
Dow Jones industrial average	3	−0.09	0.7321	No
Housing starts	0	0.37	0.8940	No
Industrial production index	0	0.18	0.2500	No
Initial unemployment insurance claims	1	0.75	0.0038	Yes
M1	4	−0.3	0.0659	Yes
M2	4	−0.17	0.0675	Yes
ISM manufacturing new orders	4	−0.46	0.0550	Yes
PMI	4	−0.48	0.0098	Yes
S&P 500 index	3	−0.1	0.6452	No
University of michigan sentiment index	2	−0.48	0.0542	Yes
Weekly hours worked by manufacturing orders	2	−0.53	0.0360	Yes
Yield (10 years minus 3 months)	0	0.39	0.9602	No
Yield (10 years minus federal fund rate)	0	0.39	0.2345	No
Unemployment rate	0	0.52	0.0776	Yes
Total credit utilization	0	−0.49	0.7390	No
Revolving credit utilization	0	−0.48	0.3667	No
Non revolving credit utilization	2	−0.39	0.7898	No

"yield (10 year minus federal fund rate" in Fig. 5. These collinearities will be handled by the feature selection step of our proposed loss forecasting model. We will show the step by step implementation of our proposed loss forecasting framework in the following subsections.

4.1. Data preparation

All the macroeconomic indicators are converted to quarterly values, and the lagged values are recorded (1 to 4 quarters). Hence, for each macroeconomic indicator, we have five columns of input data, and in total, we have 95 input columns for 19 macroeconomic indicators in this study. "bestNormalize" package in R (Peterson, 2017) is used for the normalization of each lagged input. The function "bestNormalize" in the aforementioned package performs several normalization transformations, including the Box–Cox transformation, the Yeo–Johnson transformation, the square-root transformation, log transformation, and arcsinh transformation, and uses the Pearson P test statistic for normality to select the optimal one. After performing the optimal transformation selected by "bestNormalize" function, we convert all the values for macroeconomic indicators and the loss rate to year over year changes by dividing them by corresponding values from the previous year. Now we have the input data ready for feature selection step.

4.2. Feature selection

Using the input data obtained from the data preparation step, we start performing the two versions of our feature selection procedures, "feature selection with optimal lags" and "feature selection with all lags".

4.2.1. Feature selection with optimal lags

To use the feature selection with optimal lags, we first need to find the optimal lag from the input data generated in step 1. We calculated the correlations of lagged values for each macroeconomic indicator and selected the lag with the highest correlation for each one. The results are shown in Table 2. As we can see, "initial unemployment insurance claims" and "unemployment rate" have the highest correlations with the loss rate, which is in line with what we have already seen in Fig. 2. Now we perform Lasso regression on these optimal lags to remove collinearity between variables and select the most significant features among the indicators list in Table 2. We used the feature importance list from the results of Lasso regression and selected the indicators with the importance values greater than 0.2. First, we have selected 0.2 as our cut-off point because in all our numerical experiments there was a significant gap between the variables with importance values greater than 0.2 and lower than 0.2. The second reason was that we sought to improve the level of interpretability to avoiding overfitting by using fewer variables. Hence, we selected a smaller subset of variables with higher importance values which could achieve the same level of statistical significance for the model ($p - value < 0.0001$) comparing to when we were using all the variables selected by the Lasso regression. The results for this feature selection procedure is shown in Table 3.

Table 3
Selected indicators using Lasso regression and optimal lags.

Indicators	Lag	Correlation	Relative importance
Building permits	1	−0.351927	0.84
Initial unemployment insurance claims	1	0.74811298	0.99
M1	4	−0.2975332	0.22
PMI	4	−0.4789845	0.45
Weekly hours worked by manufacturing workers	0	−0.5314578	0.48
Unemployment rate	1	0.52158091	1

Table 4
Selected indicators using Lasso regression and all lags.

Indicators	Lag	Correlation	Relative importance
Building permits	1	−0.351183	0.39
Initial unemployment insurance claims	0	0.7169161	0.93
Initial unemployment insurance claims	2	0.7262981	0.55
Initial unemployment insurance claims	4	0.570129	0.71
M1	1	−0.187398	0.47
M2	4	−0.165263	0.38
PMI	0	−0.431299	1
PMI	2	−0.36955	0.48
Weekly hours worked by manufacturing workers	0	−0.436368	0.47
Unemployment rate	0	0.5215809	0.66
Unemployment rate	4	0.1248201	0.92

Table 5
Summary statistics for machine learning models when using indicators with optimal lags.

	R^2	MSE (Train)	MSE (Validation)
Lasso regression	0.72	1.59E−02	1.90E−02
Ridge regression	0.72	1.14E−02	1.55E−02
Gradient boosting machine	0.77	4.43E−03	7.21E−03
Random forest	0.7	1.60E−02	1.86E−02

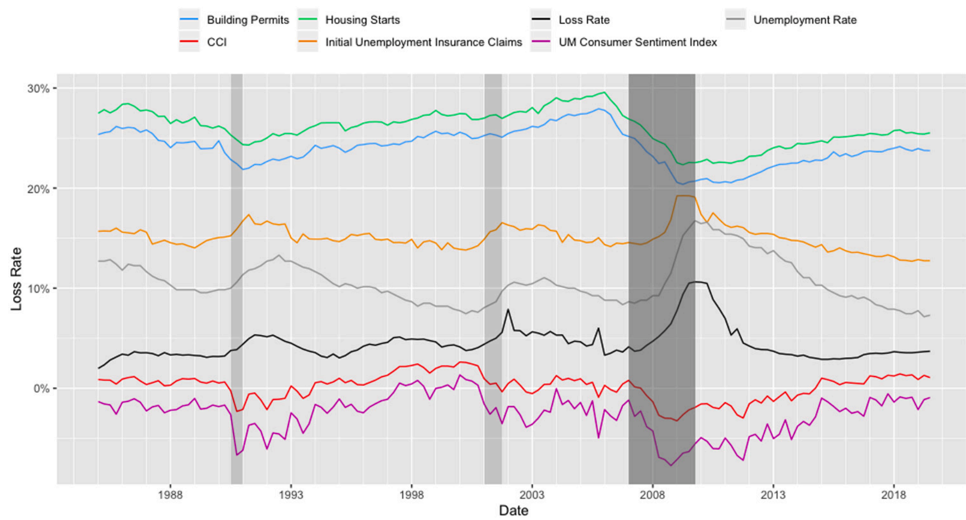

Fig. 2. Consumer related macroeconomic indicators (part 1).

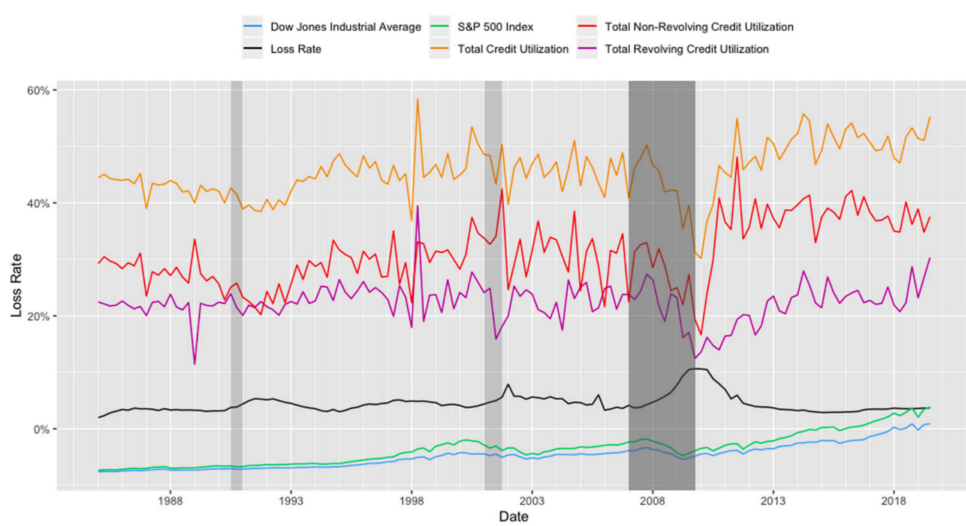

Fig. 3. Consumer related macroeconomic indicators (part 2).

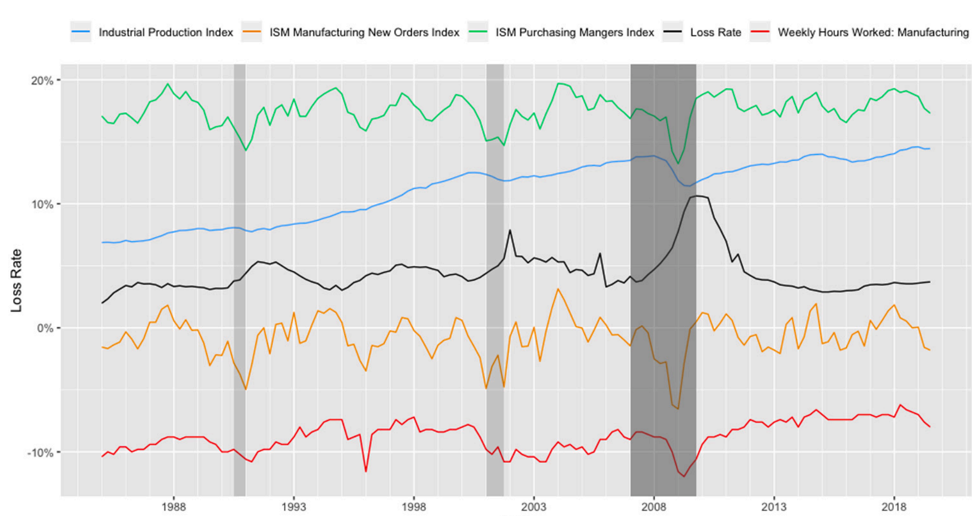

Fig. 4. Manufacturing related macroeconomic indicators.

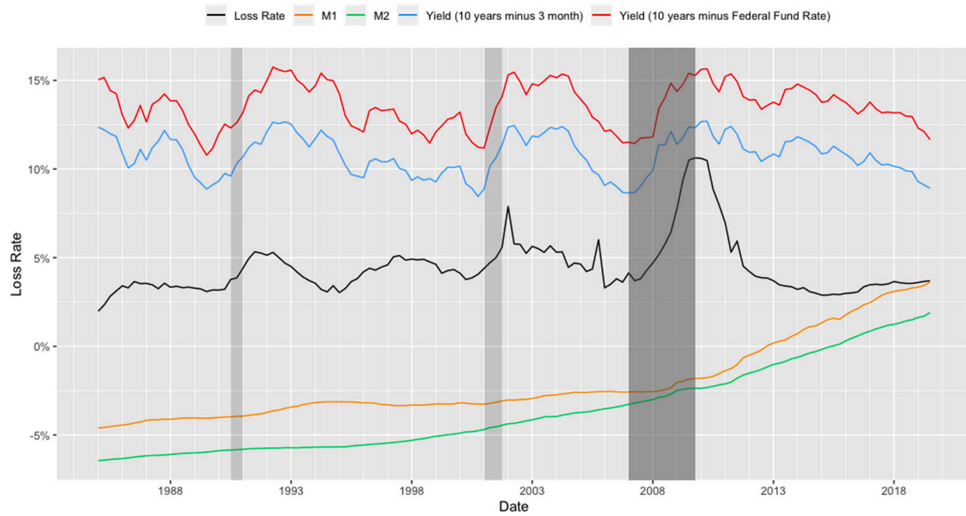

Fig. 5. Government related macroeconomic indicators.

As we can see in Table 3, only six macroeconomic indicators among the initial 19 indicators are selected using feature selection with optimal lags. These macroeconomic indicators are “buliding permits”, “initial unemployment insurance claims”, “M1”, “PMI”, “Weekly hours worked by manufacturing workers”, and “unemployment rate”. If we look at Table 1, we can see the interesting result that these indicators cover all the segments mentioned in the table. “Building permits”, “Initial unemployment insurance claims” and “unemployment rate” are from the consumer segment. “PMI” and “weekly hours worked by manufacturing workers” are from the business segment, and “M1” covers the government segment of the economy. The fact that our feature selection procedure selected indicators from all segments of the economy suggests that a holistic view of the economy is a requirement to build an effective loss forecasting framework. Additionally, we can see that “M1” is selected as a significant factor, and is in line with what we already suspected as the trend of “M1” is changed significantly after the great recession.

4.2.2. Feature selection with all lags

As opposed to the feature selection with optimal lags, in this version of feature selection, we do not select the optimal lags manually. We feed all the lagged values of macroeconomic indicators (95 input columns) to the model and let the model itself select the lagged indicators that are the most significant to predict loss. We applied Lasso regression on the input data and selected the indicators according to their relative importance. Lagged indicators with the relative importance greater than 0.2 are selected as final selection for the next step. The results are shown in Table 4. As we can see, the selected indicators are almost the same as what we have in feature selection with optimal lags, and all the indicators from feature selection with optimal lags (Table 3 are selected along with M2. Again, these macroeconomic indicators cover all the segments of the economy (consumer, business, and government segments). The main difference between the selected features in this version is that we allow multiple lags for one indicator to be selected. This way, the final model can also capture the trend of these macroeconomic indicators. It is interesting to see that in Table 4, for macroeconomic indicators that multiple lags are selected, these lags have at least two quarters difference. It means that the feature selection procedure tries to capture the most information by using the least number of variables in the cases that the trend had an important role.

4.3. Model training

The features selected by each of our feature selection procedures will be used as input to our machine learning models. The benchmark machine learning models that we use in this study are Lasso regression, Ridge regression, gradient boosting machine, and random forest. We use the data from the first quarter of 2011 to the second quarter of 2019 as the test set and develop two sets of results corresponding to each of our feature selection procedures. We report R^2 for the training set and Mean Squared Error (MSE) for both training and test sets. The results using the output of “feature selection with optimal lags” are reported in Table 5. The corresponding plots for the fit of each machine learning model are shown in Fig. 6. Comparing the values of R^2 in Table 5, we see that the gradient boosting machine shows a better performance in terms of R^2, which means that 77% of variations in the loss rate can be explained by the gradient boosting method using optimal lags. The values of MSE in training and test sets are also in line with our conclusion, and the gradient boosting machine shows the best performance in terms of MSE on both training and test sets. Hence, the gradient boosting machine is selected for making the final prediction in the next step when we use Lasso with optimal lags as our model selection procedure (see Table 6).

Table 7 shows the statistics corresponding to the result of each machine learning method when using feature selection with all lags. The final fit for each method is depicted in Fig. 7. The R^2 results in Fig. 7 suggest that both Lasso and Ridge regression have similar performance. However, looking at the values of MSE in the training and test sets, we see that Ridge regression has a better performance on both train and validation sets. Hence, we select Ridge regression for generating the final forecasts when we use the output of feature selection with all lags as the input of the machine learning model (see Table 8).

As Figs. 6 and 7 show, the uptrend of the loss rate in the last four quarters can be captured by all the benchmark models using selected features, which is not possible when the unemployment rate is the only decision variable. Additionally, we can see that all the models are able to capture the trends of loss rate with an acceptable accuracy, which shows that our loss forecasting method can provide acceptable results using any of the benchmark machine learning models. We use the selected machine learning model in this step to generate final forecasts in the next step of our loss forecasting algorithm, which is explained in the next subsection.

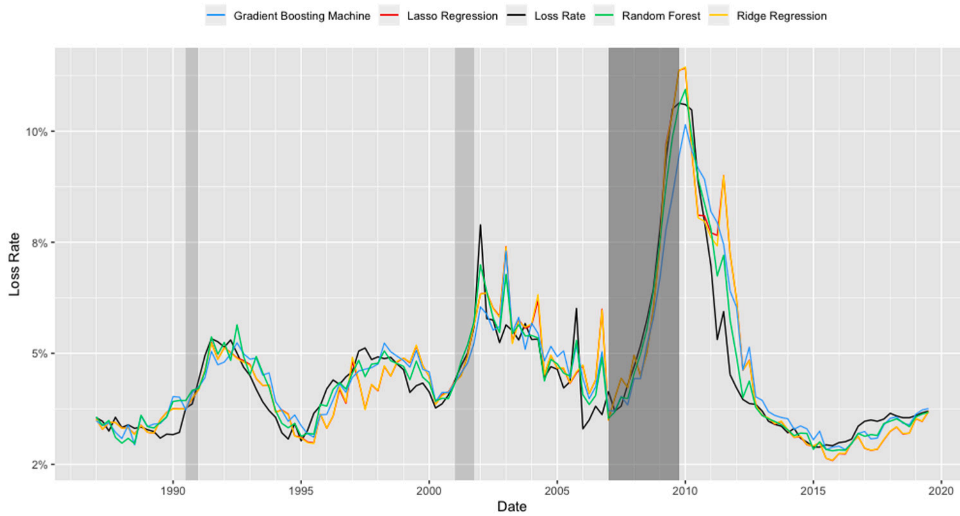

Fig. 6. Final fits for machine learning models using optimal lags as input variables.

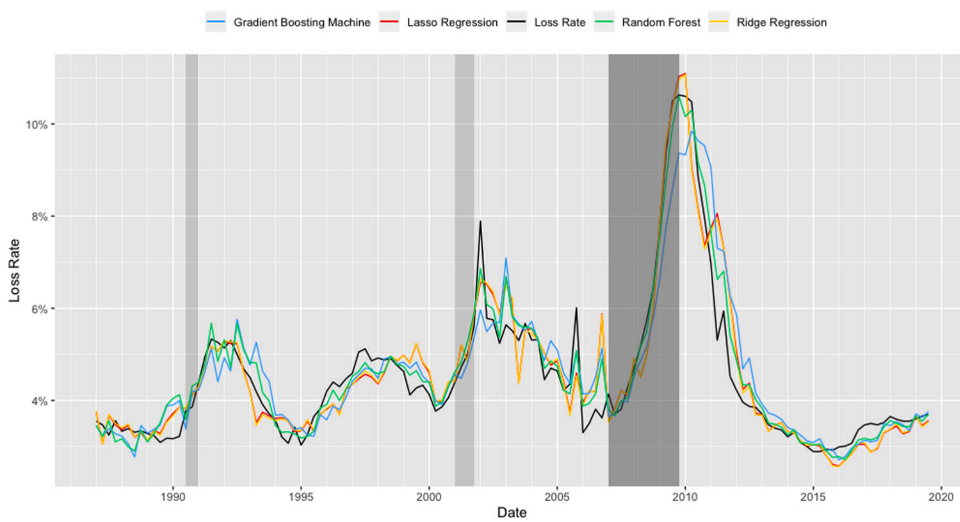

Fig. 7. Final fits for machine learning models using all lags as input variables.

Table 6
Coefficients and relative importance for machine learning models when using indicators with optimal lags.

Indicator	Lag	Coefficients		Relative importance	
		Lasso regression	Ridge regression	Gradient boosting machine	Random forest
Intercept	–	0.0388	0.0387	–	–
Building permits	1	−0.3915	−0.3922	0.1528	0.1528
Initial unemployment insurance claims	1	0.3750	0.3585	1.0000	1.0000
M1	4	0.0112	0.0135	0.0422	0.0422
PMI	4	−0.2070	−0.2094	0.0387	0.0387
Weekly hours worked by manufacturing orders	2	−2.2712	−2.5655	0.0508	0.0508
Unemployment rate	0	0.3979	0.3915	0.2061	0.2061

4.4. Forecasting

The last step to build the loss forecasting framework is to predict each macroeconomic indicator and use the trained model in step 3 to predict the future values of the loss rate. We use the second quarter of 2018 to the second quarter of 2019 (1 year) as the prediction period. We predict each macroeconomic indicator for this period, and using the trained model in step 3, we will forecast the loss rate. We will use actual values for this period to evaluate the forecasted values.

To forecast the values of each macroeconomic indicator, we need to select the most appropriate time series forecasting model. Since the performance of forecasting models highly depends on the underlying characteristics of the time series, the selection of the best is not a

Table 7
Summary statistics for machine learning models when using indicators with all lags.

Model 2: All lags	R^2	MSE (Train)	MSE (Validation)
Lasso regression	0.81	9.62E−03	1.20E−02
Ridge regression	0.81	3.82E−03	7.85E−03
Gradient boosting machine	0.77	1.01E−02	1.20E−02
Random forest	0.72	1.05E−02	1.67E−02

Table 8
Coefficients and relative importance for machine learning models when using indicators with all lags.

Indicator	Lag	Coefficients		Relative importance	
		Lasso regression	Ridge regression	Gradient boosting machine	Random forest
Intercept	–	0.0265	0.0250	–	–
Building permits	1	−0.2848	−0.2882	0.1552	0.5193
Initial unemployment insurance claims	0	0.5631	0.5355	0.7668	1.0000
Initial unemployment insurance claims	2	0.3184	0.3531	0.3061	0.5519
Initial unemployment insurance claims	4	0.4846	0.4877	0.1558	0.1674
M1	1	0.0267	0.0295	0.0436	0.0787
M2	4	0.0098	0.0157	0.0291	0.0655
PMI	0	0.5100	0.5518	0.0325	0.0979
PMI	2	0.0549	0.1134	0.0544	0.1976
Weekly hours worked by manufacturing orders	0	−2.9343	−3.1827	0.0238	0.1245
Unemployment rate	0	0.3452	0.4272	1.0000	0.6851
Unemployment rate	4	0.3910	0.4802	0.0166	0.0764

Table 9
Comparing the performance of MSIC to traditional train/validation model selection procedure using "Building Permits" data.

Building Permits	Optimal	Traditional	MSIC-LR		MSIC-SVM		MSIC-DT		Average optimal gap reduction
Train/validation separation point	MSE	MSE	MSE	Optimal gap reduction	MSE	Optimal gap reduction	MSE	Optimal gap reduction	
P1 = 24	5.86E+03	2.17E+04	7.27+03	91.07%	1.75E+04	26.20%	8.13E+03	85.61%	67.62%
P1 = 27	5.86E+03	2.97E+04	7.27E+03	94.09%	1.86E+04	46.51%	1.92E+04	44.08%	61.56%
P1 = 30	5.86E+03	4.54E+04	7.77E+03	95.15%	1.58E+04	74.83%	7.18E+03	96.65%	88.88%
P1 = 33	5.86E+03	1.13E+05	1.63E+04	90.26%	7.57E+03	98.41%	6.07E+04	49.02%	79.23%

Table 10
Comparing the performance of MSIC to traditional train/validation model selection procedure using "Initial Unemployment Insurance Claims" data.

Initial Unemployment Insurance Claims	Optimal	Traditional	MSIC-LR		MSIC-SVM		MSIC-DT		Average optimal gap reduction
Train/validation separation point	MSE	MSE	MSE	Optimal gap reduction	MSE	Optimal gap reduction	MSE	Optimal gap reduction	
P1 = 24	4.68E+10	6.08E+10	5.65E+10	30.57%	5.55E+10	37.95%	5.58E+10	35.84%	34.79%
P1 = 27	4.68E+10	5.84E+10	5.25E+10	50.43%	5.50E+10	29.59%	5.50E+10	29.59%	36.54%
P1 = 30	4.68E+10	6.08E+10	5.51E+10	41.17%	5.59E+10	35.20%	5.55E+10	37.95%	38.11%
P1 = 33	4.68E+10	6.04E+10	5.94E+10	7.54%	5.53E+10	37.78%	5.84E+10	14.96%	20.09%

Table 11
Comparing the performance of MSIC to traditional train/validation model selection procedure using "M1" data.

M1	Optimal	Traditional	MSIC-LR		MSIC-SVM		MSIC-DT		Average optimal gap reduction
Train/validation separation point	MSE	MSE	MSE	Optimal gap reduction	MSE	Optimal gap reduction	MSE	Optimal gap reduction	
P1 = 24	1.40E+02	3.84E+02	1.51E+02	95.33%	1.54E+02	94.35%	1.52E+02	94.99%	94.89%
P1 = 27	1.40E+02	1.14E+03	1.47E+02	99.28%	3.88E+02	75.08%	1.10E+03	3.60%	59.32%
P1 = 30	1.40E+02	2.25E+02	1.54E+02	83.55%	1.55E+02	82.39%	1.52E+02	85.56%	83.83%
P1 = 33	1.40E+02	1.83E+02	1.51E+02	74.65%	1.54E+02	66.52%	1.50E+02	77.43%	72.87%

simple task. As the forecasting model selection approach in Taghiyeh et al. (2020a) (MSIC algorithm) has shown promising performance, we will use this procedure to select our forecasting model for each macroeconomic indicator. Similar to Taghiyeh et al. (2020a), we select seven of the most well-known time series forecasting models as our benchmark, namely naiv̈e forecasting, moving average, ARIMA, simple exponential smoothing, Holt's linear trend, Holt-Winters, and theta. For each macroeconomic indicator, the MISC algorithm will select the optimal forecasting model, and we will use the selected optimal model to forecast future values for each macroeconomic indicator.

Since MSIC needs multiple time series as input to train its classifiers, we need to convert the time series associated with each macroeconomic indicator into several series. To achieve this goal, we use non-overlapping four-year horizons to split the data for each macroeconomic indicator. We use this input data to train the MSIC classifiers. To make final predictions, we use the entire data for the corresponding macroeconomic indicator as input to the trained classifiers of the MSIC algorithm.

To evaluate the performance of the MSIC algorithm for each macroeconomic indicator, we compare the results of the MSIC algorithm to the traditional train/validation forecasting model selection method. Three variants of the MSIC algorithm, namely MSIC with logistic regression as the classifier (MSIC-LR), MSIC with support vector machine as a classifier (MSIC-SVM), and MSIC with decision tree as a classifier (MSIC-DT) are used for this comparison, and we report MSE and optimality gap reduction as the comparison measures. To be consistent with the results reported in Taghiyeh et al. (2020a), we use different values for separations points between train and validation sets (P1). Since in the feature selection step (step 2) only 7 of the macroeconomic indicators are selected (building permits, initial unemployment insurance claims, M1, M2, purchasing managers index, weekly hours worked by manufacturing workers and unemployment rate), we only use the MSIC algorithm to predict future values for these indicators. The comparison results for the selected macroeconomic indicators are shown in Tables 9–15. The MSE results are also depicted in Figs. 8–14. The optimality gap improvements are summarized in Fig. 15.

Table 12
Comparing the performance of MSIC to traditional train/validation model selection procedure using “M2” data.

M2	Optimal	Traditional	MSIC-LR		MSIC-SVM		MSIC-DT		Average optimal gap reduction
Train/validation separation point	MSE	MSE	MSE	Optimal gap reduction	MSE	Optimal gap reduction	MSE	Optimal gap reduction	
P1 = 24	4.12E+02	3.12E+04	4.16E+02	99.99%	4.31E+02	99.94%	4.16E+02	99.99%	99.97%
P1 = 27	4.12E+02	9.33E+03	4.16E+02	99.95%	3.71E+03	63.01%	4.16E+02	99.95%	87.64%
P1 = 30	4.12E+02	3.52E+03	4.29E+02	99.47%	4.31E+02	99.40%	4.91E+02	97.46%	98.78%
P1 = 33	4.12E+02	6.54E+02	4.91E+02	67.23%	4.16E+02	98.45%	4.16E+02	98.32%	88.00%

Table 13
Comparing the performance of MSIC to traditional train/validation model selection procedure using “Purchasing Managers Index (PMI)” data.

Purchasing Managers Index (PMI)	Optimal	Traditional	MSIC-LR		MSIC-SVM		MSIC-DT		Average optimal gap reduction
Train/validation separation point	MSE	MSE	MSE	Optimal gap reduction	MSE	Optimal gap reduction	MSE	Optimal gap reduction	
P1 = 24	7.48E+00	8.52E+00	8.40E+00	10.94%	8.11E+00	39.31%	8.11E+00	39.18%	29.81%
P1 = 27	17.48E+00	3.51E+01	1.43E+01	75.22%	8.80E+00	95.19%	8.11E+00	97.70%	89.37%
P1 = 30	7.48E+00	4.19E+01	1.51E+01	77.99%	2.27E+01	55.86%	8.11E+00	98.16%	77.34%
P1 = 33	7.48E+00	2.57E+01	9.10E+00	91.10%	1.27E+01	71.22%	8.52E+00	94.28%	85.53%

Table 14
Comparing the performance of MSIC to traditional train/validation model selection procedure using “Weekly Hours Worked: Manufacturing” data.

Weekly Hours Worked: Manufacturing	Optimal	Traditional	MSIC-LR		MSIC-SVM		MSIC-DT		Average optimal gap reduction
Train/validation separation point	MSE	MSE	MSE	Optimal gap reduction	MSE	Optimal gap reduction	MSE	Optimal gap reduction	
P1 = 24	3.06E−02	5.35E−02	4.38E−02	42.35%	4.19E−02	50.87%	4.31E−02	45.48%	46.23%
P1 = 27	3.06E−02	4.55E−02	4.11E−02	30.03%	4.28E−02	18.01%	3.84E−02	47.86%	31.97%
P1 = 30	3.06E−02	4.44E−02	4.27E−02	12.37%	3.98E−02	33.18%	4.09E−02	25.13%	23.56%
P1 = 33	3.06E−02	5.33E−02	4.97E−02	15.95%	4.59E−02	32.60%	3.84E−02	65.74%	38.10%

Table 15
Comparing the performance of MSIC to traditional train/validation model selection procedure using “Unemployment Rate” data.

Unemployment Rate	Optimal	Traditional	MSIC-LR		MSIC-SVM		MSIC-DT		Average optimal gap reduction
Train/validation separation point	MSE	MSE	MSE	Optimal gap reduction	MSE	Optimal gap reduction	MSE	Optimal gap reduction	
P1 = 24	2.93E−02	1.52E+00	3.68E−02	99.50%	2.01E−01	88.50%	1.09E+00	29.01%	72.34%
P1 = 27	2.93E−02	1.30E+00	3.68E−02	99.41%	4.53E−02	98.74%	9.58E−01	27.07%	75.07%
P1 = 30	2.93E−02	1.30E+00	3.66E−02	99.43%	1.97E−01	86.83%	8.84E−01	32.84%	73.03%
P1 = 33	12.93E−02	1.30E+00	5.54E−02	97.95%	2.83E−01	80.11%	1.03E+00	21.73%	66.60%

The results suggest the same trend as numerical results in Taghiyeh et al. (2020a), as the MSIC algorithm shows a constant improvement in the optimality gap in all instances. Additionally, there is not a single winner among classifiers for the MSIC algorithm, and it is case dependent. As the overall performance in Fig. 15 shows, we can get an overall minimum of 60% improvement in optimality gap improvement using the MSIC algorithm over the traditional train/validation model selection procedure.

Now that the forecasting models are selected for each macroeconomic indicator, and the predictions are made, we use the forecast values as input to the trained models in step 3. Gradient boosting machine was selected as the best performing machine learning model using feature selection with optimal lags, and ridge regression was the winner when using feature selection with all lags. Therefore, these two models are used to generate the final forecasts for the loss rate. MSE results for final forecasts are reported in Table 16. Since all the variants of the MISC are generating the same results, we show all the predictions in one figure, which is representative of the results for all the variants of the MSIC algorithm. The prediction plots are shown in Fig. 16. As the MSE results in Table 16 show, we achieve significantly low values for MSE using our proposed loss forecasting framework that shows the efficiency of the algorithm. Moreover, looking at Fig. 16, we see that both variants of our loss forecasting model can closely predict the loss rate values, and it is able to capture the uptrend of the loss rate, which is not possible when using only unemployment rate as the decision variable. Overall, we see that ridge regression with all lags can obtain better results than gradient boosting with optimal lags. The reason is that in the feature selection with all lags, the lags are selected automatically by the model, and we allow the model to use more than one lag from each indicator. This way, more data is available to make predictions. Hence, the Ridge regression with all lags can perform better than gradient boosting with optimal lags. Moreover, Ridge regression is from the family of monotonic and linear machine learning models, which makes it highly interpretable and the assigned coefficients are available for each input variable.

Table 16
MSE for predictions resulted from two feature selection approaches (optimal lags and all lags). For each approach, MSE values are reported when using three different variants of MSIC as forecasting model selection procedure.

Model	MSIC_LR	MSIC_SVM	MSIC_DT
Gradient boosting machine (Optimal lags)	1.15E−03	1.15E−03	1.15E−03
Ridge regression (All lags)	1.04E−03	1.04E−03	1.04E−03

Now that we have selected the macroeconomic indicators with a significant correlation with the charge-off rate and built a prediction model using these values, one may bring up the question that whether the selected macroeconomic indicators are actually the ones causing

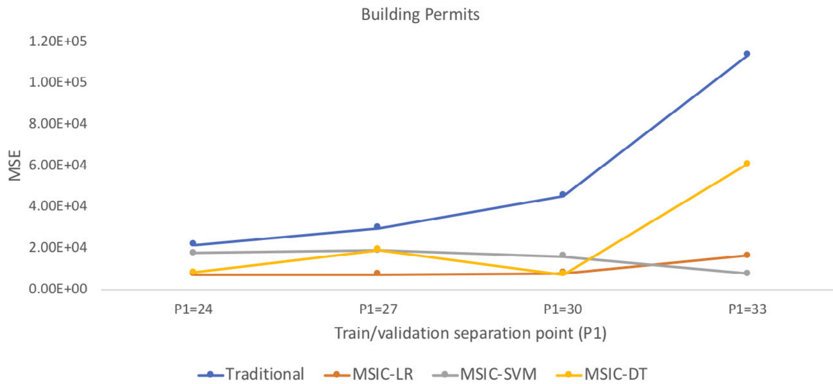

Fig. 8. Performance comparison using "Building Permits" Data.

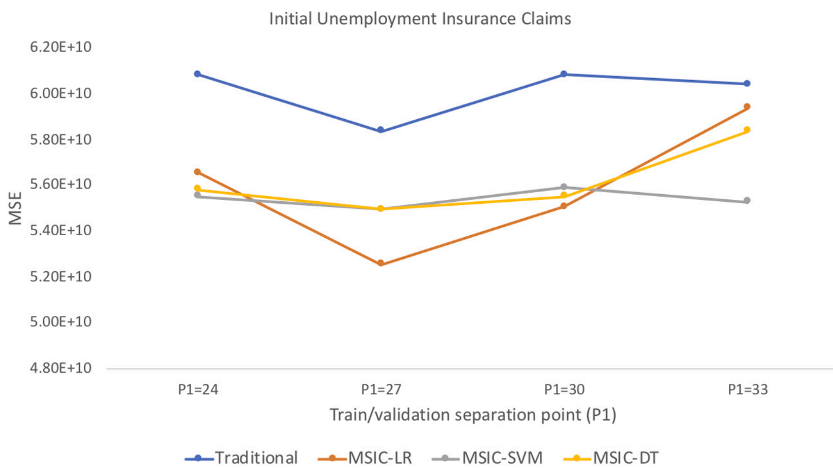

Fig. 9. Performance comparison using "Initial Unemployment Insurance Claims" Data.

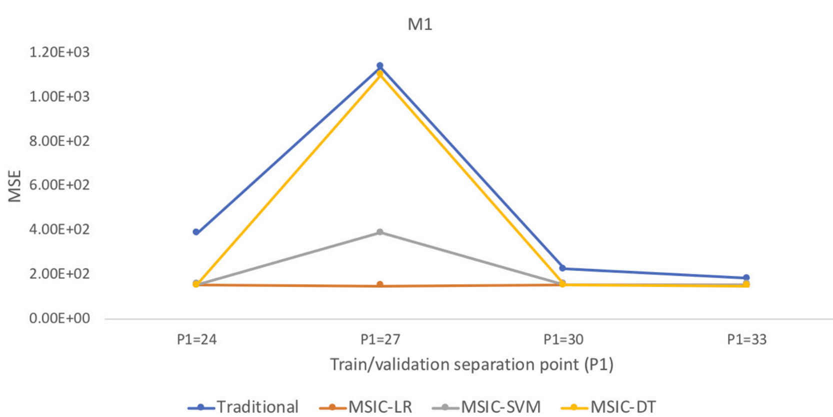

Fig. 10. Performance comparison using "M1" Data.

the fluctuations in the charge-off rate or not. While correlation and causation may exist at the same time, but the existence of a correlation does not necessarily imply causation. Causation applies in the situations that an action explicitly triggers another action, but correlation simply implies a relationship. When a correlation exists between two actions, it means that they are related to each other, but it does not necessarily mean that any of them cause the other one. We use the example from the book "Introduction to statistical learning" by James et al. (2013) to explain this issue. Suppose that we are evaluating the correlation between the sales of an ice cream vendor on a beach with a number of shark attacks. Interestingly, they have a high correlation, but it does not mean that selling ice cream on the beach causes more shark attacks or vice versa. However, when the weather is hot, people are more attracted to the beaches, and consequently, the number of ice cream sales increases. When there are more people on the beach, there is a higher chance of a shark attack, and the higher temperature is the cause of attracting more people to the beach, which results in more shark attacks. This example illustrates the difference between causation and correlation. Our primary focus in this research was on correlation rather than causality. The question regarding whether the final significant indicators that were selected to build the model are actually causing the chain of events that leads to the changes in the charge-off rate is left to the experts in the credit card industry and economists.

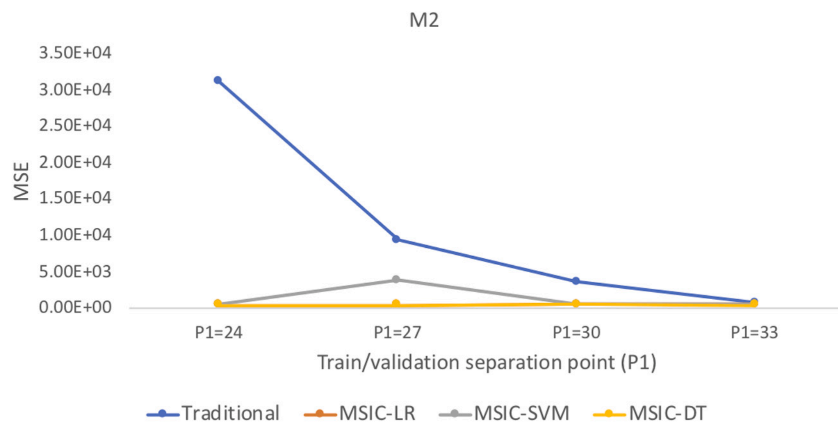

Fig. 11. Performance comparison using "M2" Data.

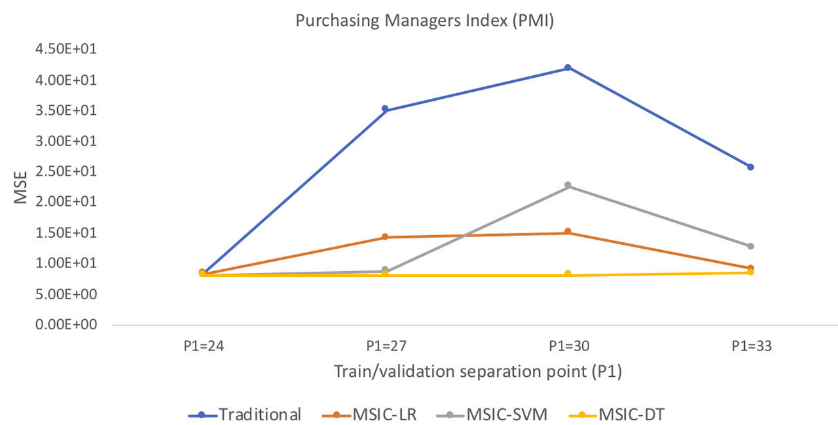

Fig. 12. Performance comparison using "Purchasing Managers Index (PMI)" Data.

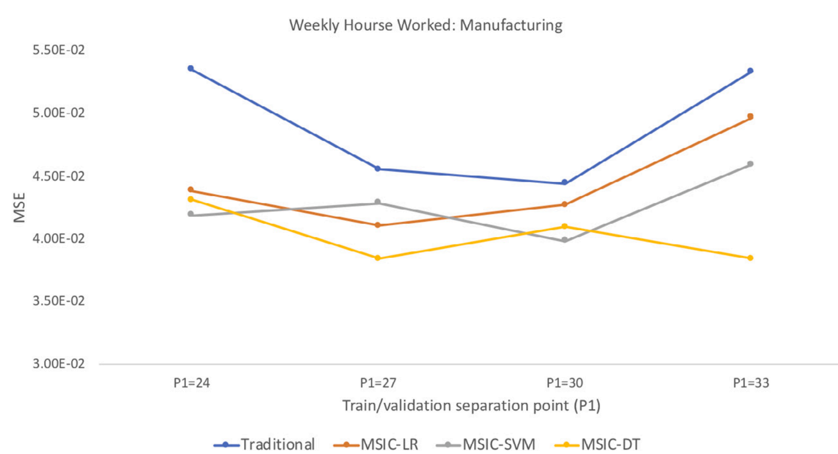

Fig. 13. Performance comparison using "Weekly Hours Worked: Manufacturing" Data.

5. Conclusions

In this paper, we have proposed a machine learning based loss forecasting framework for the credit card industry using macroeconomic indicators. Our goal was to cover macroeconomic indicators from all segments of the economy to make predictions based on a holistic view of the economic conditions. Using the review of the literature and experts' opinion, we selected 19 macroeconomic indicators, which cover consumer, business, and government sections of the economy as input to the proposed loss forecasting framework. The proposed procedure consists of four steps, data preparation, feature selection, model training, and forecasting. We used four machine learning models, namely Lasso regression, Ridge regression, gradient boosting machine, and random forest to develop two versions of the loss forecasting framework. The difference between these two versions is in the utilization of lags from input data. We also applied the proposed model selection procedure in Taghiyeh et al. (2020a) (MSIC algorithm) in the forecasting segment of the proposed loss forecasting framework to find the optimal time series forecasting model. To the best of our knowledge, this work is the first that uses an extensive number of macroeconomic indicators from all segments of the economy to build a

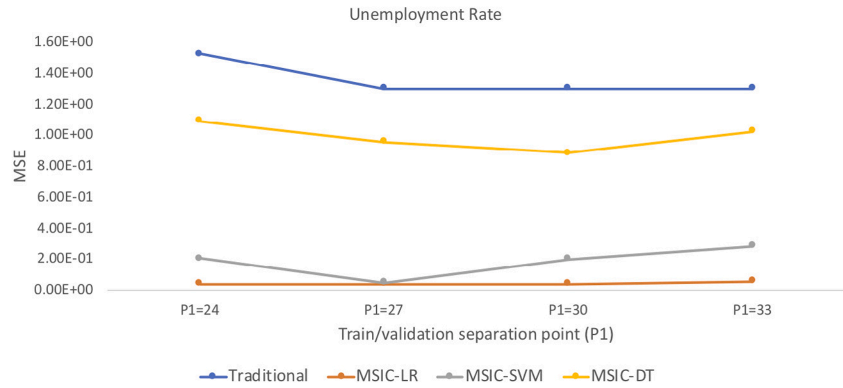

Fig. 14. Performance comparison using "Unemployment Rate" Data.

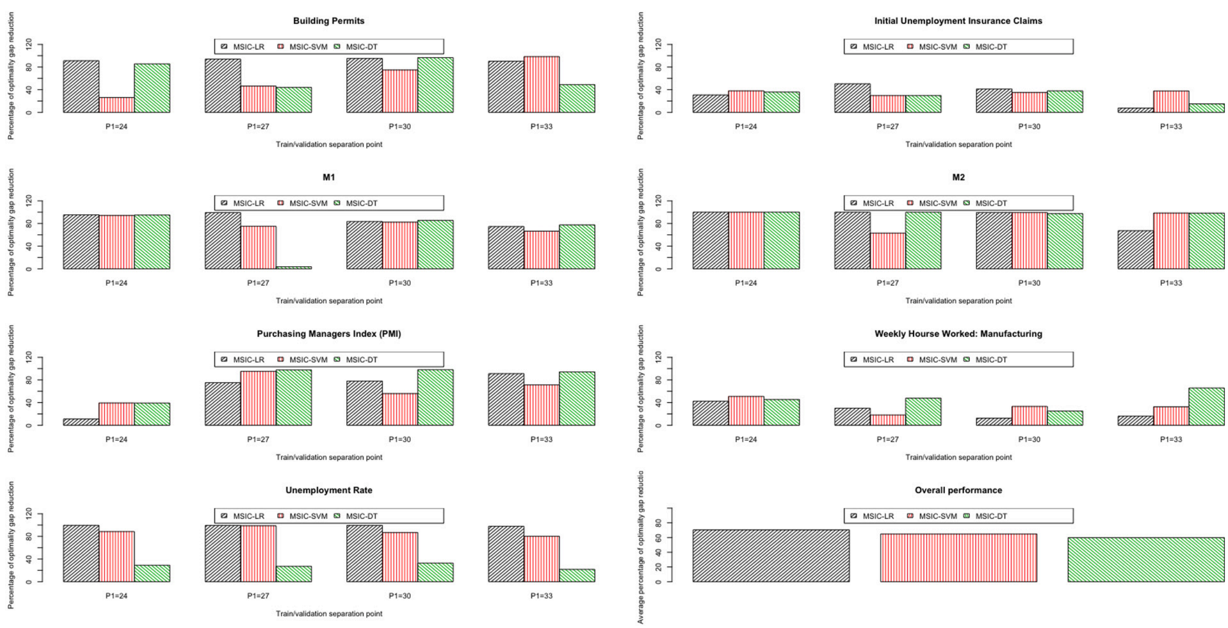

Fig. 15. Optimality gap improvement for all macro economic indicators using three versions of MSIC. Average improvements for all three versions over all categories are shown in last figure.

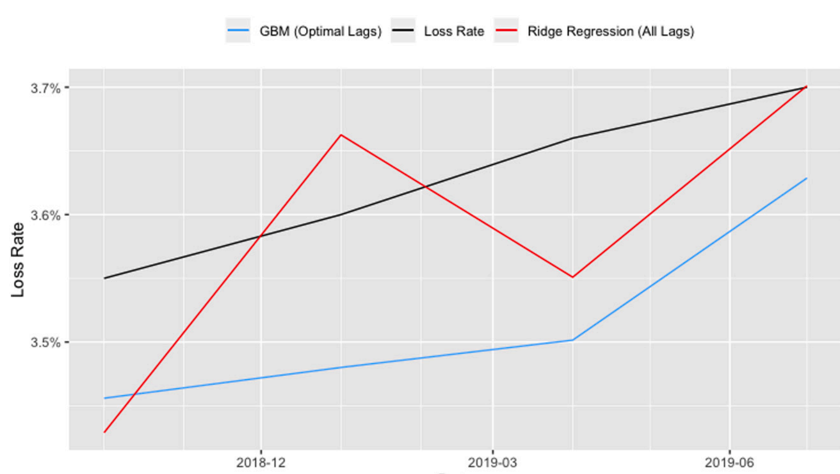

Fig. 16. Prediction plots for 2018Q2–2019Q2 when using MSIC as forecasting model selection procedure.

machine learning based loss forecasting framework for the U.S. credit card industry. To show the performance of the proposed loss forecasting framework, we used the charge-off data for the top 100 banks in the U.S. ranked by assets from 1985 to 2019, and the data corresponding

Table A.1
Reference list for macroeconomic indicators used for building the loss forecasting framework.

Indicator	Reference
Building Permits	U.S. Census Bureau and U.S Department of Housing and Urban Development (2020b)
Housing Starts	U.S. Census Bureau and U.S Department of Housing and Urban Development (2020a)
Initial Unemployment Insurance Claims	U.S. Employment and Training Administration (2020)
Unemployment Rate	U.S. Bureau of Labor Statistics (2020)
Consumer Confidence Index (CCI)	OECD (2020)
University of Michigan Sentiment Index	University of Michigan (2020)
S&P 500 Index	S & P Dow Jones Indices LLC (2020b)
Dow Jones Industrial Average	S & P Dow Jones Indices LLC (2020a)
Total Credit Utilization	Board of Governors of the Federal Reserve System (US) (2020a)
Revolving Credit Utilization	Board of Governors of the Federal Reserve System (US) (2020c)
Non Revolving Credit Utilization	Board of Governors of the Federal Reserve System (US) (2020b)
Industrial Production Index	Board of Governors of the Federal Reserve System (US) (2020)
ISM Manufacturing New Orders	Institute for Supply Management (2020b)
ISM Purchasing Managers Index	Institute for Supply Management (2020a)
Weekly Hours Worked by Manufacturing Workers	Organization for Economic Co-operation and Development (2020)
M1	Board of Governors of the Federal Reserve System (U.S.) (2020a)
M2	Board of Governors of the Federal Reserve System (U.S.) (2020b)
Yield (10 years minus 3 month)	Federal Reserve Bank of St. Louis (2020a)
Yield (10 years minus Federal Fund Rate)	Federal Reserve Bank of St. Louis (2020b)

to selected macroeconomic indicators. We applied the proposed loss forecasting framework on the data, and the final results were very promising. We could achieve the test MSE of 1.15E−03 and 1.04E−03 corresponding to feature selection with optimal lags and feature selection with all lags, respectively, which shows the effectiveness of the proposed algorithm in forecasting the loss rate. The final fit for the prediction period shows that we could closely predict the actual values of the loss rate and the uptrend of the loss rate could be captured by our proposed model, which was not possible in the conventional version of the credit card loss forecasting frameworks that only use the unemployment rate as the decision variable.

In the future, we aim to further improve the proposed loss forecasting model in this paper by adding more machine learning models, such as deep neural networks, long-short term memory (LSTM) model, and extreme gradient boosting to the benchmark models and see if we can make more accurate forecasts. Additionally, more feature selection procedures can be explored to improve the feature selection step of the loss forecasting framework. The other future line of research would be to perform a more exhaustive number of transformation for the macroeconomic indicators to see if a better data transformation can be found to improve the efficiency of the algorithm further, as the final results are sensitive to these transformations. Another interesting future research path is to analyze the credit card charge-off rates due to the rapid changes in the economy caused by the Coronavirus pandemic and adjust the model accordingly.

CRediT authorship contribution statement

Sajjad Taghiyeh: Conceptualization, Methodology, Formal analysis, Software, Visualization, Validation, Writing - original draft. **David C. Lengacher:** Data curation, Conceptualization, Methodology, Writing - review & editing. **Robert B. Handfield:** Supervision, Conceptualization, Validation, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Loss forecasting algorithm

Let n be the number of macroeconomic indicators that are selected to build the loss forecasting framework and let m be the number of machine learning models used to predict the loss rate. Feature selection method can be set to either "Optimal lags" or "All lags".

- Step 1 (data preparation): Initialize final input values as $I_1 = \{\}$. For each macroeconomic indicator i $(i = 1, \dots, n)$:
  - Step 1-1: Convert macroeconomic indicator i into quarterly values.
  - Step 1-2: Add lags from 1 to 4 quarters to indicator i and record the lagged indicator.
  - Step 1-3: Try different transformations (e.g., square root, exponential, square, log, etc.) for each lagged indicator and select the best one based on a goodness of fit statistic.
  - Step 1-4: For each lagged indicator, add the selected transformation in step 1-3 to I_1.
  - Step 1-5: Convert all the lagged indicators to year over year values by dividing them by the corresponding values from last year.
  - Step 1-6: If $i = n$, go to step 2. Else, set $i = i + 1$ and go to step 1-1.
- Step 2 (feature selection): If "Optimal lags" is selected for feature selection, go to step 2-1. Else, if "All lags" is selected go to step 2-2.
  - Step 2–1 (feature selection with optimal lags): Initialize the input data for feature selection as $I_2 = \{\}$ and list of final selected features as F. Use loss rate as dependent variable.
    - ∗ Step 2-1-1: For each macroeconomic indicator i $(i = 1, \dots, n)$, select the lag with the highest correlation with loss rate from I_1 and append it to I_2.
    - ∗ Step 2-1-2: Apply Lasso regression using the input data I_2 and loss rate. Use hyperparameter optimization to achieve the best fit in terms of R^2.
    - ∗ Step 2-1-3: Record the feature importance for each input feature from the Lasso regression model.
    - ∗ Step 2-1-4: Add features with feature importance greater than 0.2 to the list of selected features (F).
  - Step 2–2 (feature selection with all lags): Initialize the input data for feature selection as $I_2 = \{\}$ and list of final selected features as F. Use the loss rate as the dependent variable.
    - ∗ Step 2-2-1: For each macroeconomic indicator i $(i = 1, \dots, n)$, select all lagged values from I_1 and append it to I_2.
    - ∗ Step 2-2-2: Apply Lasso regression using the input data I_2 and loss rate. Use hyperparameter optimization to achieve the best fit in terms of R^2.

    * Step 2-2-3: Record the feature importance for each input feature from the Lasso regression model.
    * Step 2-2-4: Add features with feature importance greater than 0.2 to the list of selected features (F).

- Step 3 (model training): For each machine learning model j ($j = 1, \ldots, m$):
  - Step 3-1: Use input values selected from the feature selection step (F) as independent variables and loss rate as the dependent variable.
  - Step 3-2: Split the data into training and test set.
  - Step 3-3: Train model j on the training set and test it on the test set. Record R^2 for the training set and MSE for both training and test sets. Use hyperparameter optimization to achieve the best fit.
  - Step 3-4: If $j = m$, compare the performance of all machine learning models and select the best performing one to use in step 4 to generate the final predictions.

- Step 4 (Forecasting): Use the features selected in step 2 (F) as input values for the machine learning model selected in step 3. Let t be the number of macroeconomic indicators in F, and let P be a list containing the final predictions for macroeconomic indicators in F. We will use the MSIC algorithm proposed in Taghiyeh et al. (2020a) for the forecasting model selection for each macroeconomic indicator.
  - Step 4-1: For each macroeconomic indicator in F ($k = 1, \ldots, t$):
    * Step 4-1-1: Initialize the input data for the MSIC algorithm as $R = \{\}$.
    * Step 4-1-2: split the time series corresponding to macroeconomic indicator k into 4-year chunks and append it to R.
    * Step 4-1-3: Train the MSIC algorithm on R.
    * Step 4-1-4: Use the entire values for macroeconomic indicator k as input for the MISC to select the best forecasting model and make final predictions for time series k. Append the results of the MSIC algorithm to P.
  - Step 4-2: Use P as the new input to the selected machine learning model in step 3 to generate the final predictions for the loss rate.

References

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). Sanity checks for saliency maps. In *Advances in neural information processing systems* (pp. 9505–9515).

Agarwal, S., & Liu, C. (2003). Determinants of credit card delinquency and bankruptcy: Macroeconomic factors. *Journal of Economics and Finance*, *27*, 75–84.

Ausubel, L. M. (1997). Credit card defaults, credit card profits, and bankruptcy. *American Bankr. LJ*, *71*(249).

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One*, *10*.

Bellotti, T., & Crook, J. (2013). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, *29*, 563–574.

Berge, T. J., & Jordà, Ò. (2011). Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics*, *3*, 246–277.

Board of Governors of the Federal Reserve System (U. S.) (2020). Charge-off rate on credit card loans, top 100 banks ranked by assets [corcct100s]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CORCCT100S.

Board of Governors of the Federal Reserve System (US) (2020). Industrial production index [indpro]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/INDPRO.

Board of Governors of the Federal Reserve System (U. S.) (2020a). M1 money stock [m1]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CORCCT100S.

Board of Governors of the Federal Reserve System (US) (2020a). Total consumer credit owned and securitized, outstanding [totalsl]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TOTALSL.

Board of Governors of the Federal Reserve System (U. S.) (2020b). M2 money stock [m2]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/M2.

Board of Governors of the Federal Reserve System (US) (2020b). Total nonrevolving credit owned and securitized, outstanding [nonrevsl]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/NONREVSL.

Board of Governors of the Federal Reserve System (US) (2020c). Total revolving credit owned and securitized, outstanding [revolsl]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/REVOLSL.

Censky, A. (2010). Consumer confidence slumps in september. *CNN Money*.

Dalio, R. (2018). *Principles for navigating big debt crises*. Westport, CT: Bridgewater.

Debbaut, P., Ghent, A., & Kudlyak, M. (2016). The card act and young borrowers: The effects and the affected. *Journal of Money, Credit and Banking*, *48*, 1495–1513.

Desai, C. A., Elliehausen, G., & Lawrence, E. C. (2014). On the county-level credit outcome beta. *Journal of Financial Services Research*, *45*, 201–218.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The Journal of Finance*, *46*, 555–576.

Estrella, A., & Mishkin, F. S. (1998). Predicting us recessions: Financial variables as leading indicators. *The Review of Economics and Statistics*, *80*, 45–61.

Federal Reserve Bank of St. Louis (2020a). 10-year treasury constant maturity minus 3-month treasury constant maturity [t10y3m]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/T10Y3M.

Federal Reserve Bank of St. Louis (2020b). 10-year treasury constant maturity minus federal funds rate [t10yff]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/T10YFF.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems?. *The Journal of Machine Learning Research*, *15*, 3133–3181.

Figlewski, S., Frydman, H., & Liang, W. (2012). Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, *21*, 87–105.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 1–67.

Fung, T., & Wong, M. (2002). *Modeling credit card charge-off ratios: the case of Hong Kong*. City University of Hong Kong: Department of Economics & Finance.

Giesecke, K., Longstaff, F. A., Schaefer, S., & Strebulaev, I. (2011). Corporate bond default risk: A 150-year perspective. *Journal of Financial Economic*, *102*, 233–250.

Gross, D. B., & Souleles, N. S. (2002). An empirical analysis of personal bankruptcy and delinquency. *Review of Financial Studies*, *15*, 319–347.

Hall, P., & Gill, N. (2018). *Introduction to machine learning interpretability*. O'Reilly Media, Incorporated.

Institute for Supply Management (2020a). Ism manufacturing report on business, manufacturing new orders.

Institute for Supply Management (2020b). Ism manufacturing report on business, purchasing managers index.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning, Vol. 112*. Springer.

Kim, J. W., Won, S., & Kim, J. I. (2017). Additional credit for liquidity-constrained individuals: High-interest consumer credit in korea. *Emerging Markets Finance and Trade*, *53*, 109–127.

Krishnan, M. (2019). Against interpretability: a critical examination of the interpretability problem in machine learning. *Philosophy & Technology*, 1–16.

Leow, M., & Crook, J. (2014). Intensity models and transition probabilities for credit card loan delinquencies. *European Journal of Operational Research*, *236*, 685–694.

Levanon, G., Ozyildirim, A., Manini, J. C., Schaitkin, B., & Tanchua, J. (2011). *Using a leading credit index to predict turning points in the us business cycle*: *The conference board economics program working paper*.

Liu, J., & Xu, X. E. (2003). The predictive power of economic indicators in consumer credit risk management. *Rma Journal*, *86*, 40–45.

Mazibaş, M., & Tuna, Y. (2017). Understanding the recent growth in consumer loans and credit cards in emerging markets: Evidence from turkey. *Emerging Markets Finance and Trade*, *53*, 2333–2346.

Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2019). Improved support vector machine models for work zone crash injury severity prediction and analysis. *Transportation Research Record*, *2673*, 680–692.

Mokhtarimousavi, S., Talebi, D., & Asgari, H. (2018). A non-dominated sorting genetic algorithm approach for optimization of multi-objective airport gate assignment problem. *Transportation Research Record*, *2672*, 59–70.

Musto, D. K., & Souleles, N. S. (2006). A portfolio view of consumer credit. *Journal of Monetary Economics*, *53*, 59–84.

Ng, S. (2014). Boosting recessions. *Canadian Journal of Economics/Revue Canadienne d'Économique*, *47*, 1–34.

Nykodym, T., Kraljevic, T., Hussami, N., Rao, A., & Wang, A. (2016). *Generalized linear modeling with H2o*. Published by H2O. ai Inc.

OECD (2020). Consumer confidence index (cci) (indicator). http://dx.doi.org/10.1787/46434d78-en.

Organization for Economic Co-operation and Development (2020). Weekly hours worked: Manufacturing for the united states [hohwmn02usm065s]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/HOHWMN02USM065S.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research (JMLR)*, *12*, 2825–2830.

Peterson, R. (2017). Bestnormalize: A suite of normalizing transformations. R package version 3.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81–106.

Rosen, S., Salemi, P., Wickham, B., Williams, A., Harvey, C., Catlett, E., Taghiyeh, S., & Xu, J. (2016). Parallel empirical stochastic branch and bound for large-scale discrete optimization via simulation. In *2016 winter simulation conference (WSC)* (pp. 626–637). IEEE.

Rubaszek, M., & Serwa, D. (2014). Determinants of credit to households: An approach using the life-cycle model. *Economic Systems*, *38*, 572–587.

S & P Dow Jones Indices LLC (2020a). Dow jones industrial average [djia]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DJIA.

S & P Dow Jones Indices LLC (2020b). S & p 500 [sp500]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/SP500.

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*, 2507–2517.

Scholkopf, B., & Smola, A. J. (2001). *Learning with Kernels: Support vector machines, regularization, optimization, and beyond*. MIT press.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618–626).

Stavins, J. (2000). Credit card borrowing, delinquency, and personal bankruptcy. *New England Economic Review*, 15–30.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*, 1167–1179.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *Proceedings of the 34th international conference on machine learning-Volume 70* (pp. 3319–3328). JMLR. org.

Taghiyeh, S., Lengacher, D. C., & Handfield, R. B. (2020a). Forecasting model selection using intermediate classification: Application to monarchfx corporation. *Expert Systems with Applications*, Article 113371.

Taghiyeh, S., Mahmoudi, M., Fadaie, S., & Tohidi, H. (2020b). Fuzzy reliability-redundancy allocation problem of the overspeed protection system. *Engineering Reports*, Article e12221.

Taghiyeh, S., & Xu, J. (2016). A new particle swarm optimization algorithm for noisy optimization problems. *Swarm Intelligence*, *10*, 161–192.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*, 267–288.

University of Michigan (2020). University of michigan: Consumer sentiment [umcsent]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UMCSENT.

U. S. Bureau of Labor Statistics (2020). Unemployment rate [unrate]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UNRATE.

U. S. Census Bureau and U. S Department of Housing and Urban Development (2020a). Housing starts: Total: New privately owned housing units started [houst]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/HOUST.

U. S. Census Bureau and U. S Department of Housing and Urban Development (2020b). New private housing units authorized by building permits [permit]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/PERMIT.

U. S. Employment and Training Administration (2020). Initial claims [icsa]. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/ICSA.

Zintgraf, L. M., Cohen, T. S., Adel, T., & Welling, M. (2017). Visualizing deep neural network decisions: Prediction difference analysis. arXiv preprint arXiv:1702.04595.