Bayesian and Dynamic Bayesian Networks for Credit Risk Modeling in Buy Now, Pay Later

By

Sheyan Lalmohammed

An Undergraduate Thesis submitted as part of the

WHARTON RESEARCH SCHOLARS

Faculty Advisor:

Paul Sabin

Senior Analytics Fellow, Wharton Sports and Business Initiative & Lecturer, Statistics and Data Science Department

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

MAY 2025

ABSTRACT

This paper investigates the application of Bayesian Networks (BNs) and Dynamic Bayesian Networks (DBNs) on personalized credit risk assessment in the Buy Now, Pay Later (BNPL) industry. Using data from the Lending Club as a proxy for BNPL loans, the paper constructs multiple static and 5-period temporal network structures by discretizing borrower, loan, and macroeconomic characteristics. The use of score-based structure-learning methods allows for the identification of optimal network structures. A cost-sensitive bootstrap evaluation assesses model performance across multiple classification metrics. Results show that naive BNs offer high interpretability, BNs with learned structures capture important inter-dependencies in the data, and DBNs provide improved temporal inference, all while remaining computationally efficient for real-time decision making for a BNPL context. The findings suggest that using these models can serve effectively as a secondary check for default prediction flexible to the needs of the lender, balancing risk and operational constraints in short-term lending.

INTRODUCTION

This paper uses Bayesian Networks (BNs) and Dynamic Bayesian Networks (DBNs) as the primary underlying credit risk model for default risk in the Buy Now, Pay Later (BNPL) industry, examining the role BNs and DBNs can play in personalized credit risk assessment and the impact of temporal, macroeconomic, and industry data incorporation on the predictive accuracy and fairness of these models. Dynamic Bayesian Networks are an extension of Bayesian Networks whose primary purpose is to model temporal data and capture the probabilistic relationships between variables at each specified time step. Personalized credit risk assessment refers to the ability of a credit risk model to provide very specific predictions for credit risk based on the credit background and development of a specific individual. The predictive accuracy of BN and DBN models for credit risk assessments is the fraction of correct classifications they can make of defaults on BNPL loans. The Buy Now, Pay Later industry refers to the industry in which lenders provide short-term loans to be paid back in 4 to 6 installments of a value typically between $50 and $2000. This paper will contribute to the limited existing literature on credit risk in the Buy Now, Pay Later industry by examining the theoretical effectiveness of the implementation of temporal and network-based models on the ability of BNPL providers to make more profitable, less risky loaning decisions. Additionally, this paper will provide specific insights on model infrastructure for the effective use of macroeconomic and industry-level data for predictive accuracy in short-term loan models, which has previously been primarily explored only in longer-term loan formats.

Understanding the application of network-based and temporal models, such as BNs and DBNs, to short-term loan data in the BNPL industry is important both because the industry is dependent on making semi-reliable loans to generate reliable revenue and because network-based models provide more understandable and dependable predictions than most machine learning and statistical learning models in a computationally cheap way. If BNPL companies receive an increasing number of defaults or no defaults by using an unreliable or incomprehensible model structure, they void some of the assumptions about risk and the revenue structure on which the industry is built. Furthermore, if they use a computationally intensive method of determining credit risk, they risk losing sales of BNPL products due to the limited patience of BNPL consumers and the increased cost of providing their BNPL services. The BNPL industry operates significantly differently from the rest of consumer finance, with limited regulation, limited access to credit history, and more nuances in consumer financial protection, indicating the importance of providing effective modeling tactics of BN and DBN models separate from other areas of credit risk or consumer finance.

This paper will use data provided by Lending Club, a peer-to-peer lending platform, as a proxy for BNPL data. To use this data as a proxy, the data is cleaned in three steps. First, this will involve rigorously ensuring a subset of the previous data meets the credit profile requirements of a BNPL user. Next, the use of loan characteristics in the data will ensure that the loans provided in the data meet similar requirements for a BNPL loan, especially those of the value of the loan and the purpose of the loan. Finally, the temporal dataset is created by splitting the existing data on a quarter-by-quarter basis and making sure that each loan is expressed in its amount and possible late payment over time. This step will also involve the inclusion of macroeconomic data, including economic growth, unemployment, and inflation, on a quarter-by-quarter basis as reported by the Federal Reserve. After ensuring the validity of the data as a proxy for a BNPL firm, the analysis of the data first involves basic trial and error on the creation of a naive BN, a BN with structure learning, and a DBN structure using a subset of the numerous consumer and loan characteristics available in the data. Then there will be consideration of the effectiveness of the use of Partially Non-Homogeneous DBNs to model the more complex system under which BNPL firms operate. The results of the analysis are evaluated in the context of structural trends in the models, the general results and tradeoffs of each model type, and the ability to understand loaning decisions using the network-based structure of the model. In doing so, this paper will show that, in general, these models are highly interpretable in this context, work well as a secondary check for default, and can be used flexibly depending on the needs of a BNPL firm.

BRIEF REVIEW OF PREVIOUS LITERATURE

BNPL Industry, Consumer, and Loan Characteristics

To effectively consider the Dynamic Bayesian Networks as a modeling tool for credit risk modeling in the Buy Now, Pay Later industry, there must be an extended understanding of the Buy Now, Pay Later industry as a whole and the characteristics of credit risk, which begs the need for an efficient credit risk model. By formal definition, the Buy Now, Pay Later industry refers to a consumer financial sector that provides short-term installment-based loans to allow consumers to purchase retail items with little or no down payment. These loans are provided as an option to purchase an item at the point of sale. Buy Now, Pay Later providers use merchant partnerships as a way to boost their revenue throughout this process. To do this, they charge a merchant fee based on a merchant discount rate (MDR) of between 3% to 6%, which is notably higher than traditional credit products, in which MDR is 2% to 3%. This results in a merchant benefit to using BNPL providers of higher sales, conversion rates, and average order value (Equifax, 2022). The revenue from these merchants, along with late fees on provided loans, makes up the majority of the revenue of a BNPL provider. In this setup, the BNPL provider takes the credit risk as they pay the merchant the full value of the product minus the MDR.

Two recent papers by a United States Consumer Regulator Agencies (Consumer Financial Protection Bureau in September 2022 and March 2023; Philadelphia Federal Reserve Bank Consumer Finance Institute, Akana and Doubinko, 2024) established the characteristics and demographics of Buy Now, Pay Later users using data from the biggest BNPL firms as well as the primary trends seen in the industry. They first established that since 2019, the industry has grown more than 10x in size, increasing both the amount of loans given but also the risk of defaults on such loans by consumers using this BNPL. More importantly, these reports showed a fundamental difference between consumers and firms in the Buy Now, Pay Later industry compared to a traditional consumer finance industry (such as mortgage loans). The first is an imbalance in access by demographics compared to traditional finance products and the rate of approval. In terms of age, it is shown that around one-fifth of the United States's consumers have used BNPL and of those, the probability of such use is primarily focused on users under the age of 35 with over 48.5% of users falling in that age group while only 28% of nonusers fell in it. This is a sharp contrast to the average age of 35 for a first-time mortgage loan applicant (according to Experian). Also, when it comes to race and gender, there is a higher probability of female, Hispanic, and Black users on BNPL platforms compared to other races. This is again in sharp contrast to the approval of these groups lagging heavily in traditional consumer finance scenarios such as mortgages (Martinez and Kerchner 2021). Recent research has also shown that the true credit risk that BNPL firms are taking on could be significantly changed compared to traditional lenders like banks because of these characteristics in the average approval. It is indicated that, on average, there are key differences in credit delinquency by BNPL users compared to non-users at a statistically significant level. Research showed this had to do with their habits as BNPL users were substantially more likely to resort to working more, borrowing more, or reducing savings to meet monthly expenses. These users also generally report a greater inability to pay on time. This is partially due to liquidity constraints that BNPL users face compared to non-users. It is notable to state that with changes in regulatory policy between research released in 2022 and 2024, there has been a changing trend in demographics in the BNPL industry, with a greater proportion of white users compared to previously, but the general characteristics have stayed largely the same. Finally, the key disruption in the research showed up with a disconnect with the assumption for the use of BNPL products. Unlike mortgages or other traditional consumer finance products, users of BNPL do not state credit constraints as a reason for using the product, which is entirely contrasted by their higher probability of experiencing financial disruptions and higher credit use for monthly expenses. Each of the characteristics with regards to BNPL users is essentially non-existent in other traditional consumer financial industries, and therefore, the introduction of existing credit risk models to the BNPL industry seems not only inefficient for probabilistic inference but also illogical because of the introduction of a higher bias based on fundamentally different training data for such models.

The loans for BNPL providers also subsequently indicate an underlying higher risk to taking on such loans compared to traditional finance products. Firstly, there are some notable differences in the characteristics of the loans compared to other financial products. Firstly, BNPL loans do not provide access to collateral for a BNPL provider in the case that a loan is deemed to not have the ability to be paid back. In a recent paper published in a legal journal arguing for greater regulation of these loans (Mandell and Lawrence, 2023), it was discussed that BNPL loans are unsecured loans not tied to a specific tangible asset and that repossession may only occur in very rare cases, so the majority of the time, the BNPL provider does not retain ownership of the purchased items and puts these lost loans into bad-debt in their accounting. The other primary and most easily differentiating factor between most traditional financial products and BNPL loans is the value of these loans, which typically ranges between $50 and $1000 but can go up to nearly $5000 for higher-end luxury retail items. Unlike loans with specified purposes, these loans vary highly in value due to the nature of the use of such loans being for buying some consumer product, which can range from groceries to furniture to electronic devices. Also, the term conditions for these loans are different than those for typical loans. A central component of BNPL loans is that the payment schedule is set in terms of 4-6 payments, usually spread out in a span of less than 24 months.

Existing Credit Risk Models and Justification for Dynamic Bayesian Networks

There exists a vast amount of research on credit risk modeling across financial sectors. One research study (Bahtore et al., 2020) conducted a systematic review of over 136 research papers related to credit risk models based on machine learning. The researchers found that it is almost impossible to create a so-called one-size-fits-all model for all credit risk modeling. They found that the domain in which the model is applied creates key differences in the modeling approach. They also found that the majority of research, around 43%, focused on neural networks as an effective way to model credit risk. They stated that Artificial Neural Networks perform better than linear models when modeling credit risk. Another research paper (Li et al., 2022) found that credit risk modeling and prediction can be effectively done through multi-model fusion, such as a combination of Logistic Regression, Random Forest, and CatBoost. Previous studies clearly show that using probabilistic inference under uncertainty is efficient and accurate in a credit risk model. They also all indicate that network-based approaches, under some specified conditions, can actively improve model performance through the integration of real relationships and connections as well as a diversity of data.

The literature on Bayesian Networks applied to credit-risk scenarios is also robust. In an example of a Bayesian Network applied to banking data (Triki and Boujelbene, 2017), it was found that the primary reason that Bayesian networks were chosen to make default predictions for banks was for a clear graphical depiction of the interdependence of factors of default. The researchers made it clear that Bayesian Networks are more effective in creating decision systems within financial institutions. Another study (Nabende et al., 2021) considered Bayesian networks in the context of predicting credit risk on Ugandan Credit Contracts. The researchers found that prediction results from the Bayesian Networks were better than traditional classification methods, such as random forests, and that the prediction rate became more effective when considering more balanced data input to the models, such as a subset of the data. Also, a study (Morales et al., 2019) studied Bayesian Networks with application to credit risk analysis in Microfinance Institutions in Peru. The researchers found that Bayesian Networks performed better on the credit risk data in the last 5 years, at that time, compared to Neural Networks or Logistic Regression. These results confirm that Bayesian Networks can be an effective strategy for credit risk prediction based on historical data.

A very small subset of the previous literature focuses on Dynamic Bayesian Networks in credit-risk modeling scenarios. One paper (Shiguihara et al., 2021) summarized the last 20 years of development of DBNs, specifying that the structures of new DBN models are made with the intent that they remove the natural lack of discovery of information relationships that exist in machine learning black boxes in other models. The research also emphasizes deep specifics of the relationship of measured events within a system that is continuously changing. Furthermore, the research introduces a few of the most prominent DBN models that could be helpful in a credit risk modeling situation with uncertainty. Specifically, it introduces Partially Non-homogeneous Dynamic Bayesian Networks (NH-DBNs) with the advantage that they can be used with the assumption that parameters cannot be constant over time. This is particularly useful in the context of credit-risk modeling since borrowers' financial situation can change unexpectedly, and macroeconomic conditions that affect lending have been recently driven in new directions than basic economic theory would have previously suggested. Further research on NH-DBNs (Kamalabad et al., 2019) reveals that using partially NH-DBNs allows the model to operate on individual network interactions in which each interaction is based on a parameter, and the model can infer the nature of the parameter behavior from the data itself. Under the data availability in the BNPL industry and the implicit but changing connections between the parameters that exist, the research suggests that this system would more effectively work to identify the distinct subgroups of individuals (seemingly reliable borrowers that default, seemingly reliable borrowers that don't default, and borrowers who will default) that a BNPL firm is interested in.

Existing and Changing Understanding of Credit Risk Modeling for BNPL

The fundamental question of the approach to credit risk modeling in the BNPL industry comes down to finding a model that can effectively take BNPL-like data and translate it to accurate results for default prediction. While the general field of research related to credit risk research for BNPL products is extremely small, considering previous research brings up some very interesting points for further analysis for my paper. A recent research study of statistical learning models (Tas, 2023) applied to BNPL data from a provider in Turkey found that the LightGBM model was the best-performing model from the models of itself, Logistic Regression, Probit, and Random Forest. These preliminary results indicated that a Bayesian Network would be helpful since it has many similarities with a LightGBM model, especially in the way they attempt to understand the role of predictors in the decision for credit risk and the design of both models with specificity to efficiency in the calculation of a prediction. Also, a research study (Hardin and Ingre, 2021) found that it is possible to incorporate macroeconomic data into models for credit risk prediction in BNPL companies, and there is no single macroeconomic factor that stands out as most important in any model, but that real GDP growth, unemployment rate, and housing prices are usually among the most important predictors.

Further consideration of the characteristics of a Bayesian Network model indicates it clearly as a good fit for the basis of a model to predict general credit risk and, more importantly, to predict credit risk within the unique parameters provided by the BNPL industry. Recent studies (Leong, 2015; Liu et al., 2024; Lu et al., 2023) have first established that Bayesian networks address censoring, class imbalance, and real-time implementation issues in credit risk scoring. This is specifically relevant because the Buy Now, Pay Later industry contains many applicants with limited to no significant credit history. Furthermore, the consideration of class imbalance comes in when measuring the degree to which the default rate plays a role in the BNPL industry. Bayesian networks have a Bayesian classifier that can partition the data according to a class variable, which ensures that there is representation of all classes in the training data, making a less biased model. In terms of real-time updating, Bayesian networks are more efficient for scaling since they can take a general trained structure established originally and use subsequent data to change probabilistic weights and make more informed decisions. This is especially necessary in the BNPL industry, where the average decision for an approval or denial of a loan is made only in a few seconds. Furthermore, these research studies establish that Bayesian networks are particularly important because they require much less computational power. This is helpful for BNPL providers who can only afford a few seconds to decide to take advantage of the consumers' psychological bias related to BNPL spending which prompts possibly unhealthy spending habits but promotes greater use of the product for BNPL firms (Relja et al., 2023).

There are two extremely important takeaways from the previous research. The first, and most important, of these is that the research supports the consideration of a Dynamic Bayesian Network model as a possible more effective solution for a credit risk model in the BNPL industry. In all cases of previous research on credit risk within the BNPL industry, the researchers were unable to verify that the approach to modeling they took was the most effective. Therefore, there remains a key gap in the literature on the effectiveness of network-based models that use temporal data. This is a key gap that this solution fits and which is beneficial to this small field of research. Furthermore, the literature interestingly shows the possibility of beneficial use of external non-credit macroeconomic data on credit risk prediction. This fact drives the justification for the incorporation of macroeconomic indicators as part of the predictors to be used across time in the DBN model this paper uses.

BACKGROUND ON BAYESIAN NETWORKS AND DYNAMIC BAYESIAN NETWORKS

Before presenting the modeling and data decisions of this paper, there is the presentation of some basic notation and a very brief background on Bayesian Networks (BNs) and Dynamic Bayesian Networks (DBNs). In doing so, there is the adoption of a similar notation and structure to Leao et al. (2021) and definitions from Mihajlovic and Petkovic (2001).

Bayesian Networks

Bayesian Networks are directed Probabilistic Graphical Models (PGMs) which graphically represent the joint probability distribution over a set of random variables. For the purposes of this paper, Bayesian Networks will represent the joint probability distribution over the set of available personal, loan, and macroeconomic variables.

Notation and Example

In mathematical and theoretical terms, a Bayesian network, as presented in Leao, Madiera, Gromicho, M.D. Carvalho, A.M. Carvalho (2021), can be represented as a triple $\boldsymbol{B} = \{\boldsymbol{X}, \boldsymbol{G}, \boldsymbol{\theta}\}$ with the following characteristics:

- $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n})$ represents a vector of all personal, loan, and macroeconomic variables relevant to the credit risk profile of an individual and loan. As an example, consider a very basic credit risk model in the context of a Bayesian Network. Let $X_1 = Loan\ Amount$, $X_2 = FICO\ Score$, and $X_3 = Default$. Then X represents all possible nodes in the graph, including both independent (X_1, X_2) and dependent nodes (X_3). Note, in this definition, the description of dependent indicates that the variable is determined due to factors observed in the model, and independent indicates the variables existed independently and were exogenously determined outside of the model. Previous research has referred to these nodes as attributes of the Bayesian Network.
- $\boldsymbol{G} = \{\boldsymbol{X}, \boldsymbol{E}\}$ represents a directed acyclic graph (DAG) where each of the variables X_i is the random variable related to the credit risk profile of the loan and consumer. $\boldsymbol{E}$ represents the set of directed edges in the graph. Any dependent variable, as defined previously, must have some set of other variables, commonly referred to as the parents, which directly influence its value. In notation, the parents of X_i are written as $\boldsymbol{pa}(X_i)$.

Continuing the example of the simple credit risk model, let $E = \{(X_1, X_3), (X_2, X_3)\}$. Then there is a directed edge from loan amount to default and from FICO score to default. As a result, it is the case that $\boldsymbol{pa}(X_3) = \{X_1, X_2\}$ while $\boldsymbol{pa}(X_1) = \emptyset$ and $\boldsymbol{pa}(X_2) = \emptyset$.

- $\boldsymbol{\theta}$ is the set of all conditional probabilities that are required to describe the joint probability distribution of $\boldsymbol{X}$**.** For the case of the theoretical models that this paper adapts for credit risk modeling, the focus is placed on discretizing the attributes of the BN such that every variable X_i has at most r_i states. In the previous example, default has two states so that $X_3 \in \{0,1\}$. $\boldsymbol{\theta}$ is made up of the values given as θ_{ijk}**,** which are given as

$$\theta_{\mathrm{ijk}} = P\big(X_i = x_{ik} \big| \mathbf{pa}(X_i) = w_{ij}\big)$$

where $i \in \{1, \dots, n\}$, $k \in \{1, \dots, r_i\}$, and $j \in \{1, \dots, q_i\}$ where $q_i = \prod_{X_m \in \mathbf{pa}(X_i)} r_m$. This states that θ_{ijk} is the probability of some node X_i taking on a specific state k given the parents of X_i are taking on some state j which is one of many possible combinations of states of the parents. In the example credit risk model, if $X_1 \in \{100{,}200\}$ and $X_2 \in \{700{,}800\}$, $i = 3$, $k = 1$, $j = 1 \in \{1{,}2{,}3{,}4\}$ and

$$w_{3j} \in \{1 = (X_1 = 100, X_2 = 700), 2 = (X_1 = 100, X_2 = 800),$$
$$3 = (X_1 = 200, X_2 = 700), 4 = (X_1 = 200, X_2 = 800)\}$$

then let $w_{31} = 1$. The result is that

$$\theta_{311} = P\big(X_3 = x_{31} \big| \mathbf{pa}(X_3) = w_{31} = (X_1 = 100, X_2 = 700)\big)$$

which is the probability that default is the value of 1 given that the loan amount is 100 and the FICO Score is 700.

In general, in many probabilistic graphical models, and specifically Bayesian Networks, as stated in Leao et al. (2021), there is an assumption that given all the parents of a node, the node is conditionally independent to all other nodes in the graph. Then the result is that, using the definitions previously given, the joint probability distribution for all nodes in the BN is given as (using the chain rule):

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P\big(X_i \big| \boldsymbol{pa}(X_i)\big)$$

Dynamic Bayesian Networks

In the simplest of terms, Dynamic Bayesian Networks are effectively multiple Bayesian Networks that are connected across time. In doing so, DBNs are specifically designed as to handle temporal data. In particular, DBNs are designed to model how variables evolve, capturing the complex dependencies between variables at any given time step as well as the relationships between time steps. This connectivity allows for efficient inference and parameter estimation by exploiting the network's structured temporal decomposition.

Notation

Again following very similarly the notation from Leao et al. (2021), this paper defines a DBN as a pair $(\boldsymbol{B}, \boldsymbol{B}_{\rightarrow})$ where $\boldsymbol{B}$ is the base Bayesian Network, this may technically be derived from any time point but within the context of the model will represents the network observed structure at the initial time point, which is referred to as time point zero. It is defined as $\boldsymbol{B} = P(X[0])$ which represents the probability of the observed BN at the initial time point. $\boldsymbol{B}_{\rightarrow}$ represents the set of all transition networks. Each transition network can be defined as $\boldsymbol{B}_{\rightarrow}[\boldsymbol{0}, \boldsymbol{t}]$ for some discrete time period $\boldsymbol{t} \in \{\boldsymbol{1}, \dots, \boldsymbol{T}\}$. Each of these transition networks is defined as $\boldsymbol{B}_{\rightarrow}[\boldsymbol{0}, \boldsymbol{t}] =$

$\boldsymbol{P(X[t]|X[0,t-1])}$ where this represents the probability of the node values seen in the BN at timestep t given all the values of the nodes in the BNs from all previous timesteps. Then, using the chain rule, the DBN gives the joint probability distribution of the whole network over all possible values of the variables over the entire time period. This is written as

$$P(X[0,T]) = \boldsymbol{B}\prod_{t=1}^{T}\boldsymbol{B}_{\rightarrow}[0,t] = P(X[0])\prod_{t=1}^{T}P(X[t]|X[0,t-1])$$

Effectively, in this paper's study of DBNs in the context of credit risk, there are two directed connections between nodes that are important. These are defined and explained as follows:

- Intra-slice directed edges are those edges which exist within each BN at each time step. A simple way to think about these edges is that they have the same properties as the edges in the BN notation described previously. The importance of these directed edges is that they must showcase a relationship which remains consistent over time. For example, if it is the case that there is a directed edge from FICO Score to Default in the BN at time period 0, then there will be a directed edge from FICO Score to Default in time periods 1,2, and so on. Therefore, when considering the edges that should and should not exist in the initial BN, it is extremely important to find an optimal of a structure of the network so as to minimize the possibility of getting inaccurate probabilities of default due to dependency through multiple periods of time being miscalculated.
- Inter-slice directed edges are those edges which go from nodes in one time period to nodes in the next time step. It must be the case that these edges go only forward one step. This means a FICO Score node in time step 1 cannot connect to a Default node in time step 3. A general assumption made in homogenous DBNs is that these inter-slice edges, like the intra-slice edges, remain consistent over time. This means that any relationships

that are shown to occur between time step 0 and time step 1 should be replicated to occur between time step 1 and time step 2 and each pair of consecutive time slices after.

Structure Learning

A key component of this paper is to understand the optimal structure for a BN given credit risk BNPL data. This task is extremely complicated given the problem is *NP*-hard due to a large search-space of possible solutions (Berretta et al., 2018). To do this, there is the use of structure learning methods widely discussed in the literature. Structure learning methods aim to find the optimal set of edges for the graph. To do this, there are many popular algorithms which are generally split up into two categories of constraint-based and score-based. The primary one that this paper employs is Hill Climb Search, a score-based algorithm. Hill Climb Search is a value-based algorithm in a directed graph space and includes a heuristic search method that works greedily which can be used to learn BN structure (Adhitama and Saputro, 2022).

In using such a structure learning algorithm, there is a need to differentiate between each BN to determine which BN is a better fit for the data compared to another one. To do this, there is a reliance on a scoring function. A scoring function provides a quantitative measure of the network quality to the fitting of the data under the structure being considered. In this paper, there is an exploration of the use of multiple scoring functions. Each of these is briefly introduced as follows:

- **Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC):** BIC and AIC are cases of information-theoretic scores. These scoring functions try to quantitatively balance goodness-of-fit and model dimensionality and aim to avoid overfitting by using the data directly (Kitson and Constantinou, 2023). Both structural scores are relatively similar. AIC effectively evaluates the network structure by combining a maximum log-likelihood function with a penalty term which is proportional to the number of model parameters (Liu, Malone, and Yuan, 2014). BIC alternatively uses a logarithmic function which penalizes the model based on sample size. Note, Liu et al. (2014) found that BIC consistently outperforms other scoring functions, such as AIC, in practice. Also, it is the case that BIC, because of its use of the logarithmic penalty term, results in generally sparser networks than AIC. Therefore, this paper uses BIC as the primary scoring function when considering score-based network structure learning algorithms in the creation of the credit risk model.

- **Bayesian Dirichlet Equivalent Uniform (BDeu) and Bayesian Dirichlet Sparse (BDs):** BDeu is a scoring function that extends the Bayesian Dirichlet scoring function but allows for score equivalence and uniform priors which means it assumes that all networks are equally likely and will get the same score for two networks which have the same set of independence assumptions (Liu et al., 2014). BDs is a scoring function that adjusts the Dirichlet prior so that it can favor models that have conditional probability distributions that are sparse. It tends to favor models with simpler final dependency structure that captures the dependencies of the data (Scutari, 2016).

METHODS

Dataset and Sample Identification

In this analysis, archival data provided by the Lending Club will be used. The Lending Club is a consumer financial organization that publicly releases information regarding its peer-to-peer lending business. The dataset to be used is an archival dataset that contains both the accepted and rejected loan data from 2007 to 2018 for Lending Club. The dataset contains well over 2 million loans that have been given out. The dataset contains 150+ variables describing the nature of the loan as well as the credit characteristics of the individual. This includes employment status, annual income, home ownership status, debt-to-income ratio, credit delinquencies, credit inquiries, credit utilization, historical fail-to-pay counts, and credit accounts for credit cards and other credit instruments. The data also includes the nature of the loan being individual or joint, the loan amount, the term of the loan, the interest rate, the installment amount, an assigned grade of the loan, a sub-grade of the loan, the loan status, form of disbursement, the late fees on the loan, the balance on the loans if they are not yet complete, and the breakdown of the payment of principal and interest. The rich depth of the dataset is central to the creation of time series data from static data and the similarities drawn from the available data to that of BNPL-like data.

BNPL Characteristics Breakdown and Sample Selection Methods

Using the dataset of Lending Club personal loans to narrow down relevant loans which are BNPL-like started with the consideration of loans that fall in the $1000-$5000 range. Unlike the larger loans on Lending Club, according to the summary located in columns related to the use and justification of the loan, these smaller value loans, like BNPL use loans, were used to finance purchases or for personal expenses on a monthly basis rather than for significant personal purchases, such as that in the case of a mortgage loan for example. These loans make up roughly 295,000 of the existing loans in the data from the Lending Club and thus become our primary set of data from which there will be a creation of the relevant samples to train a Bayesian network. Using the primary set of loans that meet the BNPL characteristics, there is consideration for those loans that provide valuable data relevant to the prediction of default probability and default itself. Similar to the conclusion reached in previous research using the Lending Club data (Hou, 2020), the loans that can be used effectively are those whose status is listed as “Fully Paid” or as “Defaulted”, “Charged Off”, “In Grace Period”, and “Late”. The former refers to loans that have been completely paid back, while the latter refers to those that have defaulted. There are loans classified as “Current”, but using these loans as part of the analysis would be misleading to the truth of the data regarding individual repayment ability. Thus, loans marked with a status of “Current” are removed from the primary set of data. This results in approximately 184,000 samples of loans that meet the criteria.

To create relevant proxy data that accurately reflects those values represented by BNPL-like users, there is a reference to the January 2025 report by the Consumer Financial Protection Bureau (CFPB). In this report, the CFPB recaps the most prominent characteristics of BNPL users on moderate and extreme scales according to the most recent data collected in a mass survey in 2022. In this report, the CFPB mentions that there are two primary characteristics under which default, credit history, and other credit-related and behavior-related characteristics can be classified. These two are the age group and FICO score category. For the age group, the CFPB decided the primary age-group buckets are given as 18-24, 25-33, 34-40, 41-50, 51-64, and 65+. As for the FICO score categories, the relevant categories are given as deep subprime (300-579), subprime (580-619), near prime (620-659), prime (660-719), and super-prime (720-850). All aggregate statistics to the data relevant to these two categories reported by the CFPB are those used for the creation of the sample for BNPL-like data. These aggregate statistics can be seen in the tables presented from the CFPB report, which have been placed in the Appendix. In constructing the dataset, there is an inference in the age of the consumer, as this information is not explicitly included in the publicly released data, using the date of the earliest credit account.

There is a rough assumption similar to that in previous research (Serrano-Cinca and Gutiérrez-Nieto, 2016) that a borrower's age can be inferred by subtracting the date of the requested loan from the earliest credit line and adding roughly 18-23 years to it as according to the Federal Reserve (2023), over 70% of Americans have a credit card by this time. This inferred age is used to create the buckets for age to relate to the BNPL data. Also, from a qualitative perspective, this decision makes sense, considering the acceptance of use for a credit platform like Lending Club is an extreme case for someone without a longer credit history. As for the FICO-score category, there is a much stronger assumption made. Specifically, the assumption made is that the borrowing profile distribution of those borrowers using personal loans for BNPL-like purposes (because BNPL did not exist as an alternative option to the use of these personal loans) is directly comparable to the current distribution of BNPL users. In making this assumption, the data is relevantly split up into categories directly defined by percentages from the CFPB report on the score category breakdown which can be found in **Table 2** of Section **A.1** in the Appendix**.** This is done in order of the existing FICO scores available in the data.

Primary Samples of Interest

After aggregating the values of the FICO scores to buckets relevant to the data sample, the sample from previously is modified to create multiple samples which present interesting takeaways about the effectiveness of the data under the current understanding of the BNPL consumer profile. To do this, there is no strict assumption on the age group profile of the BNPL users. While the BNPL report from the CFPB provides some details, as in **Table 9** of section **A.1** in the Appendix, there are no strict details given on a straightforward breakdown of total BNPL transactions by age group. Thus, an assumption is applied that by capturing multiple points of possible BNPL age-demographic breakdowns along with the true breakdown given by the age calculation previously described, there is the ability to capture the case that an incorrectly assumed age group breakdown will be relevant to the results of a network-model structure and effectiveness. In doing so, the definitions of the breakdown are summarized in Table 1:

Table 1: Age-Group by Sample - Breakdown (%)

	Original	Sample 1	Sample 2
18–24	8.83%	30%	25%
25–33	49.7%	30%	35%
34–40	25.4%	20%	25%
41–50	12.5%	10%	10%
51–64	3.28%	10%	5%
65+	0.28%	0%	0%

To keep logistically certain that age-group movement in these artificial samples occurs consistently, the new age groups for each sample previously defined are adhered to as close as possible by only reassigning groups to adjacent levels depending on which group is closer in value to the bound of the next age group. The result of doing so is established in Table 2:

Table 2: Age-Group by Sample – Post-Balancing Counts

	Original	Sample 1	Sample 2
18–24	16262	55112	45926
25–33	91581	55112	64297
34–40	46788	36741	45926
41–50	23007	18370	18370
51–64	6045	18348	9164
65+	507	0	0

Note, It may very well be the case that the existing (original) profile of age in the data is relevant and correct for a BNPL firm, but this measure serves as a precaution if that may not be the case.

Extending Static Data to Dynamic Components

With the basic samples created, the data relevant to training and testing the Bayesian Networks of interest (without macroeconomic variable incorporation) has been fully discussed. For DBNs however, it is the case that the static loan data is transitioned to be made dynamic with a varying set of consistent values from each time to the next. In the context of this problem, BNPL is most commonly found in a pay-in-four structure where no upfront payment is made at the initiation of the loan, and there are four payments made of equal size up to 4 periods away. For the purposes of this analysis, the assumption is that the payment occurs quarterly, so it takes one year for the BNPL firm to recoup its capital. Multiple variables are extended to five time steps (the initial and the four additional payment time steps). Each one and the assumed logic for extension is discussed in detail as follows:

- **Default Indicator**: The time period for default for a loan is chosen randomly to exist at one of the four time periods of payment if the individual loan was defaulted on. This is a relatively sound assumption that can be made since the goal is predicting the value of the default at time period 4, and the intermediary nodes simply provide a framework for parameter estimation of the DBN.
- **Total Payment**: The total payment dynamic variable captures the progress made on the repayment of the loan. It assumes that, as is common in loans, the payment is fixed at some amount which is one-fourth of the total loan amount. Then progress is achieved towards the total payment if the borrower does not default, and no progress is made if the borrower does default during the time period of the payment. This value starts at 0 and should end at the value of the total payment.

- **Revolving Balance**: Like the total payment, revolving balance captured the balance of outstanding loans for the consumer, and this value is made dynamic by decreasing with respect to time at a proportional level when payments are made, and the borrower does not default on the loan at the time period and going forward.
- **Last Payment Amount**: This captures the true last payment value given the true value of the last payment in the existing data. It is meant to indicate the amount per period that the consumer is paying towards their loan and is heavily influenced by the installment on the loan they must pay.
- **Fico Bucket**: As previously discussed, FICO buckets are crucial to understanding the way the general credit risk profile of an individual develops over time. As a result, the bucket in which an individual falls should change depending on their ability to effectively pay back their BNPL-like loan. As a result, this paper assumes (for illustration purposes of a real effective change in borrowing behavior on FICO score) that those that default had falling FICO scores after defaulting with up to 30-100 points loss due to the high values of the loans and those that didn't default could gain up to 30-100 points or the maximum FICO score. At each point in time, the FICO buckets by percentage are calculated as described previously in the dataset section of the paper. In doing so, the FICO bucket an individual falls in can change over time based on known characteristics of their repayment behavior.
- **Revolving Utilization:** Revolving utilization is made dynamic by considering the value of the dynamic revolving balance and dividing it by the static total revolving credit limit in which case the result is the dynamic utilization of the individual at that period.

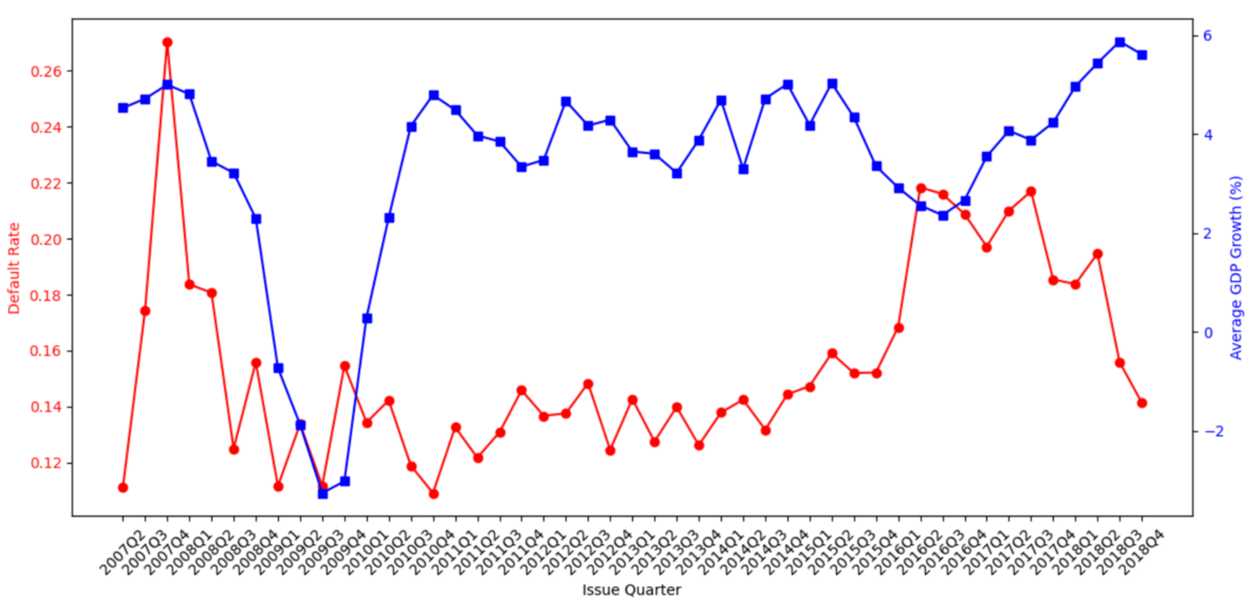
- **Outstanding Principal:** Like total payment, outstanding principal is made dynamic by proportionally decreasing the remaining principal to be returned on the loan when individuals choose not to default and leaving it unchanged when the default does occur on a loan.
- **Delinquency Count:** The delinquency count keeps track of the repayment behavior of the consumer. If the individual defaults on the loan, their delinquency count increases by 1, and if they do not default on the loan, then the delinquency count remains consistent and does not change.

In creating these dynamic variables, the focus was primarily on the basic interpretability of the dynamic values and capturing the credit risk evolution of the consumer over the one-year time frame. From a theoretical perspective, the absolute value of the changes has little importance in the evaluation of a network-based credit risk model based on these attributes. Instead, more important is that these variables accurately capture the intermediate changes expected in the individual's credit risk and the loan repayment based on the outcome of the loan known to be true.

Incorporation of Lagged Macroeconomic Variables

Based on previous literature (Hardin and Ingre, 2021) in credit risk modeling with network-based models, it can be extremely important to include lagged macroeconomic variables that capture the underlying credit risk associated with external changes to the economic environment that cause increases in default rates. In doing so, this paper takes on five lagged macroeconomic indicators, which are GDP growth, housing prices, unemployment rate, inflation rate, and the federal funds rate. GDP growth, housing prices, and the unemployment rate were previously shown to have a large and notable effect on default rates for alternative loaning environments, such as mortgages. The inflation rate and federal funds rate act as additional information to lending practices and price levels which may impact GDP growth and housing prices and provide additional information regarding monetary policy changes. Each of these indicators is taken from the St. Louis Federal Reserve Database and summarized according to the mean over the relevant quarterly time periods. They are then matched in a lagged format to the period when the loans are made. For example, for a loan made in Q3 of 2014. The indicators relevant are the macroeconomic indicator values for Q2 2014, Q3 2014, Q4 2014, Q1 2015, and Q2 2015. These represent time steps 0 through 4 respectively in the data. To illustrate a meaningful relationship, consider Figure 1, which shows the development of the default rate at the initial time period at each quarter compared to the lagged GDP growth in that time period:

Figure 1: Default Rate and Average GDP Growth by Issue Quarter (Time Slice 0)



Visually, an inverse relationship post-2009 with the default rate could exist, which is relevant to the credit-loaning decisions of the lender. This behavior indicates that, as expected, GDP growth may play an integral role in the default probability in each quarter and for specific loans. More information about the remaining indicators' relevance is presented in a similar graphical form in section **A.2** of the Appendix.

Justification of Aggregate Profile as BNPL-Like

With a completed and transformed dataset, there is the need for justification as to why the conclusions from such a dataset may be generalizable to the BNPL industry and that it truly represents BNPL-like data. To do this, there is a reference to the aggregate profile of the consumer demographics in the context of loan assignment and default probability. This paper assumes that BNPL companies, as is reported by the CFPB, only have access to basic attributes available in a soft credit check report and not many sub-details. Thus, for the static case of modeling a network to use these attributes, the paper assumes that BNPL companies can access credit scores, credit utilization, open credit accounts, total credit accounts, and general delinquencies over the past short-term and long-term periods. They also have some basic demographic and details to the short-term loan, which are the age, employment length, and homeownership status of the borrower along with the purpose of the loan. Finally, an important assumption is that, like most credit-providing institutions (Federal Deposit Insurance Corporation), a BNPL firm provides some sort of grade to the loan using the characteristics, which impacts its ability for default in the context of all other available information. This builds on the fact that the Bayesian and Dynamic Bayesian Network models considered for default risk provide additional information to the importance. Another important fact about the context of these available data parameters is that they take advantage of a benefit of Bayesian Networks in that they can create predictions without having all the available information. Therefore, it is the case that the resulting BN and DBN models presented later in this paper can still be applied to available credit and loan information.

To make the argument regarding validity of the data, there is first the presentation of the expected default risk profile of the consumers taking out these loans in the dataset compared to that which would be expected had it been BNPL-like consumers. In summarizing the data, it is found that approximately 14-15% of the data includes defaulted loans where it is assumed each loan is originated by a different individual, making the default rate around this value. Comparing this to the true default rate for BNPL is irrelevant because of the true fundamental difference in the structure and payment profile of the loans. However, it is notable to state that this value is very close to that of the charged-off amount to the credit card amounts for these consumers which is shown in **Table 10** of section **A.1** of the appendix. Therefore, it may be the case that the same type of consumers is using these credit products as the BNPL products.

Additionally, the aggregate credit risk profile and diversity of the accepted loan consumers looks relatively like what would be expected from a BNPL firm. To illustrate this, consider the summary of the data characteristics in the following tables:

Table 3: FICO Bucket Breakdown by Age Group

Age Group	FICO Bucket	Count	Percentage (%)
18–24	Deep Subprime	7,104	43.68
	Near Prime	2,181	13.41
	Prime	2,805	17.25
	Subprime	3,001	18.45
	Super-prime	1,171	7.20
25–33	Deep Subprime	43,823	47.85
	Near Prime	11,355	12.40
	Prime	11,517	12.58
	Subprime	16,707	18.24
	Super-prime	8,179	8.93
34–40	Deep Subprime	22,159	47.36
	Near Prime	5,744	12.28

	Prime	6,080	12.99
	Subprime	8,128	17.37
	Super-prime	4,677	10.00
41–50	Deep Subprime	10,608	46.11
	Near Prime	2,839	12.34
	Prime	3,067	13.33
	Subprime	3,823	16.62
	Super-prime	2,670	11.61
51–64	Deep Subprime	2,756	45.59
	Near Prime	669	11.07
	Prime	829	13.71
	Subprime	1,020	16.87
	Super-prime	771	12.75
65+	Deep Subprime	220	43.39
	Near Prime	60	11.83
	Prime	69	13.61
	Subprime	97	19.13
	Super-prime	61	12.03

Table 4: Default and Fully Paid Loans by Age Group

Age Group	Count Paid	Count Default	% Paid	% Default
18–24	13,158	3,104	80.91	19.09
25–33	75,884	15,697	82.86	17.14
34–40	39,393	7,395	84.19	15.81
41–50	19,413	3,594	84.38	15.62
51–64	4,950	1,095	81.89	18.11
65+	371	136	73.18	26.82

Table 5: Credit Grading by Age Group

Age Group	Grade	Count	Percentage (%)
18–24	A	1,318	8.10
	B	4,148	25.51
	C	5,771	35.49
	D	3,329	20.47
	E	1,229	7.56
	F	392	2.41
	G	75	0.46
25–33	A	13,543	14.79
	B	27,522	30.05
	C	29,453	32.16
	D	14,823	16.19

	E	4,817	5.26
	F	1,226	1.34
	G	197	0.22
34–40	A	8,280	17.70
	B	15,462	33.05
	C	14,124	30.19
	D	6,499	13.89
	E	1,949	4.17
	F	410	0.88
	G	64	0.14
41–50	A	4,650	20.21
	B	7,942	34.52
	C	6,469	28.12
	D	2,872	12.48
	E	852	3.70
	F	204	0.89
	G	18	0.08
51–64	A	1,348	22.30
	B	2,096	34.67
	C	1,661	27.48
	D	707	11.70
	E	175	2.89
	F	50	0.83
	G	8	0.13
65+	A	122	24.06
	B	174	34.32
	C	129	25.44
	D	60	11.83
	E	17	3.35
	F	4	0.79
	G	1	0.20

From the above tables, there are a couple of key characteristics to point out. Firstly, the evidence suggests that age group and default are not directly related, which is a true statement and thus correctly identifies the relationship as not being directly correlated. This is seen with the true BNPL data in **Table 5** in section **A.1** of the Appendix. On top of this, the aggregate grading scheme, used as an original check on the individuals before using a BN or DBN as introduced in this paper, shows that there is a relationship between grade and age group. Generally, lower age groups receive worse grades on the loans, and thus, age is implicitly captured as a factor in that grading scheme. On top of this, as expected, it is relatively straightforward to tell that generally, those with lower age groups also have higher percentages of borrowers falling into lower FICO score buckets. This is clearly a good sign that the data realization is consistent with how the FICO score considers credit history length. Additionally, in considering the summary statistics of the data, the average loan size of the BNPL loans is relatively comparable to the average installment size of the personal loans. This is relevant since this means that when looking at payment cases for higher-end products, considering the installment size may give more relevant information than even considering the loan amount itself since the difference between the two installment sizes for the personal loans from the Lending Club and that from the BNPL providers is smaller than the difference between the loan amounts. Also, as an additional point, the data here ends in 2018. Thus, it is notable to point out that in looking at BNPL consumers during that period, the default rates of these consumers on other credit products, like credit cards, is similar to the data of the default rates by age group seen in the data sample as summarized above. In attaining all these characteristics, this paper assumes that the data cleaned and shown above is somewhat representative of what the aggregate risk profile of consumers and diversity of loans would look like to a standard BNPL firm operating loans for a wide variety of products and services. This is a very strong and central assumption to the results discussed in the rest of this paper.

Discretizing the Data

As mentioned in the background on Bayesian Networks, this paper is focused on the use of discrete Bayesian Networks. Thus, it is the case that continuous variables, such as loan amount, interest rate, and macroeconomic indicators, must be discretized to a countable number of states. This paper does not make any strong assumptions about the parametric nature of these variables, so unlike using an approach that assumes the continuous variables follow some sort of distribution, there is only an assumption that these variables may be discretized to a smaller number of categories. This directly reduces computational complexity, which is exponential in the number of states for a discrete Bayesian Network (Kungurtsev et al., 2024), and is central to the estimation of default risk which is computationally efficient and justifiable for a BNPL firm. Due to computational constraints with an increased number of possible states and to obtain results for a variety of dependence connections, this paper assumes that continuous variables may be discretized in up to three states. This assumption is important for three reasons. The primary reason is that in doing so, the relative amount of possible states for the conditional probability distribution in the discrete Bayesian Network case is reduced to a factor of three to the power of the number of available continuous variables multiplied by the number of discrete states in each of the categorical variables. The second is that the models discussed in this paper, in both a theoretical and practical sense, are meant to provide only an additional level of credit risk protection. In assuming previous checks of credit risk have been completed and a subset of would-be accepted individuals are being considered to be given the loan, the model's reliance on only some categories of the continuous data adds a soft check to avoid losses incurred from default. The final factor in this decision has to do with consumer behavior. BNPL firms accept and reject loans immediately in the moment they are requested by a consumer. In providing additional time required to purchase an item using a BNPL loan, there is a risk of possibly changing the psychology driving the purchase in the first place and cutting into the revenue-making activities of the BNPL firm (Ang and Maesen, 2024). Thus, reducing the time of computation of the network-based models and their predictions is central to the decision for a minimal number of discretized states for continuous variables. Also, in making this assumption, the computational constraints were kept within the bounds of available resources for this paper.

The algorithm applied to the existing set of continuous variables is the Kbins discretize algorithm using a Kmeans clustering approach. In doing so, each column is treated. The algorithm assumes that each continuous variable must fall into one of three bins. These bins are assumed to be categories but are not given a categorical name. The algorithm finds three centroids that minimize the sum of squares for the continuous data, then uses those centroids to find cutoff points that exist halfway consecutively between the centroids and uses those cutoff points to assign the continuous variable to one of the bins (AixCAPE, 2016).

Imbalanced and Cost-Sensitive Data Approach

The interest in the network model performance is in its ability to correctly classify those values that are likely to default because this model approach is secondary to preliminary credit risk screening algorithms. A key part of this is using the model to make classification decisions regarding the default status of an individual, particularly with the predicted probability that they will default on the loan. Note, there is an imbalance between the cost of misclassification. In the case that the model predicts that the loan will be defaulted on but it is not defaulted on, the BNPL firm has lost no value. Instead, they actually profited from the fee that they charged the merchant. However, if it is the case that the model predicts the loan will not default, and it does default, then the BNPL firm is forced to write that charge off as a loss. Therefore, it is the case that there must be consideration given to the cost-sensitive learning approach taken when considering what loans should be classified as default or non-default. In particular, a bias is introduced into this model as the model is shown an imbalanced dataset of defaulting to non-defaulting loans, making the network-based model increase the weight of non-default probabilities beyond the true probability for any given loan.

In doing so, this paper uses a method suggested by Elkin 2001, in which the author states that generally changing the balance of the default to non-default classes has little effect on the resulting classification model when considering Bayesian or decision tree learning methods. The author suggests that one can empirically change the value of the optimal decision threshold p^* so as to classify all those samples as a default that meet the criteria that $p \geq p^*$ where $p = P(Default = 1|X)$. As previously represented in notation, *X* represents the other variables in the data. This paper leverages this finding by using multiple thresholds in the analysis with the goal of indicating various points of results and the tradeoff between the performance of the model, which is shown to suffer when underestimating the true probability of default from the Bayesian Network. In doing so, this also keeps away from making the assumption that the trained BN or DBN must be reliant on a specific mechanism to training.

Interval Calculation of Model Performance Indicators

As mentioned in the previous section, there is a need to accurately assess the performance of both BN and DBN models as suitable options for use by BNPL firms. To do this, first consider the metrics by which the results of the classification task are considered within the context of this paper and usually within the field of prediction as a whole:

- **True Positives, True Negatives, False Negatives, False Positives**: Assuming that there were two categories named true and false, True Positives (TPs) are those values that are predicted to be true by the model and are indeed true. True Negatives (TNs) are those values that are predicted to be false by the model and are indeed false. False Negatives (FNs) are those values that are predicted to be false but are actually true. False Positives (FPs) are those values that are predicted to be true but are in actually false. This paper uses the False Negative Rate (FNR) and False Positive Rate (FPR) as relevant indicators of model performance when it comes to default and non-default classification.

- **Precision and Recall**: Precision is defined as $\frac{TP}{TP+FP}$ which is effectively used to measure the ratio of correctly predicted values of the case of positive class when considering two-class cases, such as in default scores. For this paper, the precision of both the default and non-default classes is considered. Recall is defined as $\frac{TP}{TP+TN}$ which is meant to capture the fraction of the predictions of the positive class that are correct. Again, for the purposes of this paper, the recall of both the default and non-default classes is considered.
- **F1 Score**: The F1 score is defined as $\frac{2*Precision*Recall}{Precision.+Recall}$ which, according to Hossin and Sulaiman 2015, represents the harmonic mean between precision and recall. Using the value of precision and recall for both default and non-default classes, metrics of both F1 score for default and non-default are considered.
- **Classification Accuracy**: Classification accuracy is simply defined as the fraction of correct predictions to all predictions made by the classification model. It measures the correctness of the model in an absolute sense.

In attempting to compare such metrics across models and accurately capturing the distribution of these metrics without making parametric assumptions, the paper employs a bootstrap approach. For each case of calculating the confidence intervals of the previously mentioned metrics, there is a random sample of the data selected as the test set of data, while the remaining data is used to train the network-based model. Then, using the test set of data, a set of predictions are made and the indicators are calculated. This process is iteratively repeated a set number of times, and using the results of all iterations, the means are calculated directly while the bootstrap confidence intervals are computed using the percentiles of the results. Note, the goal of this paper is not to determine whether the intervals or mean values of these statistical methods falls warrants the use of the model at all, but rather to use these aggregate statistics on the classification accuracy and thus the underlying accuracy of the conditional probability distribution, as a mechanism to assess the performance of various structures of network-based models.

ANALYSIS

Naive Bayesian Network Approach

The first type of model this paper considers for BNPL lenders and the simplest of all models discussed within the context of discrete Bayesian Networks is the Naive Bayesian Network. Simply defined, it is a directed graph with edges outgoing from all nodes except the variable that is to be predicted into the variable to be predicted. These are the only edges in the graph. In using the naive approach, this paper considers four types of models. The first of these models is based solely on the personal characteristics and motivation of the individual requesting the loan. The second is based solely on the credit profile of the individual borrower. The third uses only lagged macroeconomic indicators in the initial time period to predict the default. The final one uses a mixture of variables from each of the first three models. The motivation behind the development of such models is that they are the most direct and simple type of network-based models for Bayesian Networks as they simplify the relationship for any nodes in the graph to at most one edge, either incoming or outgoing from the node. While they may not always be the most computationally efficient while being discretized, they are an extremely interpretable type of model, though they lack showing the true complex relationship between many of the observed characteristics that affect the credit risk of an individual and a loan to the lender.

Personal Characteristics-Based Bayesian Network

The first model this paper considers is that which solely uses the personal characteristics of the individual along with the previous information regarding a grading scheme for the loans expected to be accepted as valid loans. This again aligns with the assumption to this credit risk model as secondary to screening applicants for BNPL-like loans in order to reduce losses caused by default. To do this, let this Bayesian Network be defined by $\boldsymbol{B_p} = (\boldsymbol{X_p}, \boldsymbol{G_p}, \boldsymbol{\theta_p})$**.**

- $\boldsymbol{X_p}$ represents those characteristics of the individual that are relevant to the default setting. For the exemplification purposes of this paper, these characteristics include age group, employment length, purpose, homeownership, and the prespecified grading of the loan. Included in these characteristics is the default indication variable. Formally,

$$\boldsymbol{X_p} = (AgeGroup, EmpLength, HomeOwnership, Grade, Purpose\ Default)$$

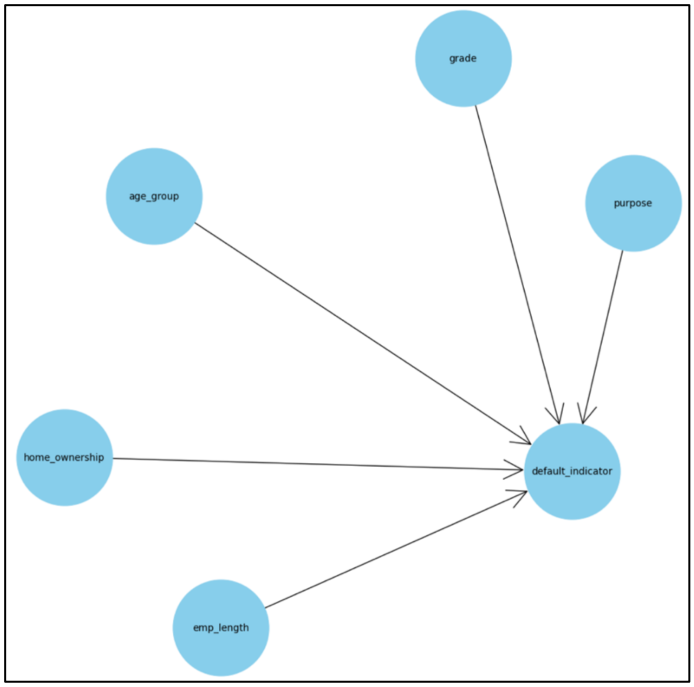
- $\boldsymbol{G_p}$ represents the graphical model made up of the directed edges $\boldsymbol{E_p}$ which is defined as

$$\boldsymbol{E_p} = [(AgeGroup, Default), (\ EmpLength, Default),$$

$$(HomeOwnership, Default), (Grade, Default), (Purpose, Default)]$$

  The nodes $\boldsymbol{X_p}$ are defined as previously. In doing so, the assumption of the graphical structure of this mode is shown in Figure 2:

Figure 2: Network Structure of Naive BN based on Personal Characteristic



- $\boldsymbol{\theta_p}$ **,** as previously defined**,** is the set of all conditional probabilities that are required to describe the joint probability distribution of $\boldsymbol{X_p}$ where $\boldsymbol{x_p}$ is one of 38,808 possible combinations of states for the 5 nodes with directed edges into the default. To estimate the parameters of $\boldsymbol{B_p}$, there is the use of Bayesian estimation with a Dirichlet Uniform prior (BDeu). This is a standard estimation technique in the literature used uniformly throughout the rest of the modeling for the various BNs considered in this paper. Note, an empirical test of changing the equivalent sample size of the estimator had no significant effect on the intervals observed. The resulting conditional probability distribution tables for the example of the original sample on one run of the model are given in section **A.4** of the Appendix.

Following these steps and using the sampling and interval calculations mechanisms previously discussed in the methods section, the results are presented on the original dataset sample with the thresholds as discussed in the methods section:

Table 6: Mean and 95% Intervals for Naive Personal BN

Threshold	Accuracy	FPR	FNR
0.15	0.5412 (0.5246–0.5575)	0.4765 (0.4602–0.4953)	0.3636 (0.3316–0.4013)
0.20	0.6720 (0.6591–0.6864)	0.2803 (0.2641–0.2938)	0.5817 (0.5346–0.6265)
0.25	0.7648 (0.7516–0.7773)	0.1303 (0.1178–0.1461)	0.7909 (0.7530–0.8274)
0.30	0.8047 (0.7925–0.8187)	0.0646 (0.0552–0.0760)	0.8917 (0.8683–0.9175)
0.35	0.8202 (0.8089–0.8293)	0.0379 (0.0313–0.0467)	0.9387 (0.9196–0.9594)
0.40	0.8243 (0.8126–0.8367)	0.0309 (0.0248–0.0379)	0.9496 (0.9316–0.9666)
0.45	0.8261 (0.8153–0.8431)	0.0251 (0.0193–0.0316)	0.9580 (0.9397–0.9717)
0.50	0.8374 (0.8251–0.8509)	0.0061 (0.0034–0.0089)	0.9904 (0.9819–0.9963)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8850 (0.8705–0.8970)	0.5235 (0.5047–0.5398)	0.6578 (0.6428–0.6707)
0.20	0.8681 (0.8553–0.8826)	0.7197 (0.7062–0.7359)	0.7869 (0.7760–0.7977)
0.25	0.8535 (0.8412–0.8645)	0.8697 (0.8539–0.8822)	0.8615 (0.8518–0.8702)
0.30	0.8482 (0.8372–0.8607)	0.9354 (0.9240–0.9448)	0.8897 (0.8820–0.8983)
0.35	0.8457 (0.8371–0.8556)	0.9621 (0.9533–0.9687)	0.9001 (0.8933–0.9057)
0.40	0.8450 (0.8331–0.8576)	0.9691 (0.9621–0.9752)	0.9028 (0.8957–0.9104)
0.45	0.8429 (0.8317–0.8577)	0.9749 (0.9684–0.9807)	0.9041 (0.8974–0.9143)
0.50	0.8414 (0.8295–0.8547)	0.9939 (0.9911–0.9966)	0.9113 (0.9039–0.9193)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1999 (0.1794–0.2233)	0.6364 (0.5987–0.6684)	0.3041 (0.2787–0.3340)
0.20	0.2191 (0.1972–0.2400)	0.4183 (0.3735–0.4654)	0.2874 (0.2620–0.3169)
0.25	0.2322 (0.1977–0.2685)	0.2091 (0.1726–0.2470)	0.2199 (0.1884–0.2532)
0.30	0.2400 (0.1858–0.2901)	0.1083 (0.0825–0.1317)	0.1490 (0.1142–0.1795)
0.35	0.2319 (0.1642–0.2980)	0.0613 (0.0406–0.0804)	0.0968 (0.0650–0.1248)
0.40	0.2340 (0.1583–0.3131)	0.0504 (0.0334–0.0684)	0.0829 (0.0555–0.1108)
0.45	0.2410 (0.1623–0.3287)	0.0420 (0.0283–0.0603)	0.0714 (0.0484–0.1016)
0.50	0.2319 (0.0952–0.4193)	0.0096 (0.0037–0.0181)	0.0184 (0.0072–0.0344)

Note, the results from the other samples are presented in section **A.4** of the Appendix. The notable idea to point out is that the intervals for all samples overlap which indicates, as expected, the age group, in the context of all other available personal information, has little effect on the default since it is most likely the type of individual credit and spending habits that draw consumers to this loan format and a wholistic picture of purpose and employment, and homeownership play larger roles in dictating this.

Loan and Credit Characteristics-Based Bayesian Network

The second model this paper considers is that which uses the loan characteristics along with the credit profile of the borrower. Let this Bayesian Network be defined by $\boldsymbol{B_{LC}} = (\boldsymbol{X_{LC}}, \boldsymbol{G_{LC}}, \boldsymbol{\theta_{LC}})$**.**

- $\boldsymbol{X_{LC}}$ represents those characteristics of the loan and credit background that are relevant to the default risk. For the exemplification purposes of this paper, these characteristics include loan amount, FICO score bucket, interest rate, installment, and the public record of bankruptcies, revolving balance, revolving utilization, total credit accounts, annual income, delinquencies in the last 2 years, credit inquiries in the last 6 months, and open credit accounts. Included in these characteristics is the default indication variable. Formally, this is defined as

$$\boldsymbol{X_{LC}} = \begin{pmatrix} LoanAmt, FICOBucket, IntRate, Installment, PubRec, RevBal, \\ RevUtil, TAcc, AnnualInc, Delinq2Y, Inq6M, OpenAcc, Default \end{pmatrix}$$

- $\boldsymbol{G_{LC}}$ represents the graphical model made up of the directed edges $\boldsymbol{E_{LC}}$ which is defined as

$$\boldsymbol{E_{LC}} = [(LoanAmt, Default), (\ FICOBucket, Default), (\ IntRate, Default),$$

$$(Installment, Default), (PubRec, Default), (RevBal, Default), (RevUtil, Default),$$

$$(TAcc, Default), (\ AnnualInc, Default), (\ Delinq2Y, Default), (\ Inq6M, Default),$$

$$(\ OpenAcc, Default)]$$

The nodes $\boldsymbol{X_{LC}}$ are defined as previously. In doing so, the assumption of the graphical structure of this mode is shown in Figure 3:

Figure 3: Network Structure of Naive BN based on Loan and Credit Characteristics

- $\boldsymbol{\theta}_{LC}$ **,** as previously defined**,** is the set of all conditional probabilities that are required to describe the joint probability distribution of $\boldsymbol{X}_{LC}$ where $\boldsymbol{x}_{lc}$ is one of 1,062,882 possible combinations of states for the 12 nodes with directed edges into the default.

In following these steps and using the sampling and interval calculations mechanisms previously discussed in the methods section, the following results are presented on the original dataset sample with the thresholds as discussed in the methods section:

Table 7: Mean and 95% Intervals for Naive Loan and Credit Risk BN

Threshold	Accuracy	FPR	FNR
0.15	0.4066 (0.3905–0.4221)	0.6626 (0.6465–0.6810)	0.2509 (0.2109–0.2772)
0.20	0.6364 (0.6200–0.6517)	0.3080 (0.2904–0.3297)	0.6385 (0.6081–0.6732)
0.25	0.7882 (0.7810–0.7960)	0.0669 (0.0581–0.0762)	0.9289 (0.9076–0.9474)
0.30	0.8078 (0.8014–0.8122)	0.0362 (0.0312–0.0431)	0.9640 (0.9479–0.9757)
0.35	0.8152 (0.8104–0.8198)	0.0251 (0.0204–0.0295)	0.9753 (0.9601–0.9854)
0.40	0.8162 (0.8109–0.8205)	0.0232 (0.0187–0.0285)	0.9787 (0.9651–0.9887)

0.45	0.8186 (0.8142–0.8223)	0.0197 (0.0157–0.0242)	0.9820 (0.9708–0.9911)
0.50	0.8185 (0.8142–0.8229)	0.0197 (0.0154–0.0254)	0.9815 (0.9700–0.9927)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8694 (0.8563–0.8865)	0.3374 (0.3190–0.3535)	0.4860 (0.4645–0.5042)
0.20	0.8428 (0.8353–0.8495)	0.6920 (0.6703–0.7096)	0.7600 (0.7464–0.7718)
0.25	0.8325 (0.8297–0.8352)	0.9331 (0.9238–0.9419)	0.8800 (0.8752–0.8847)
0.30	0.8319 (0.8297–0.8344)	0.9638 (0.9569–0.9688)	0.8929 (0.8891–0.8956)
0.35	0.8318 (0.8297–0.8342)	0.9749 (0.9705–0.9796)	0.8977 (0.8949–0.9003)
0.40	0.8316 (0.8284–0.8337)	0.9768 (0.9715–0.9813)	0.8984 (0.8953–0.9009)
0.45	0.8317 (0.8302–0.8334)	0.9803 (0.9758–0.9843)	0.8999 (0.8973–0.9022)
0.50	0.8316 (0.8300–0.8337)	0.9803 (0.9746–0.9846)	0.8999 (0.8971–0.9026)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1860 (0.1795–0.1933)	0.7491 (0.7228–0.7891)	0.2980 (0.2880–0.3098)
0.20	0.1918 (0.1760–0.2064)	0.3615 (0.3268–0.3919)	0.2506 (0.2304–0.2703)
0.25	0.1766 (0.1391–0.2171)	0.0711 (0.0526–0.0924)	0.1013 (0.0764–0.1291)
0.30	0.1671 (0.1131–0.2231)	0.0360 (0.0243–0.0521)	0.0591 (0.0394–0.0845)
0.35	0.1655 (0.0913–0.2500)	0.0247 (0.0146–0.0399)	0.0429 (0.0251–0.0683)
0.40	0.1568 (0.0843–0.2437)	0.0213 (0.0113–0.0349)	0.0375 (0.0200–0.0604)
0.45	0.1557 (0.0857–0.2383)	0.0180 (0.0089–0.0292)	0.0322 (0.0161–0.0515)
0.50	0.1586 (0.0660–0.2455)	0.0185 (0.0073–0.0300)	0.0331 (0.0131–0.0531)

Macroeconomic-Based Bayesian Network

The third model this paper considers is that which uses solely the observed lagged macroeconomic indicators to achieve an extremely naive model. Let this Bayesian Network be defined by $\boldsymbol{B_M} = (\boldsymbol{X_M}, \boldsymbol{G_M}, \boldsymbol{\theta_M})$.

- $\boldsymbol{X_M}$ represents those characteristics of the economy that are relevant to the default of the loan. For the exemplification purposes of this paper, these characteristics include GDP growth, unemployment rate, inflation rate, federal funds rate, and housing prices. Included in these characteristics is the default indication variable. Formally,

  $$\boldsymbol{X_M} = (GDPGrowth, Unemp, Inflation, FedFunds, HousePrices, Default)$$

- $\boldsymbol{G_M}$ represents the graphical model made up of the directed edges $\boldsymbol{E_M}$ which is defined as

$$\boldsymbol{E_M} = [(GDPGrowth, Default), (Unemp, Default),$$

$$(Inflation, Default), (FedFunds, Default), (HousePrices, Default)]$$

The nodes $\boldsymbol{X_M}$ are defined as previously. In doing so, the assumption of the graphical structure of this mode is shown in Figure 4:

Figure 4: Network Structure of Naive BN based on Macroeconomics

- $\boldsymbol{\theta_M}$ **,** as previously defined**,** is the set of all conditional probabilities that are required to describe the joint probability distribution of $\boldsymbol{X_M}$ where $\boldsymbol{x_M}$ is one of 243 possible combinations of states for the 5 nodes with directed edges into the default. The following is an example of the resulting conditional probability distribution for Default:

Table 7: Example of Conditional Probability Distribution Values for Default

FedFunds (t–1)	GDP Growth (t–1)	Housing Price (t–1)	Inflation (t–1)	Unemp. Rate (t–1)	P(0)	P(1)
0.0	0.0	2.0	1.0	2.0	0.5	0.5

0.0	0.0	0.0	1.0	2.0	0.5	0.5
1.0	2.0	1.0	2.0	2.0	0.5	0.5
2.0	1.0	2.0	1.0	0.0	0.5	0.5
2.0	1.0	0.0	2.0	2.0	0.5	0.5

In following these steps and using the sampling and interval calculations mechanisms previously discussed in the methods section, the following results are presented on the original dataset sample with the thresholds as discussed in the methods section:

Table 8: Mean and 95% Intervals for Macroeconomic BN

Threshold	Accuracy	FPR	FNR
0.20	0.6958 (0.6839–0.7085)	0.2227 (0.2076–0.2374)	0.7075 (0.6754–0.7447)
0.25	0.8319 (0.8314–0.8325)	0.0000 (0.0000–0.0000)	1.0000 (1.0000–1.0000)
0.30	0.8319 (0.8314–0.8326)	0.0000 (0.0000–0.0000)	1.0000 (1.0000–1.0000)
0.35	0.8319 (0.8314–0.8324)	0.0000 (0.0000–0.0000)	1.0000 (1.0000–1.0000)
0.40	0.8319 (0.8314–0.8325)	0.0000 (0.0000–0.0000)	1.0000 (1.0000–1.0000)
0.45	0.8319 (0.8313–0.8325)	0.0000 (0.0000–0.0000)	1.0000 (1.0000–1.0000)
0.50	0.8318 (0.8314–0.8323)	0.0000 (0.0000–0.0000)	1.0000 (1.0000–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.20	0.8446 (0.8371–0.8517)	0.7773 (0.7626–0.7924)	0.8095 (0.8005–0.8185)
0.25	0.8319 (0.8314–0.8325)	1.0000 (1.0000–1.0000)	0.9082 (0.9079–0.9086)
0.30	0.8319 (0.8314–0.8326)	1.0000 (1.0000–1.0000)	0.9082 (0.9080–0.9086)
0.35	0.8319 (0.8314–0.8324)	1.0000 (1.0000–1.0000)	0.9082 (0.9079–0.9086)
0.40	0.8319 (0.8314–0.8325)	1.0000 (1.0000–1.0000)	0.9082 (0.9079–0.9086)
0.45	0.8319 (0.8313–0.8325)	1.0000 (1.0000–1.0000)	0.9082 (0.9079–0.9086)
0.50	0.8318 (0.8314–0.8323)	1.0000 (1.0000–1.0000)	0.9082 (0.9080–0.9085)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.20	0.2098 (0.1871–0.2329)	0.2925 (0.2553–0.3246)	0.2443 (0.2159–0.2707)
0.25	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)
0.30	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)
0.35	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)
0.40	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)
0.45	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)
0.50	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)

__Mixed Approach__

To take advantage of the predictive abilities of multiple types of indicators available to BNPL and credit firms, the fourth model this paper proposes is that which uses macroeconomic, credit, and personal characteristics. The motivation behind such a model is to capture the benefits in considering a diversity of indication mechanisms. To do this, let this Bayesian Network be defined by $\boldsymbol{B_{MI}} = (\boldsymbol{X_{MI}}, \boldsymbol{G_{MI}}, \boldsymbol{\theta_{MI}})$.

- $\boldsymbol{X_{MI}}$ represents the macroeconomic, credit, and personal characteristics relevant to the default risk. For the exemplification purposes of this paper, the macroeconomic characteristic included is GDP growth. The credit-related characteristics include FICO score bucket, revolving utilization, total credit accounts, annual income, delinquencies in the last 2 years, and open credit accounts. The personal characteristics included are the purpose, employment length, and homeownership. Included in these characteristics is the default indication variable. Formally, this is defined as

$$\boldsymbol{X_{MI}} = \begin{pmatrix} FICOBucket, RevUtil, TAcc, AnnualInc, Delinq2Y, OpenAcc \\ EmpLength, HomeOwnership, Purpose, GDPGrowth, Default \end{pmatrix}$$

- $\boldsymbol{G_{MI}}$ represents the graphical model made up of the directed edges $\boldsymbol{E_{MI}}$ which is defined as

$$\boldsymbol{E_{MI}} = [(FICOBucket, Default), (RevUtil, Default), (TAcc, Default),$$

$$(AnnualInc, Default), (Delinq2Y, Default), (OpenAcc, Default),$$

$$(GDPGrowth, Default)(EmpLength, Default), (HomeOwnership, Default),$$

$$(Purpose, Default)]$$

The nodes $\boldsymbol{X_{MI}}$ are defined as previously. In doing so, the assumption of the graphical structure of this mode is shown in Figure 5:

Figure 5: Network Structure of Naive BN based on Mixed Characteristics

- $\boldsymbol{\theta_{MI}}$ **,** as previously defined**,** is the set of all conditional probabilities that are required to describe the joint probability distribution of $\boldsymbol{X_{MI}}$ where $\boldsymbol{x_{MI}}$ is one of 1,924,560 possible combinations of states for the 10 nodes with directed edges into the default.

In following these steps and using the sampling and interval calculations mechanisms previously discussed in the methods section, the following results are presented on the original dataset sample with the thresholds as discussed in the methods section:

Table 9: Mean and 95% Intervals for Mixed BN

Threshold	Accuracy	FPR	FNR
0.15	0.4853 (0.4693–0.5045)	0.5383 (0.5204–0.5563)	0.3889 (0.3448–0.4294)

0.20	0.5910 (0.5747–0.6066)	0.3799 (0.3616–0.3964)	0.5640 (0.5199–0.6071)
0.25	0.6686 (0.6542–0.6844)	0.2591 (0.2403–0.2748)	0.7158 (0.6768–0.7534)
0.30	0.7180 (0.7059–0.7277)	0.1826 (0.1713–0.1957)	0.8107 (0.7780–0.8373)
0.35	0.7422 (0.7298–0.7558)	0.1444 (0.1313–0.1569)	0.8606 (0.8302–0.8952)
0.40	0.7488 (0.7389–0.7578)	0.1351 (0.1237–0.1467)	0.8689 (0.8359–0.8943)
0.45	0.7538 (0.7439–0.7657)	0.1265 (0.1147–0.1363)	0.8827 (0.8627–0.9152)
0.50	0.7548 (0.7430–0.7659)	0.1253 (0.1144–0.1386)	0.8829 (0.8542–0.9076)

Threshold	Precision (0)	Recall (0)	F (0)
0.15	0.8633 (0.8494–0.8766)	0.4617 (0.4437–0.4796)	0.6016 (0.5855–0.6198)
0.20	0.8540 (0.8436–0.8642)	0.6201 (0.6036–0.6384)	0.7184 (0.7050–0.7312)
0.25	0.8463 (0.8386–0.8533)	0.7409 (0.7252–0.7597)	0.7900 (0.7797–0.8016)
0.30	0.8429 (0.8371–0.8473)	0.8174 (0.8043–0.8287)	0.8299 (0.8219–0.8365)
0.35	0.8410 (0.8353–0.8465)	0.8556 (0.8431–0.8687)	0.8482 (0.8407–0.8571)
0.40	0.8412 (0.8376–0.8465)	0.8649 (0.8533–0.8763)	0.8528 (0.8462–0.8583)
0.45	0.8403 (0.8352–0.8444)	0.8735 (0.8637–0.8853)	0.8566 (0.8502–0.8640)
0.50	0.8405 (0.8366–0.8446)	0.8747 (0.8614–0.8856)	0.8573 (0.8497–0.8641)

Threshold	Precision (1)	Recall (1)	F (1)
0.15	0.1759 (0.1643–0.1872)	0.6111 (0.5706–0.6552)	0.2731 (0.2551–0.2904)
0.20	0.1775 (0.1612–0.1925)	0.4360 (0.3929–0.4801)	0.2523 (0.2288–0.2747)
0.25	0.1709 (0.1501–0.1899)	0.2842 (0.2466–0.3232)	0.2134 (0.1860–0.2387)
0.30	0.1630 (0.1395–0.1829)	0.1893 (0.1627–0.2220)	0.1751 (0.1506–0.1981)
0.35	0.1536 (0.1169–0.1870)	0.1394 (0.1048–0.1698)	0.1460 (0.1105–0.1757)
0.40	0.1541 (0.1304–0.1875)	0.1311 (0.1057–0.1641)	0.1424 (0.1313–0.1569)
0.45	0.1484 (0.1091–0.1768)	0.1173 (0.0848–0.1373)	0.1310 (0.0954–0.1532)
0.50	0.1495 (0.1206–0.1801)	0.1171 (0.0924–0.1458)	0.1313 (0.1051–0.1599)

Static Bayesian Network with Structure Learning

The second type of network-based model this paper considers as a plausible credit risk model is a Bayesian Network whose structure is inferred from the data. The benefits of this model are straightforward. In the naive approach, there are limitations to both data missing values and the scale by which these values affect the result. Because the conditional probability distribution for the static approach with structure learning does not depend on all configurations of all previously known variables, there is some flexibility for the BNPL-like firm to use a limited number of core information and still have meaningful results in terms of possible default predictions as a secondary credit model. With the use of a learned structure, this limited information is used optimally to predict default.

To learn the structure of the network, a non-parametric bootstrap approach is implemented. In this approach, the network structure is learned on data sampled by replacement with stratification to avoid overlearning the negative class edges only. This is repeated 100 times, and the resulting network structure's directed edges are saved and then compared. An edge is only implemented in the final network structure if it is contained in greater than 95% of the bootstrap-sampled edges. In following this process for all of the following structures in this section of the paper, there is the use of a robust way to ensure that the final learned network structure is as accurate as possible. Note, in the following section, rather than formally defining the edge set in its entirety (due to a large number of variables and edges), a description of the restrictions on the set of all variables to create the relevant edge set is provided.

Once a network structure has been learned, the use of the bootstrap approach described in the methods section regarding the classification metrics is used. In doing so, for the primary models of interest with the Bayesian Network and BIC scoring function approach, 100 bootstrap samples are used as in the previous section with the naive BNs. However, for those additional cases of tests that are conducted, such as that with the structure scores, while 100 bootstrap samples are used to learn the network structures as accurately as possible, in testing the models, there is a smaller number of samples for illustration purposes and due to computations constraints of this paper.

It is notable to additionally point out that in learning the network structure and conducting the bootstrap sampling approach, it was generally less computationally expensive to do so compared to the more complex naive BNs considered previously in this section of the paper, such as that which took a mixed approach on variable inclusion.

Strict Restriction on Order Development

The first type of static BN with a learned structure is that with a strict restriction on the assumed temporal development of the available nodes. To explain this in a more interpretable sense, this model focuses on introducing an edge set that has edges that may not occur and may occur. The way in which the edges that may not occur in the network are determined is by the expected order in which they occur over time. Particularly, the model assumes that the personal characteristics and the motivation for the loan occur, and then the macroeconomic variables are taken into account. Once that has been accomplished, in the face of the economic environment and the personal characteristics of the lender, the credit characteristics are developed from which the lender derives the arbitrary grading scheme of the loans. Then the loan occurs with the amount, interest rate, installment, and other unique characteristics. Finally, the default value is witnessed. In following this temporal order scheme, restriction is not placed only on the ability of the nodes in a previous category to go to the next category, but also on the order in which the nodes may exist in a single category. Within the personal characteristics of the borrower, an assumption is made that the age group exists, and then comes employment length, the FICO score bucket, the purpose of the loan, the annual income, and homeownership. Then the issue quarter exists for the loan at which time, the lagged GDP growth, unemployment rate, federal funds rate, inflation rate, and housing price follow suit. Then the number of open accounts, public records, revolving total balance, revolving total utility, delinquencies, inquiries, total accounts, grade, subgrade, and loan details follow.

Consider the two following Bayesian Networks:

$$\boldsymbol{B_{S1}} = (\boldsymbol{X_{S1}}, \boldsymbol{G_{S1}}, \boldsymbol{\theta_{S1}}) \text{ and } \boldsymbol{B_{S2}} = (\boldsymbol{X_{S2}}, \boldsymbol{G_{S2}}, \boldsymbol{\theta_{S2}})$$

$\boldsymbol{B_{S1}}$ and $\boldsymbol{B_{S2}}$ are nearly the same except with one difference in the restriction of the edges with regard to default. Both have the previously described order of development of the edges. For $\boldsymbol{B_{S1}}$, the restriction that occurs for the edges is that a directed edge between $\boldsymbol{X_i} \in \boldsymbol{X_{S1}}$ to $\boldsymbol{X_j} \in \boldsymbol{X_{S1}}$ can only exist if $\boldsymbol{i} > \boldsymbol{j}$. This is the same for $\boldsymbol{B_{S2}}$. The primary difference between the two is that $\boldsymbol{G_{S1}} = \{\boldsymbol{X_{S1}}, \boldsymbol{E_{S1}}\}$ has the added restriction on edges that must be contained in $\boldsymbol{E_{S1}}$. Specifically, $\boldsymbol{E_{S1}}$ must contain directed edges from age group, employment length, annual income, FICO bucket, and purpose to the default indicator. This is to effectively include personal and credit effects for certain in the default probability conclusions. This restriction does not exist in $\boldsymbol{B_{S2}}$. Note, $\boldsymbol{B_{S1}}$ will be referred to as Strict Order BN 1 in the context of this section and $\boldsymbol{B_{S2}}$ will be referred to as Strict Order BN 2.

In following the previously stated steps for the structure learning of the BN in the introduction of this section, the following are the learned structures of the BNs:

Figure 6: Network Structure of Strict Order BN 1 with BIC Scoring

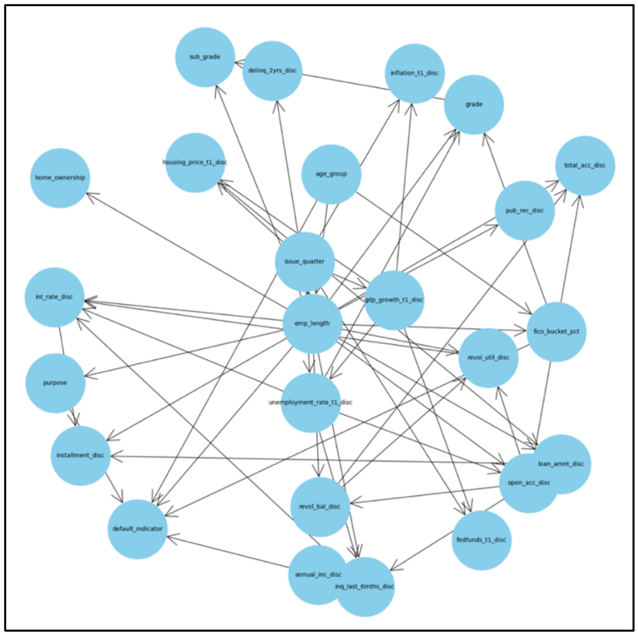
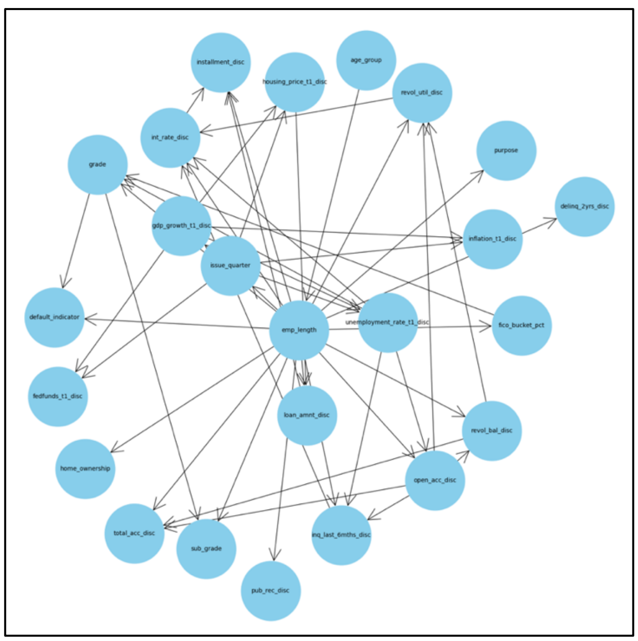


Figure 7: Network Structure of Strict Order BN 2 with BIC Scoring



The following are the means and intervals of the various classification metrics for the BIC case in both the strict order models:

Table 10: Mean and 95% Intervals for Strict Order BN 1 with BIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.4863 (0.4690–0.5020)	0.5493 (0.5323–0.5696)	0.3241 (0.2854–0.3593)
0.20	0.6978 (0.6830–0.7117)	0.2281 (0.2090–0.2452)	0.6957 (0.6490–0.7308)
0.25	0.8127 (0.8051–0.8222)	0.0469 (0.0394–0.0548)	0.9344 (0.9116–0.9510)
0.30	0.8285 (0.8222–0.8344)	0.0209 (0.0160–0.0257)	0.9733 (0.9608–0.9854)
0.35	0.8339 (0.8274–0.8393)	0.0115 (0.0079–0.0143)	0.9868 (0.9779–0.9961)
0.40	0.8362 (0.8310–0.8419)	0.0088 (0.0054–0.0126)	0.9898 (0.9801–0.9962)
0.45	0.8370 (0.8317–0.8439)	0.0066 (0.0038–0.0093)	0.9933 (0.9907–0.9962)
0.50	0.8407 (0.8362–0.8454)	0.0015 (0.0004–0.0031)	0.9983 (0.9942–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8811 (0.8670–0.8956)	0.4507 (0.4304–0.4677)	0.5963 (0.5771–0.6131)
0.20	0.8549 (0.8470–0.8637)	0.7719 (0.7548–0.7910)	0.8112 (0.8009–0.8225)
0.25	0.8445 (0.8400–0.8511)	0.9531 (0.9452–0.9606)	0.8955 (0.8908–0.9010)
0.30	0.8427 (0.8380–0.8471)	0.9791 (0.9743–0.9840)	0.9058 (0.9019–0.9093)
0.35	0.8417 (0.8367–0.8465)	0.9885 (0.9857–0.9921)	0.9092 (0.9053–0.9124)
0.40	0.8421 (0.8373–0.8466)	0.9912 (0.9874–0.9946)	0.9106 (0.9076–0.9139)
0.45	0.8415 (0.8373–0.8471)	0.9934 (0.9907–0.9962)	0.9112 (0.9080–0.9153)
0.50	0.8418 (0.8371–0.8458)	0.9985 (0.9969–0.9996)	0.9134 (0.9108–0.9162)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1876 (0.1781–0.1990)	0.6759 (0.6407–0.7146)	0.2937 (0.2792–0.3093)
0.20	0.2007 (0.1796–0.2265)	0.3043 (0.2692–0.3510)	0.2418 (0.2162–0.2742)
0.25	0.2082 (0.1617–0.2654)	0.0656 (0.0490–0.0884)	0.0996 (0.0757–0.1326)
0.30	0.1934 (0.1122–0.2840)	0.0267 (0.0146–0.0392)	0.0469 (0.0258–0.0684)
0.35	0.1773 (0.0628–0.2850)	0.0132 (0.0039–0.0221)	0.0246 (0.0074–0.0404)
0.40	0.1803 (0.0657–0.3287)	0.0102 (0.0038–0.0199)	0.0088 (0.0054–0.0126)
0.45	0.1618 (0.0198–0.3333)	0.0067 (0.0009–0.0116)	0.0066 (0.0038–0.0093)
0.50	0.1711 (0.0000–0.7500)	0.0017 (0.0000–0.0058)	0.0034 (0.0000–0.0114)

Table 11: Mean and 95% Intervals for Strict Order BN 2 with BIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.5341 (0.5177–0.5523)	0.4953 (0.4780–0.5135)	0.3089 (0.2693–0.3483)
0.20	0.7118 (0.6908–0.7300)	0.2208 (0.1958–0.2475)	0.6477 (0.5992–0.6993)
0.25	0.8166 (0.8064–0.8255)	0.0478 (0.0393–0.0608)	0.9049 (0.8793–0.9283)
0.30	0.8394 (0.8347–0.8462)	0.0048 (0.0023–0.0083)	0.9896 (0.9807–0.9962)
0.35	0.8406 (0.8359–0.8452)	0.0014 (0.0004–0.0025)	0.9959 (0.9904–1.0000)
0.40	0.8362 (0.8310–0.8419)	0.0005 (0.0000–0.0016)	0.9982 (0.9943–1.0000)
0.45	0.8370 (0.8317–0.8439)	0.0002 (0.0000–0.0007)	0.9989 (0.9961–1.0000)
0.50	0.8418 (0.8377–0.8463)	0.0002 (0.0000–0.0009)	0.9995 (0.9980–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8970 (0.8855–0.9101)	0.5047 (0.4865–0.5220)	0.6459 (0.6302–0.6621)
0.20	0.8651 (0.8579–0.8735)	0.7792 (0.7525–0.8042)	0.8199 (0.8034–0.8331)
0.25	0.8485 (0.8431–0.8549)	0.9522 (0.9392–0.9607)	0.8973 (0.8908–0.9026)
0.30	0.8426 (0.8383–0.8471)	0.9952 (0.9917–0.9977)	0.9126 (0.9098–0.9165)
0.35	0.8417 (0.8369–0.8465)	0.9885 (0.9857–0.9921)	0.9134 (0.9053–0.9161)
0.40	0.8421 (0.8373–0.8466)	0.9995 (0.9984–1.0000)	0.9142 (0.9116–0.9163)
0.45	0.8420 (0.8373–0.8471)	0.9998 (0.9993–1.0000)	0.9141 (0.9080–0.9153)
0.50	0.8419 (0.8371–0.8458)	0.9985 (0.9969–0.9996)	0.9134 (0.9108–0.9162)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2074 (0.1945–0.2196)	0.6911 (0.6517–0.7307)	0.3191 (0.3003–0.3364)
0.20	0.2304 (0.2076–0.2522)	0.3523 (0.3007–0.4008)	0.2784 (0.2522–0.3042)
0.25	0.2727 (0.2272–0.3315)	0.0951 (0.0717–0.1207)	0.1408 (0.1078–0.1713)
0.30	0.2952 (0.0952–0.5511)	0.0104 (0.0038–0.0193)	0.0200 (0.0073–0.0369)
0.35	0.3542 (0.0000–0.6917)	0.0041 (0.0000–0.0096)	0.0081 (0.0000–0.0188)
0.40	0.4140 (0.0000–1.0000)	0.0018 (0.0000–0.0057)	0.0037 (0.0000–0.0126)
0.45	0.3533 (0.0000–1.0000)	0.0011 (0.0000–0.0039)	0.0022 (0.0000–0.0078)
0.50	0.1850 (0.0000–1.0000)	0.0005 (0.0000–0.0019)	0.0010 (0.0000–0.0039)

Loose Restriction on Order Development

In following a strict order of temporal development, there is always the possibility that there is an incorrect ordering of the variables or the temporal scheme does not hold up to real restrictions in borrowing and lending requirements. Therefore, in this subsection, there is consideration of a much more relaxed model in terms of restrictions. Specifically, in this section, the paper assumes that the primary restrictions to existing edges in the BN is that there may be no edges going out of the default indicator and there may be no edges going into age group. The reasoning for these very relaxed decisions is straightforward. Default is the variable of interest and does not include any of the previously seen nodes in the network. Age group is determined completely outside of the network, as it was determined by factors described in the methods section, which were not included in the network. In making these restrictions, the BNs that follow in this subsection may learn a structure that assumes any order of dependence between age group and default.

Consider the two following Bayesian Networks:

$$\boldsymbol{B_{L1}} = (\boldsymbol{X_{L1}}, \boldsymbol{G_{L1}}, \boldsymbol{\theta_{L1}}) \text{ and } \boldsymbol{B_{L2}} = (\boldsymbol{X_{L2}}, \boldsymbol{G_{L2}}, \boldsymbol{\theta_{L2}})$$

$\boldsymbol{B_{L1}}$ and $\boldsymbol{B_{L2}}$ are nearly the same except with one difference in the restriction of the edges with regard to default. This restriction difference is defined exactly the same as the difference between $\boldsymbol{B_{S1}}$ and $\boldsymbol{B_{S2}}$ from the previous subsection**.** Note, $\boldsymbol{B_{L1}}$ will be referred to as Loose Order BN 1 in the context of this section and $\boldsymbol{B_{L2}}$ will be referred to as Loose Order BN 2.

In following the previously stated steps for the structure learning of the BN in the introduction of this section, the following are the learned structures of the BNs:

Figure 8: Network Structure of Loose Order BN 1 with BIC Scoring

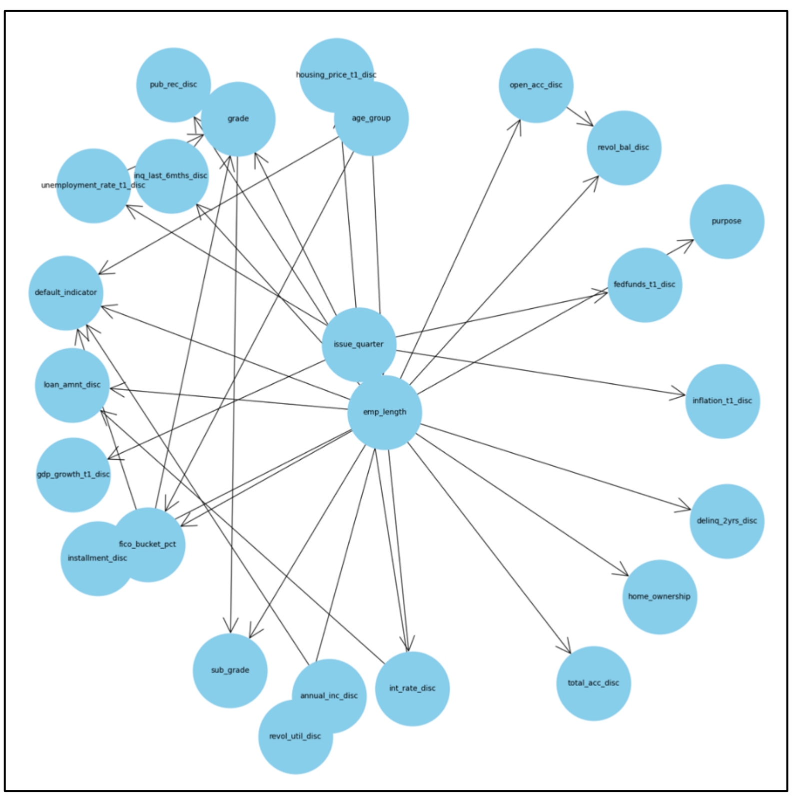
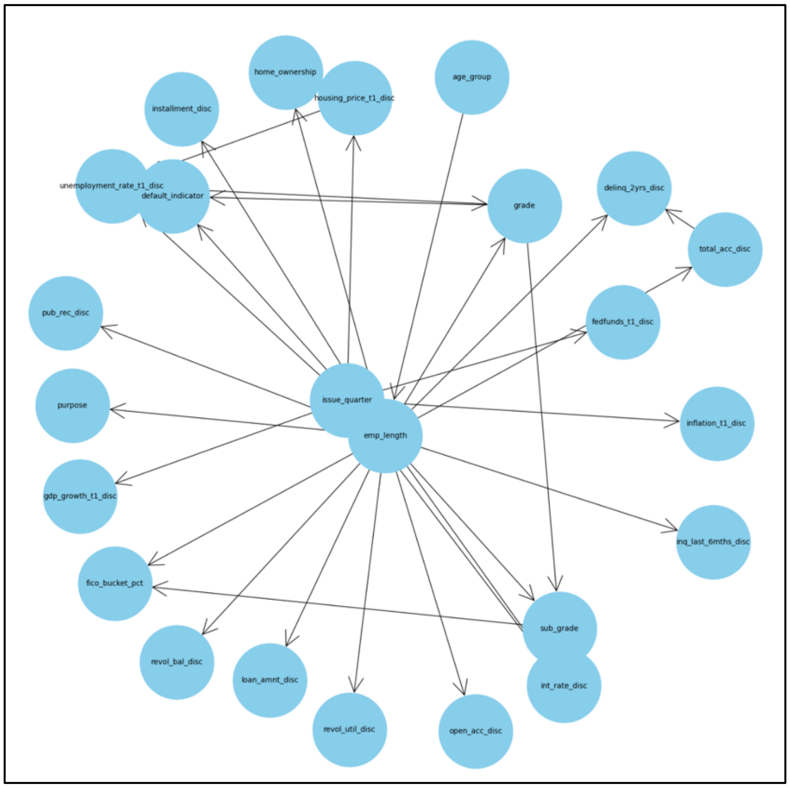


Figure 9: Network Structure of Loose Order BN 2 with BIC Scoring



The following are the means and intervals of the various classification metrics for the BIC case in both the loose order models:

Table 12: Mean and 95% Intervals for Loose Order BN 1 with BIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.4413 (0.4233–0.4580)	0.6147 (0.5949–0.6355)	0.2596 (0.2192–0.2980)
0.20	0.7726 (0.7629–0.7833)	0.1109 (0.0976–0.1225)	0.8464 (0.8182–0.8694)
0.25	0.8355 (0.8308–0.8406)	0.0098 (0.0066–0.0139)	0.9860 (0.9754–0.9943)
0.30	0.8413 (0.8377–0.8462)	0.0010 (0.0000–0.0022)	0.9988 (0.9968–1.0000)
0.35	0.8411 (0.8369–0.8449)	0.0005 (0.0000–0.0014)	0.9995 (0.9959–1.0000)
0.40	0.8417 (0.8365–0.8463)	0.0003 (0.0000–0.0009)	0.9982 (0.9943–1.0000)
0.45	0.8414 (0.8368–0.8462)	0.0002 (0.0000–0.0007)	0.9995 (0.9983–1.0000)
0.50	0.8419 (0.8371–0.8471)	0.0001 (0.0000–0.0007)	0.9999 (0.9980–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8878 (0.8734–0.9036)	0.3853 (0.3645–0.4051)	0.5373 (0.5169–0.5569)
0.20	0.8481 (0.8424–0.8536)	0.8891 (0.8775–0.9024)	0.8681 (0.8617–0.8753)

0.25	0.8422 (0.8383–0.8468)	0.9902 (0.9861–0.9934)	0.9102 (0.9074–0.9132)
0.30	0.8420 (0.8383–0.8468)	0.9990 (0.9978–1.0000)	0.9138 (0.9117–0.9166)
0.35	0.8414 (0.8372–0.8452)	0.9986 (0.9980–1.0000)	0.9134 (0.9112–0.9159)
0.40	0.8421 (0.8379–0.8465)	0.9995 (0.9984–1.0000)	0.9142 (0.9116–0.9163)
0.45	0.8415 (0.8372–0.8471)	0.9998 (0.9993–1.0000)	0.9141 (0.9111–0.9153)
0.50	0.8419 (0.8371–0.8458)	0.9999 (0.9993–1.0000)	0.9141 (0.9108–0.9168)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1843 (0.1747–0.1974)	0.7404 (0.7020–0.7808)	0.2951 (0.2812–0.3147)
0.20	0.2067 (0.1781–0.2385)	0.1536 (0.1306–0.1818)	0.1761 (0.1518–0.2015)
0.25	0.2106 (0.1025–0.3497)	0.0140 (0.0057–0.0246)	0.0262 (0.0109–0.0455)
0.30	0.1974 (0.0000–0.8417)	0.0012 (0.0000–0.0038)	0.0023 (0.0000–0.0075)
0.35	0.1653 (0.0000–1.0000)	0.0006 (0.0000–0.0037)	0.0013 (0.0000–0.0074)
0.40	0.2150 (0.0000–1.0000)	0.0006 (0.0000–0.0057)	0.0013 (0.0000–0.0126)
0.45	0.1667 (0.0000–1.0000)	0.0005 (0.0000–0.0019)	0.0009 (0.0000–0.0039)
0.50	0.1533 (0.0000–1.0000)	0.0004 (0.0000–0.0007)	0.0002 (0.0000–0.0007)

Table 13: Mean and 95% Intervals for Loose Order BN 2 with BIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.5341 (0.5187–0.5503)	0.4956 (0.4810–0.5117)	0.3086 (0.2705–0.3490)
0.20	0.7117 (0.6847–0.7340)	0.2215 (0.1901–0.2547)	0.6442 (0.5985–0.6961)
0.25	0.8163 (0.8077–0.8247)	0.0476 (0.0386–0.0599)	0.9069 (0.8768–0.9337)
0.30	0.8399 (0.8347–0.8443)	0.0045 (0.0023–0.0082)	0.9883 (0.9780–0.9972)
0.35	0.8407 (0.8360–0.8469)	0.0016 (0.0002–0.0031)	0.9956 (0.9885–1.0000)
0.40	0.8414 (0.8363–0.8463)	0.0005 (0.0000–0.0014)	0.9982 (0.9944–1.0000)
0.45	0.8416 (0.8374–0.8460)	0.0002 (0.0000–0.0009)	0.9994 (0.9962–1.0000)
0.50	0.8416 (0.8373–0.8468)	0.0002 (0.0000–0.0007)	0.9995 (0.9981–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8966 (0.8841–0.9103)	0.5044 (0.4883–0.5190)	0.6456 (0.6309–0.6602)
0.20	0.8655 (0.8575–0.8746)	0.7785 (0.7453–0.8099)	0.8196 (0.7989–0.8361)
0.25	0.8480 (0.8428–0.8543)	0.9524 (0.9401–0.9614)	0.8972 (0.8916–0.9023)
0.30	0.8428 (0.8382–0.8467)	0.9955 (0.9918–0.9980)	0.9128 (0.9097–0.9154)
0.35	0.8418 (0.8368–0.8473)	0.9984 (0.9969–1.0000)	0.9134 (0.9106–0.9170)
0.40	0.8417 (0.8368–0.8465)	0.9995 (0.9986–1.0000)	0.9142 (0.9116–0.9167)
0.45	0.8418 (0.8376–0.8459)	0.9998 (0.9991–1.0000)	0.9140 (0.9115–0.9165)
0.50	0.8417 (0.8373–0.8468)	0.9998 (0.9993–1.0000)	0.9141 (0.9093–0.9171)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2083 (0.1979–0.2187)	0.6914 (0.6510–0.7295)	0.3201 (0.3040–0.3363)
0.20	0.2319 (0.2125–0.2599)	0.3558 (0.3039–0.4015)	0.2806 (0.2554–0.3113)

0.25	0.2693 (0.2126–0.3343)	0.0931 (0.0663–0.1232)	0.1381 (0.1015–0.1729)
0.30	0.3303 (0.1111–0.6052)	0.0117 (0.0028–0.0220)	0.0225 (0.0055–0.0419)
0.35	0.3300 (0.0000–0.7367)	0.0044 (0.0000–0.0115)	0.0087 (0.0000–0.0224)
0.40	0.3840 (0.0000–1.0000)	0.0018 (0.0000–0.0056)	0.0036 (0.0000–0.0111)
0.45	0.2107 (0.0000–1.0000)	0.0006 (0.0000–0.0038)	0.0012 (0.0000–0.0075)
0.50	0.1833 (0.0000–1.0000)	0.0005 (0.0000–0.0007)	0.0009 (0.0000–0.0007)

Consideration to the Structure Score

As mentioned in the section of this paper on notation and explanation of BNs and DBNs, there are multiple scoring functions that can be used in conjunction with the Hill Climb Search algorithm. In this subsection, there is consideration to the network structures proposed by the these functions and what the preliminary results indicate. The three additional scoring functions which are considered, as mentioned previously, are AIC, BDeu, and BDs. The optimal network structures in the case of the BN with strict restriction on order development is given as follows (the left side is with strict edge definitions of required edges to default while the right side is with the looser set of restrictions):

Figure 10: Network Structure of Strict Order BN with AIC Scoring

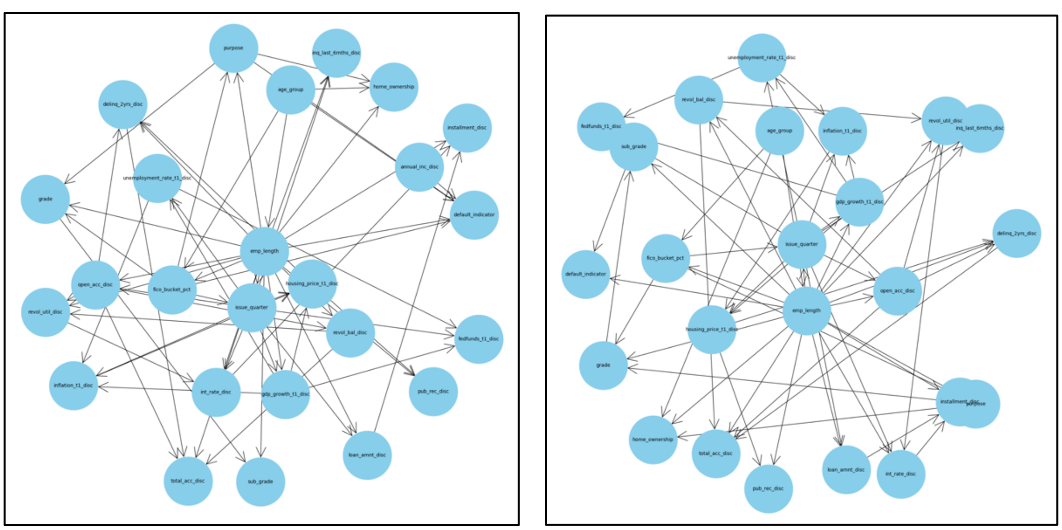


Figure 11: Network Structure of Strict Order BN with BDeu Scoring

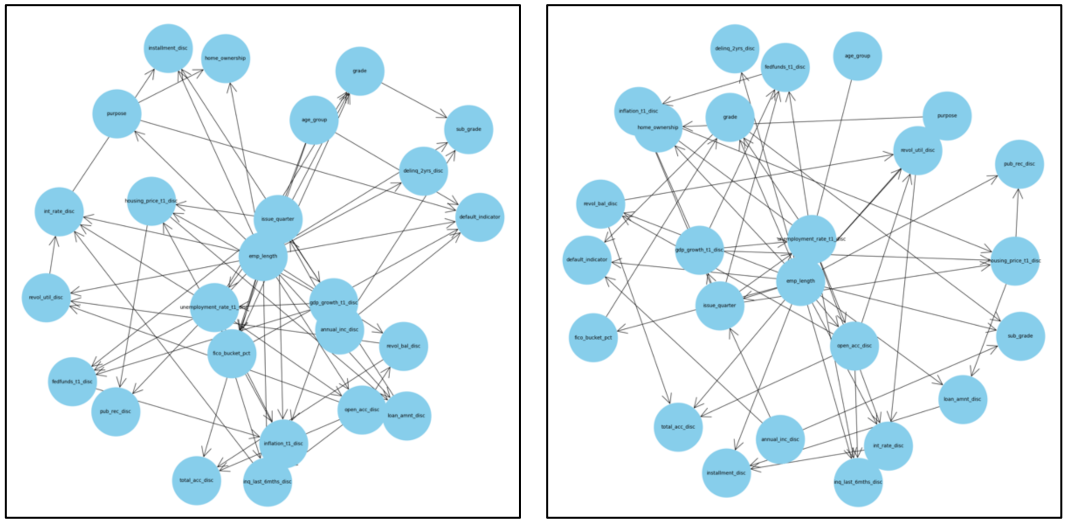
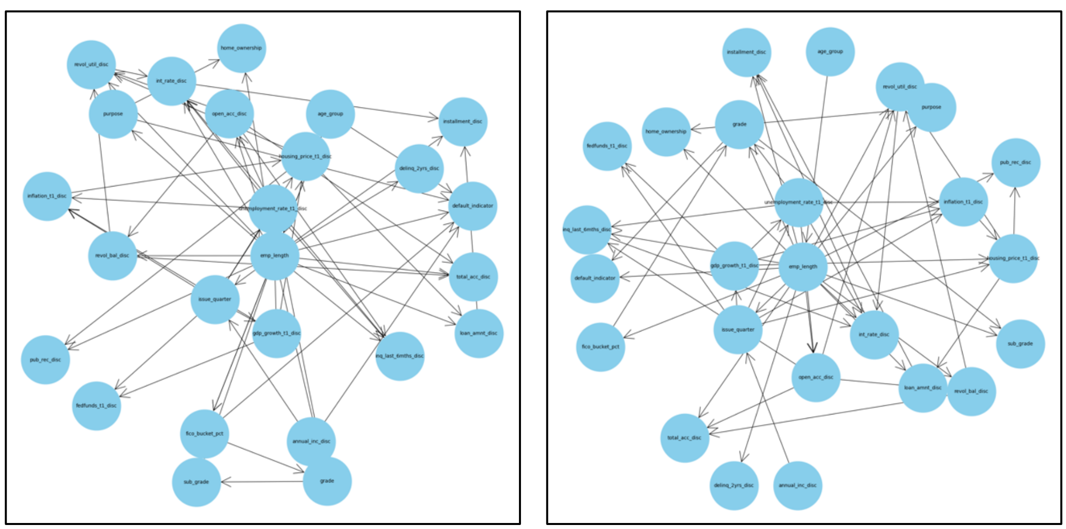


Figure 12: Network Structure of Strict Order BN with BDs Scoring



In conducting these network structures, there is a lot of similarity, particularly in the way that default is predicted due to the initial edge set to default being specified in the way it was before. This is not the necessarily the case when considering the much more loose restriction on the edges to default which can exists in the second variation of this model where network structure may play a more important role. For exemplification purposes, the following are the means and intervals of the various classification metrics for the AIC case in both the strict order models (additional score metrics and their corresponding results can also be found in section A.5 of the Appendix):

Table 14: Mean and 95% Intervals for Strict Order BN 1 with AIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.4974 (0.4827–0.5088)	0.5387 (0.5228–0.5560)	0.3108 (0.2851–0.3337)
0.20	0.7033 (0.6917–0.7102)	0.2242 (0.2131–0.2371)	0.6841 (0.6535–0.7072)
0.25	0.8118 (0.8063–0.8178)	0.0477 (0.0428–0.0518)	0.9328 (0.9199–0.9455)
0.30	0.8276 (0.8202–0.8327)	0.0209 (0.0169–0.0242)	0.9730 (0.9675–0.9831)
0.35	0.8353 (0.8320–0.8380)	0.0105 (0.0083–0.0119)	0.9856 (0.9813–0.9917)
0.40	0.8370 (0.8350–0.8406)	0.0083 (0.0058–0.0104)	0.9863 (0.9790–0.9942)
0.45	0.8361 (0.8322–0.8426)	0.0067 (0.0037–0.0094)	0.9941 (0.9906–0.9963)
0.50	0.8416 (0.8383–0.8444)	0.0012 (0.0004–0.0024)	0.9987 (0.9947–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8875 (0.8811–0.8971)	0.4613 (0.4440–0.4772)	0.6070 (0.5911–0.6203)
0.20	0.8582 (0.8535–0.8634)	0.7758 (0.7629–0.7869)	0.8149 (0.8065–0.8206)
0.25	0.8440 (0.8407–0.8484)	0.9523 (0.9482–0.9572)	0.8949 (0.8917–0.8985)
0.30	0.8418 (0.8365–0.8470)	0.9791 (0.9758–0.9831)	0.9052 (0.9008–0.9083)
0.35	0.8424 (0.8394–0.8450)	0.9895 (0.9881–0.9917)	0.9100 (0.9082–0.9117)
0.40	0.8426 (0.8408–0.8445)	0.9917 (0.9896–0.9950)	0.9111 (0.9099–0.9133)
0.45	0.8407 (0.8385–0.8462)	0.9933 (0.9905–0.9961)	0.9107 (0.9083–0.9145)
0.50	0.8424 (0.8373–0.8468)	0.9988 (0.9976–0.9996)	0.9139 (0.9120–0.9156)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1941 (0.1876–0.2024)	0.6892 (0.6663–0.7149)	0.3029 (0.2928–0.3142)
0.20	0.2089 (0.1964–0.2265)	0.3159 (0.2928–0.3465)	0.2514 (0.2364–0.2739)
0.25	0.2103 (0.1627–0.2501)	0.0672 (0.0545–0.0801)	0.1017 (0.0816–0.1200)
0.30	0.1980 (0.1519–0.2394)	0.0270 (0.0222–0.0325)	0.0475 (0.0387–0.0572)
0.35	0.2030 (0.1283–0.2394)	0.0144 (0.0083–0.0187)	0.0268 (0.0157–0.0346)
0.40	0.2335 (0.1283–0.3696)	0.0137 (0.0058–0.0210)	0.0258 (0.0110–0.0394)
0.45	0.1448 (0.0892–0.2154)	0.0059 (0.0037–0.0094)	0.0067 (0.0039–0.0095)
0.50	0.1150 (0.0000–0.3944)	0.0013 (0.0004–0.0024)	0.0026 (0.0004–0.0050)

Table 15: Mean and 95% Intervals for Strict Order BN 2 with AIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.5133 (0.4934–0.5307)	0.5256 (0.5058–0.5476)	0.2792 (0.2536–0.3026)
0.20	0.6916 (0.6847–0.7084)	0.2526 (0.2405–0.2639)	0.6093 (0.5604–0.6429)

0.25	0.8135 (0.8080–0.8200)	0.0563 (0.0495–0.0613)	0.8862 (0.8675–0.9093)
0.30	0.8347 (0.8310–0.8384)	0.0131 (0.0102–0.0165)	0.9726 (0.9646–0.9811)
0.35	0.8400 (0.8350–0.8432)	0.0034 (0.0019–0.0052)	0.9899 (0.9850–0.9961)
0.40	0.8418 (0.8364–0.8463)	0.0005 (0.0000–0.0029)	0.9942 (0.9872–0.9981)
0.45	0.8410 (0.8388–0.8438)	0.0008 (0.0000–0.0020)	0.9983 (0.9961–1.0000)
0.50	0.8421 (0.8393–0.8473)	0.0004 (0.0000–0.0010)	0.9996 (0.9981–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.9008 (0.8910–0.9107)	0.4744 (0.4524–0.4942)	0.6214 (0.6005–0.6406)
0.20	0.8686 (0.8609–0.8791)	0.7474 (0.7361–0.7595)	0.8034 (0.7945–0.8143)
0.25	0.8513 (0.8471–0.8573)	0.9437 (0.9387–0.9505)	0.8951 (0.8917–0.8990)
0.30	0.8433 (0.8407–0.8482)	0.9869 (0.9835–0.9898)	0.9095 (0.9072–0.9116)
0.35	0.8421 (0.8381–0.8454)	0.9966 (0.9948–0.9981)	0.9129 (0.9106–0.9149)
0.40	0.8425 (0.8372–0.8460)	0.9987 (0.9971–1.0000)	0.9140 (0.9108–0.9162)
0.45	0.8415 (0.8394–0.8448)	0.9992 (0.9984–1.0000)	0.9136 (0.9123–0.9152)
0.50	0.8423 (0.8397–0.8473)	0.9996 (0.9990–1.0000)	0.9142 (0.9126–0.9174)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2042 (0.1958–0.2121)	0.7208 (0.6974–0.7464)	0.3182 (0.3062–0.3293)
0.20	0.2231 (0.2042–0.2529)	0.3907 (0.3571–0.4396)	0.2839 (0.2614–0.3211)
0.25	0.2728 (0.2312–0.3138)	0.1138 (0.0907–0.1325)	0.1605 (0.1303–0.1815)
0.30	0.2829 (0.1111–0.6052)	0.0274 (0.0189–0.0354)	0.0499 (0.0350–0.0642)
0.35	0.3602 (0.2151–0.5646)	0.0101 (0.0043–0.0150)	0.0195 (0.0085–0.0289)
0.40	0.5086 (0.2113–1.0000)	0.0058 (0.0019–0.0128)	0.0114 (0.0038–0.0250)
0.45	0.3519 (0.0000–1.0000)	0.0017 (0.0000–0.0039)	0.0034 (0.0000–0.0095)
0.50	0.2000 (0.0000–1.0000)	0.0004 (0.0000–0.0007)	0.0008 (0.0000–0.0010)

For the sake of comparison, the process outlined above for the strict ordering BN is repeated for the loose ordering BNs. The following is the results of the network structures as well as examples of the results for the AIC scoring function for the aggregate classification metrics (the left side is with strict edge definitions of required edges to default while the right side is with the looser set of restrictions on default edges):

Figure 13: Network Structure of Loose Order BN with AIC Scoring

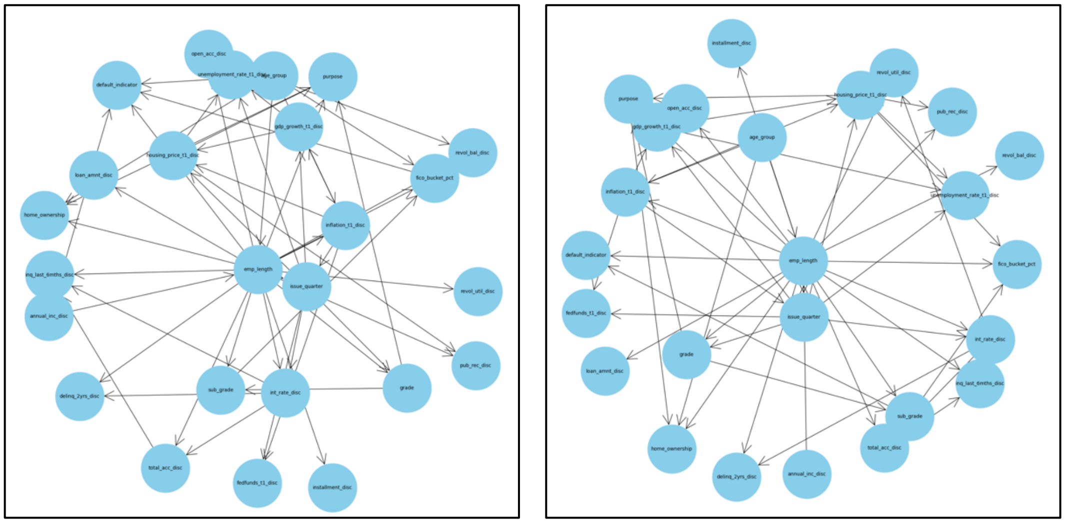


Figure 14: Network Structure of Loose Order BN with BDeu Scoring

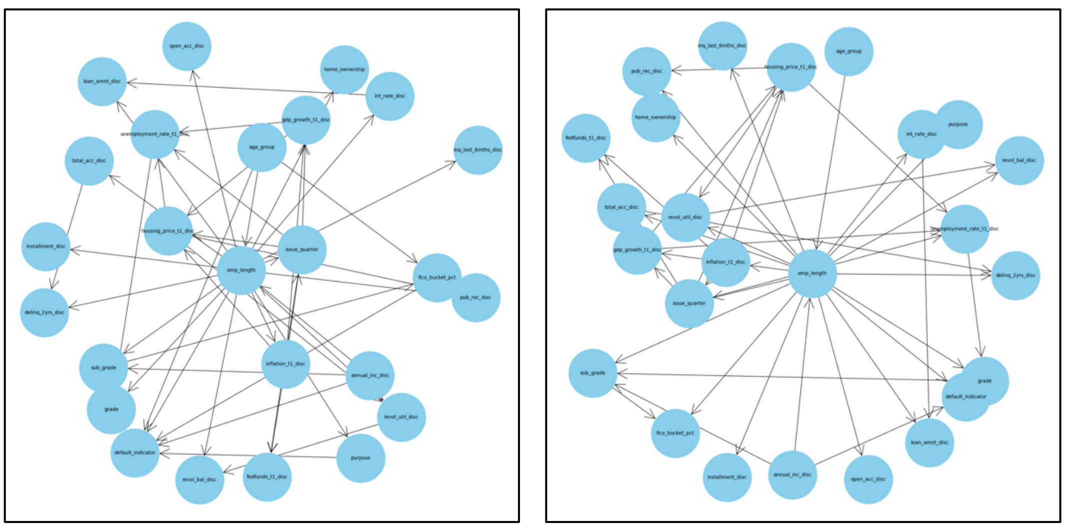


Figure 15: Network Structure of Loose Order BN with BDs Scoring

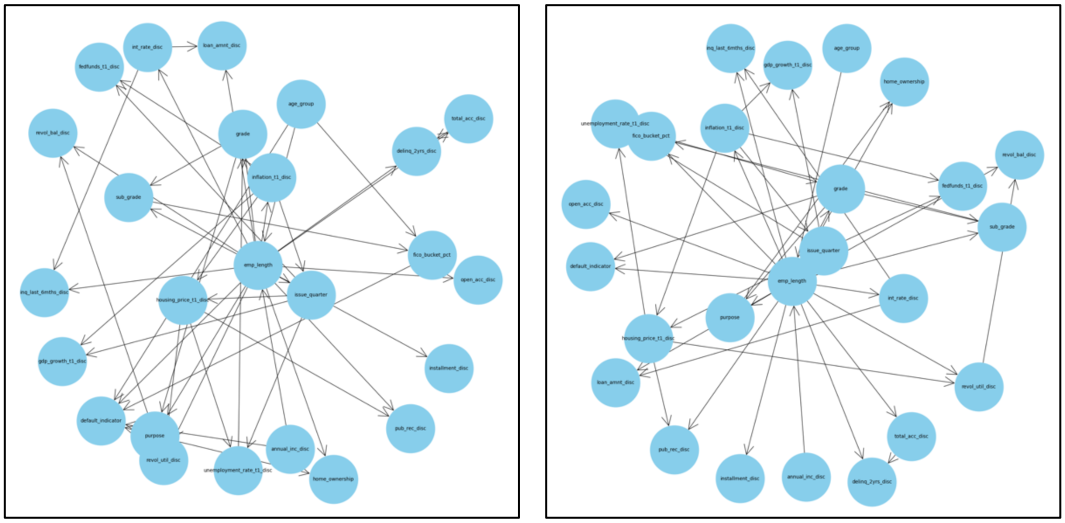


Table 16: Mean and 95% Intervals for Loose Order BN 1 with AIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.4364 (0.4179–0.4538)	0.6191 (0.5996–0.6441)	0.2679 (0.2300–0.2983)
0.20	0.7745 (0.7577–0.7864)	0.1085 (0.0969–0.1245)	0.8462 (0.8169–0.8691)
0.25	0.8359 (0.8319–0.8402)	0.0101 (0.0077–0.0147)	0.9853 (0.9778–0.9925)
0.30	0.8409 (0.8361–0.8450)	0.0010 (0.0004–0.0017)	0.9993 (0.9967–1.0000)
0.35	0.8412 (0.8377–0.8451)	0.0003 (0.0000–0.0010)	1.0000 (1.0000–1.0000)
0.40	0.8416 (0.8372–0.8456)	0.0013 (0.0000–0.0029)	0.9942 (0.9872–1.0000)
0.45	0.8412 (0.8377–0.8438)	0.0008 (0.0000–0.0020)	0.9983 (0.9961–1.0000)
0.50	0.8412 (0.8386–0.8433)	0.0002 (0.0000–0.0007)	0.9998 (0.9985–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8834 (0.8701–0.8963)	0.3809 (0.3559–0.4004)	0.5321 (0.5078–0.5521)
0.20	0.8482 (0.8415–0.8561)	0.8915 (0.8755–0.9031)	0.8693 (0.8583–0.8766)
0.25	0.8427 (0.8388–0.8468)	0.9899 (0.9853–0.9923)	0.9104 (0.9080–0.9129)
0.30	0.8415 (0.8369–0.8452)	0.9990 (0.9983–1.0000)	0.9136 (0.9108–0.9160)
0.35	0.8415 (0.8384–0.8451)	0.9995 (0.9990–1.0000)	0.9137 (0.9117–0.9160)
0.40	0.8419 (0.8379–0.8460)	0.9987 (0.9971–1.0000)	0.9140 (0.9114–0.9163)
0.45	0.8412 (0.8384–0.8443)	0.9992 (0.9984–1.0000)	0.9136 (0.9123–0.9152)
0.50	0.8420 (0.8393–0.8473)	0.9996 (0.9990–1.0000)	0.9142 (0.9126–0.9174)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1817 (0.1729–0.1898)	0.7321 (0.7017–0.7700)	0.2912 (0.2776–0.3147)
0.20	0.2113 (0.1825–0.2457)	0.1538 (0.1309–0.1831)	0.1778 (0.1576–0.2098)
0.25	0.2142 (0.1355–0.3143)	0.0147 (0.0075–0.0222)	0.0275 (0.0143–0.0412)
0.30	0.0786 (0.0000–0.2777)	0.0007 (0.0000–0.0033)	0.0015 (0.0000–0.0065)
0.35	0.0000 (0.0000–0.0000)	0.0005 (0.0000–0.0010)	0.0000 (0.0000–0.0000)
0.40	0.2250 (0.0000–1.0000)	0.0008 (0.0000–0.0034)	0.0015 (0.0000–0.0067)

0.45	0.2333 (0.0000–1.0000)	0.0006 (0.0000–0.0019)	0.0011 (0.0000–0.0038)
0.50	0.1000 (0.0000–0.7750)	0.0004 (0.0000–0.0007)	0.0008 (0.0000–0.0010)

Table 17: Mean and 95% Intervals for Loose Order BN 2 with AIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.5155 (0.5097–0.5205)	0.5219 (0.5117–0.5297)	0.2860 (0.2553–0.3110)
0.20	0.6869 (0.6794–0.6983)	0.2569 (0.2484–0.2639)	0.6142 (0.5826–0.6432)
0.25	0.8128 (0.8050–0.8207)	0.0547 (0.0448–0.0621)	0.8921 (0.8683–0.9142)
0.30	0.8368 (0.8282–0.8434)	0.0125 (0.0083–0.0168)	0.9723 (0.9598–0.9865)
0.35	0.8388 (0.8340–0.8435)	0.0038 (0.0019–0.0061)	0.9913 (0.9873–0.9943)
0.40	0.8418 (0.8374–0.8440)	0.0011 (0.0001–0.0022)	0.9968 (0.9913–1.0000)
0.45	0.8412 (0.8382–0.8468)	0.0010 (0.0001–0.0020)	0.9973 (0.9927–1.0000)
0.50	0.8421 (0.8380–0.8453)	0.0005 (0.0000–0.0010)	0.9998 (0.9985–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8988 (0.8926–0.9097)	0.4781 (0.4703–0.4883)	0.6242 (0.6178–0.6315)
0.20	0.8663 (0.8608–0.8733)	0.7431 (0.7361–0.7516)	0.8000 (0.7948–0.8074)
0.25	0.8493 (0.8470–0.8538)	0.9453 (0.9379–0.9552)	0.8947 (0.8902–0.8996)
0.30	0.8451 (0.8387–0.8512)	0.9875 (0.9832–0.9917)	0.9107 (0.9058–0.9145)
0.35	0.8413 (0.8378–0.8446)	0.9962 (0.9948–0.9981)	0.9122 (0.9100–0.9150)
0.40	0.8419 (0.8386–0.8444)	0.9989 (0.9978–1.0000)	0.9134 (0.9115–0.9163)
0.45	0.8418 (0.8386–0.8472)	0.9990 (0.9980–1.0000)	0.9137 (0.9120–0.9170)
0.50	0.8424 (0.8397–0.8473)	0.9995 (0.9990–1.0000)	0.9142 (0.9126–0.9174)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2048 (0.1987–0.2098)	0.7140 (0.6890–0.7447)	0.3183 (0.3093–0.3272)
0.20	0.2190 (0.2071–0.2399)	0.3858 (0.3568–0.4174)	0.2794 (0.2621–0.3047)
0.25	0.2714 (0.2157–0.3215)	0.1079 (0.0858–0.1317)	0.1542 (0.1229–0.1860)
0.30	0.2936 (0.1447–0.4491)	0.0277 (0.0135–0.0402)	0.0504 (0.0248–0.0738)
0.35	0.3195 (0.1942–0.5277)	0.0087 (0.0057–0.0127)	0.0169 (0.0111–0.0248)
0.40	0.5086 (0.2113–1.0000)	0.0032 (0.0000–0.0087)	0.0114 (0.0038–0.0250)
0.45	0.2905 (0.0000–0.8875)	0.0027 (0.0000–0.0073)	0.0053 (0.0000–0.0144)
0.50	0.0333 (0.0000–0.2583)	0.0002 (0.0000–0.0007)	0.0008 (0.0000–0.0010)

Additional information regarding results for the BDeu and BDs scoring functions can also be found in section A.5 of the Appendix. Note, the network structure in the case of many of these scoring functions now diverges heavily. This is likely because with decreased restrictions, the scoring functions take on a different route to finding the optimal network structure.

Homogeneous Dynamic Bayesian Network with Structure Learning

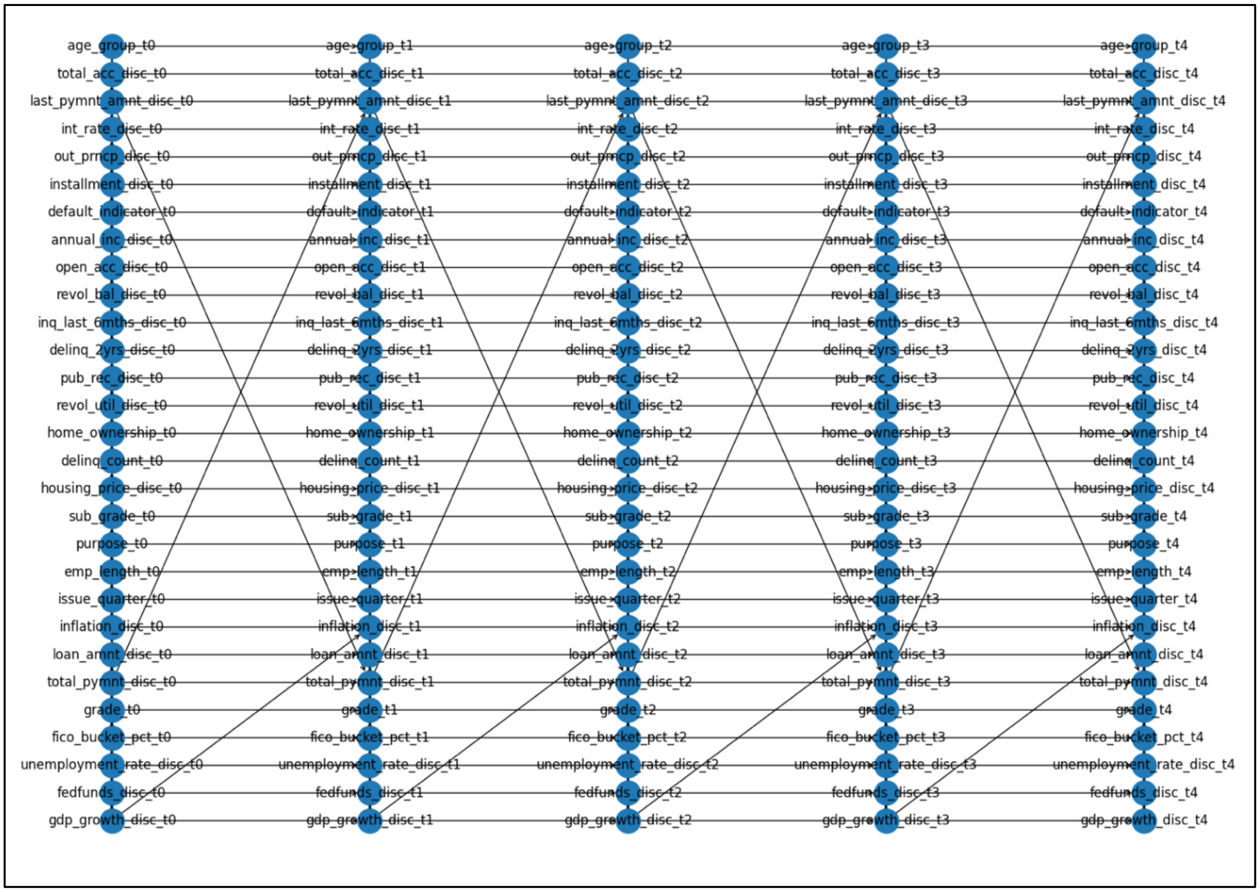
This paper will now extend the analysis on static BNs to DBNs. In this section, there is an exploration of the structure and effectiveness of Homogeneous DBNs. In doing so, the assumed structure of any DBN model in this section is that which was described in detail in the overview of BNs and DBNs earlier in this paper. Specifically, in a homogeneous DBN, the inter-slice and intra-slice edges are assumed to remain consistent over time, and each slice represents a static BN representative of that time step. The DBN structure is learned entirely from the available data as was done in the previous case with BNs.

To learn the network structure of the DBN, there are two levels of learning that must be conducted. The first level of learning is for the intra-slice edges. This process is done the same as was described for the static case of the BN. Specifically, a nonparametric bootstrap approach is taken on the time step 1 BN nodes. The resulting edges are then applied to the BNs present at each time step. The second level of learning is for the inter-slice edges, those edges which are meant to represent dependency over time between each of the BNs at each time step. The edges are learned using a nonparametric bootstrap approach as well. The edges that show up in more than 95% of the iterations are then used in the final interslice structure. These edges are learned originally between time step 1 and time step 2 BNs and then subsequently applied to all time periods. It is notable to additionally point out that in learning the network structure and conducting the bootstrap sampling approach, it was generally much more computationally expensive to do so compared to the static BNs considered previously in this section of the paper.

Demonstrating Validity

To demonstrate the validity of the network structure to accurately work in this setting with perfect information, consider the case when the intra-slice structure is the same as optimal for the Loose Structure BN 1 described previously in the paper. Then the interslice structure is learned using a BIC scoring function with the restriction that all edges from and to a node of the same type must be included in the interslice structure. The resulting network is learned below:

Figure 16: Network Structure of Example Model for Validity



When considering the results of a bootstrap approach to the classification metrics for this model, the results indicate a perfect classification rate. The reason for this is the restrictions in the model allow for perfect information in this approach, because it gives access to both the delinquency and the default at each time step to the BN at the next time step. Simply put, this model assumes that the BNPL-like firm has access to perfect information regarding the intermediary default and loan payment statuses of every individual and loan. To avoid this and simulate the results with realistic assumptions going forward, additional restrictions on knowledge to the credit firm will be placed going forward.

Building the Network

Two different DBN structures are considered in this paper. These are directly related to the previous sections with the BN. Let the first network be represented by $\boldsymbol{DBN_{HS1}} = (\boldsymbol{B_{S1}}, \boldsymbol{B_{\rightarrow}})$**.** Note here, that $\boldsymbol{B_{S1}}$ is BN Strict 1 as previously defined in the section on BN models with structure learning. Let the edges in $\boldsymbol{B_{\rightarrow}}$ **,** the inter-slice edges, be defined as having a restriction using the static and dynamic nodes at each time step. Assume that $\boldsymbol{X_{S1}} \in \boldsymbol{B_{S1}}$ can be partitioned into sets of nodes. This partition is defined as $\boldsymbol{X_{S1}} = \boldsymbol{X_{S1}^{S}} \cup \boldsymbol{X_{S1}^{D}}$ where $\boldsymbol{X_{S1}^{S}}$ represents the static nodes present at each time step and $\boldsymbol{X_{S1}^{D}}$ refers to the dynamic nodes. The static nodes are those nodes whose value is the same regardless of the time step. An example is the loan amount. The dynamic nodes are those variables which change over time. An example is GDP growth which does change on a quarterly basis. There are three important restrictions placed on the edges allowed to be learned for the inter-slice structure:

1. There may not be a directed edge from $\boldsymbol{x} \in \boldsymbol{t_i}(\boldsymbol{X_{S1}^{S}})$ to $\boldsymbol{y} \in \boldsymbol{t_{i+1}}(\boldsymbol{X_{S1}^{S}})$. This means there may not be any inter-slice edges from a static variable in the previous time period to a static variable in the next time period. These values remain the same so doing this only results in more dependency connections and increases the computation time of the model with no performance changes.

2. There may not be an edge from default in the previous time step to the delinquency count in the next time step. As previously shown when proving the validity of the model, allowing this to occur essentially inherently allows for perfect predictions based on the model's assumption of perfect information.
3. For the same reason as restriction 2, there may not be an edge from delinquency count in the previous time step to default in the next time step.

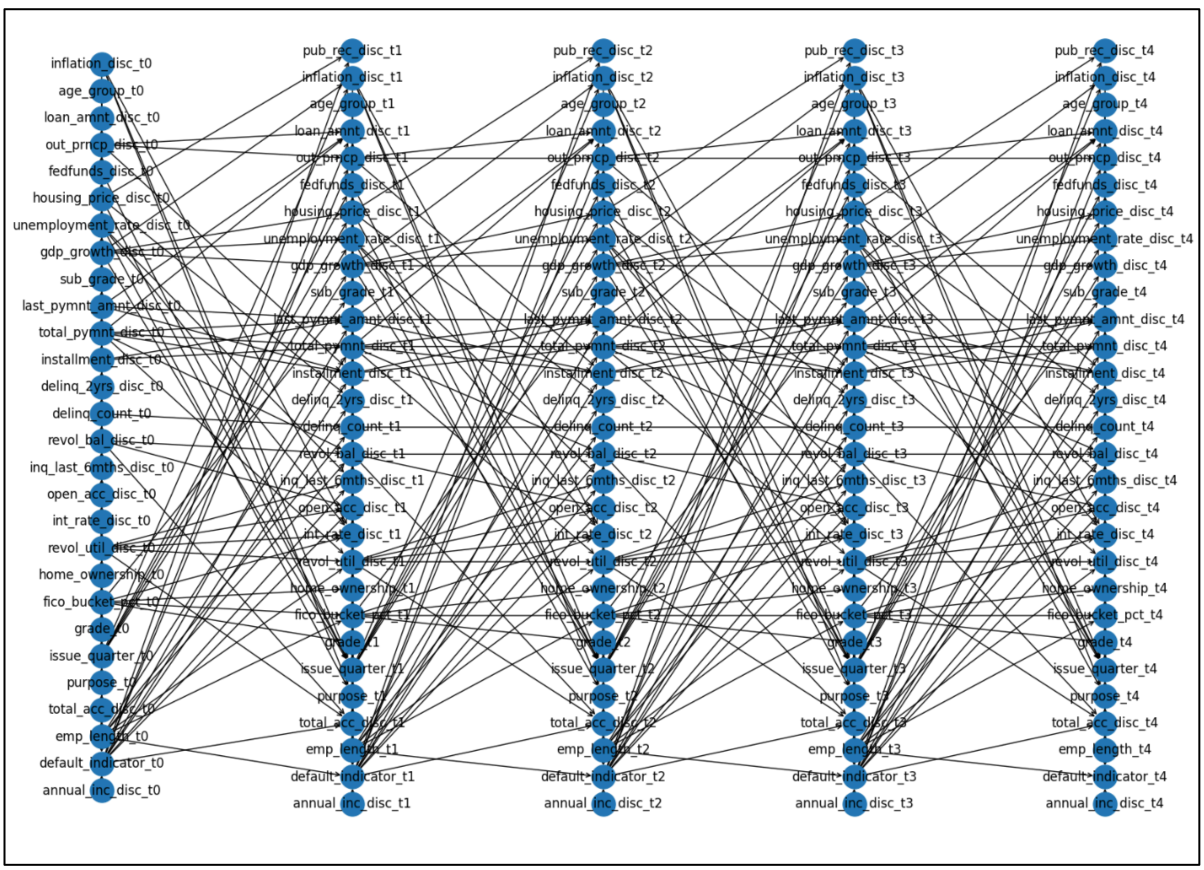
The DBN model described above, given as $\boldsymbol{DBN_{HS1}}$, will be referred to as DBN Strict 1.

Let the second network be given by $\boldsymbol{DBN_{HL1} = (B_{L1}, B_{\rightarrow})}$**.** Note, $\boldsymbol{B_{L1}}$ is BN Loose 1 as previously defined in the BN model with structure learning section. For this model, the same set of restriction to the inter-slice edges as before and those restrictions to the intra-slice edges present in BN strict 2 are the same when we consider learning the network structure of the model. In doing so, the restrictions on the edges in $\boldsymbol{B_{\rightarrow}}$ are also followed by the same as that in $\boldsymbol{DBN_{HL1}}$. The DBN model described above, given as $\boldsymbol{DBN_{HL1}}$, will be referred to as DBN Strict 2.

As mentioned in the introduction of this subsection, a non-parametric bootstrap approach is applied to find the optimal network structure under these restrictions. The example network structure for one of the DBNs for the intra-edges and inter-edges is given as follows (BIC as standardized scoring function is used for structure learning):

Figure 17: Intra-Network Structure of DBN Strict 1

Figure 18: Inter-Network Structure of DBN Strict 1



In testing the performance of these models, because there is a need to avoid the false assumption that the BNPL-like lender has access to perfect information for the future of the loan payment and the credit of the borrower, there is the use of test data with the intermediate information removed. The only information assumed that the lender would have is access to the general outlook of the economy since the macroeconomic indicators are discretized into very distinct subgroups. This results in the following means and bootstrap confidence intervals of the classification metrics:

Table 18: Mean and 95% Intervals for Strict Order DBN 1 with BIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.4461 (0.4250–0.4698)	0.5823 (0.5569–0.6071)	0.4014 (0.3316–0.4600)
0.20	0.7878 (0.7710–0.8018)	0.0815 (0.0654–0.0961)	0.9127 (0.8762–0.9490)
0.25	0.8235 (0.8125–0.8366)	0.0273 (0.0189–0.0365)	0.9707 (0.9427–0.9905)
0.30	0.8329 (0.8223–0.8444)	0.0156 (0.0083–0.0224)	0.9840 (0.9659–0.9979)
0.35	0.8354 (0.8259–0.8466)	0.0099 (0.0042–0.0164)	0.9900 (0.9753–1.0000)
0.40	0.8378 (0.8270–0.8476)	0.0066 (0.0034–0.0106)	0.9929 (0.9815–1.0000)
0.45	0.8394 (0.8307–0.8477)	0.0046 (0.0004–0.0096)	0.9961 (0.9861–1.0000)
0.50	0.8405 (0.8321–0.8497)	0.0025 (0.0000–0.0093)	0.9975 (0.9907–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8464 (0.8242–0.8674)	0.4177 (0.3942–0.4449)	0.5592 (0.5333–0.5849)
0.20	0.8438 (0.8325–0.8548)	0.9185 (0.9030–0.9346)	0.8795 (0.8706–0.8901)
0.25	0.8431 (0.8340–0.8524)	0.9727 (0.9635–0.9811)	0.9033 (0.8965–0.9104)
0.30	0.8428 (0.8342–0.8526)	0.9844 (0.9776–0.9921)	0.9081 (0.9018–0.9142)
0.35	0.8425 (0.8337–0.8518)	0.9900 (0.9836–0.9958)	0.9103 (0.9046–0.9167)
0.40	0.8425 (0.8335–0.8520)	0.9934 (0.9894–0.9966)	0.9117 (0.9052–0.9174)
0.45	0.8426 (0.8346–0.8520)	0.9954 (0.9919–0.9992)	0.9126 (0.9074–0.9175)
0.50	0.8428 (0.8344–0.8514)	0.9966 (0.9933–0.9992)	0.9133 (0.9083–0.9187)

Threshold	Precision (1)	Recall (1)	F1 (0)
0.15	0.1617 (0.1443–0.1791)	0.5971 (0.5395–0.6534)	0.2537 (0.2287–0.2806)
0.20	0.1703 (0.1048–0.2336)	0.0897 (0.0510–0.1238)	0.1144 (0.0687–0.1598)
0.25	0.1671 (0.0600–0.2941)	0.0293 (0.0095–0.0573)	0.0497 (0.0163–0.0954)
0.30	0.1616 (0.0198–0.3497)	0.0160 (0.0021–0.0341)	0.0290 (0.0037–0.0621)
0.35	0.1560 (0.0000–0.3636)	0.0100 (0.0000–0.0247)	0.0187 (0.0000–0.0451)
0.40	0.1597 (0.0000–0.3750)	0.0071 (0.0000–0.0185)	0.0136 (0.0000–0.0353)
0.45	0.1265 (0.0000–0.4286)	0.0039 (0.0000–0.0139)	0.0075 (0.0000–0.0269)

0.50	0.1175 (0.0000–0.5000)	0.0025 (0.0000–0.0093)	0.0049 (0.0000–0.0183)

Table 19: Mean and 95% Intervals for Loose Order DBN 1 with BIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.4465 (0.4274–0.4670)	0.5819 (0.5569–0.6071)	0.4014 (0.3316–0.4600)
0.20	0.7878 (0.7739–0.8056)	0.0815 (0.0654–0.0961)	0.9127 (0.8762–0.9490)
0.25	0.8244 (0.8132–0.8366)	0.0273 (0.0189–0.0365)	0.9707 (0.9427–0.9905)
0.30	0.8322 (0.8215–0.8423)	0.0156 (0.0079–0.0224)	0.9840 (0.9659–0.9979)
0.35	0.8357 (0.8259–0.8466)	0.0010 (0.0042–0.0164)	0.9900 (0.9753–1.0000)
0.40	0.8380 (0.8270–0.8476)	0.0066 (0.0034–0.0106)	0.9929 (0.9894–0.9966)
0.45	0.8394 (0.8307–0.8477)	0.0046 (0.0004–0.0139)	0.9954 (0.9917–0.9992)
0.50	0.8405 (0.8321–0.8497)	0.0034 (0.0004–0.0071)	0.9966 (0.9927–1.0000)

Threshold	Precision (0)	Recall (0)	F1(0)
0.15	0.8482 (0.8295–0.8664)	0.4181 (0.3929–0.4431)	0.5599 (0.5366–0.5843)
0.20	0.8436 (0.8328–0.8543)	0.9185 (0.9039–0.9346)	0.8795 (0.8706–0.8901)
0.25	0.8431 (0.8340–0.8524)	0.9727 (0.9635–0.9811)	0.9033 (0.8965–0.9104)
0.30	0.8428 (0.8342–0.8526)	0.9844 (0.9776–0.9921)	0.9081 (0.9018–0.9142)
0.35	0.8425 (0.8337–0.8518)	0.9900 (0.9836–0.9958)	0.9103 (0.9046–0.9167)
0.40	0.8425 (0.8335–0.8520)	0.9934 (0.9894–0.9966)	0.9117 (0.9052–0.9174)
0.45	0.8426 (0.8346–0.8490)	0.9954 (0.9923–0.9992)	0.9138 (0.9091–0.9185)
0.50	0.8419 (0.8392–0.8445)	0.9999 (0.9993–1.0000)	0.9150 (0.9118–0.9172)

Threshold	Precision (1)	Recall (1)	F1(1)
0.15	0.1610 (0.1459–0.1766)	0.5986 (0.5400–0.6684)	0.2537 (0.2291–0.2761)
0.20	0.1668 (0.1115–0.2451)	0.0873 (0.0510–0.1238)	0.1144 (0.0687–0.1598)
0.25	0.1671 (0.0600–0.2941)	0.0293 (0.0095–0.0573)	0.0497 (0.0163–0.0954)
0.30	0.1616 (0.0198–0.3497)	0.0160 (0.0021–0.0341)	0.0290 (0.0037–0.0621)
0.35	0.1560 (0.0000–0.3636)	0.0100 (0.0000–0.0247)	0.0187 (0.0000–0.0451)
0.40	0.1597 (0.0000–0.3750)	0.0071 (0.0000–0.0185)	0.0147 (0.0042–0.0164)
0.45	0.1265 (0.0000–0.4286)	0.0039 (0.0000–0.0139)	0.0096 (0.0000–0.0189)
0.50	0.1175 (0.0000–0.5000)	0.0025 (0.0000–0.0093)	0.0049 (0.0004–0.0169)

Additional results regarding loosening some of these restrictions on access of information to the lender is presented in section A.6 of the Appendix. While the results are relatively similar to that of the Static BN case, this simplified case only accounts for the fact that there is a limited credit effect from defaulting on a BNPL-like loan. There is no existing research on the cross-effect of the default rates with other multiple credit instruments with smaller consumer or BNPL loans and the CFPB does not publicly release this information. However, it is important to consider the case that defaulting on a BNPL loan at any given time period resulted from or leads to a cascading effect to the credit behavior of an individual. For example, it is possible that when defaulting on a smaller consumer loan that individuals may be simultaneously defaulting on multiple other loans, such as their mortgages or credit card debt. As a result, instead of the moderate FICO score fluctuations originally assumed in the above networks, it would be the case that there would be large swings in the credit scores of the individual under which the FICO score buckets along with loan repayment information predicted forward may be extremely relevant. An example of the results of such an analysis, assuming using previous FICO score development information of consumers, the BNPL-like firm could predict forward the FICO scores of consumers at 70% accuracy, is presented in section A.6 of the Appendix.

Partially Non-Homogeneous Dynamic Bayesian Network with Structure Learning

Next, the paper considers an alternative DBN model, that which is referred to as a partially Non-Homogenous DBN (NH-DBN). In this type of model, unlike the more restrictive homogenous DBN, it is the case that there need not be continuously replicated relationships over time or between time periods. Specifically, it may be the case that the inter-edges and intra-edges are different between each time step. For the purposes of this paper, these remain unique to the five periods of issuance of the loan and repayment. This means that while there may be changes between the time steps, there is no consideration to network structure based on when the actual date of the loan was issued to create such a network. This is because it would be highly inefficient computationally to consider each of these time periods when there already exist time varying characteristics in the data, such as the macroeconomic indicators. More importantly though, this model allows for prediction forward in a practical setting which would not be the case on using only existing, non-arbitrary time steps in terms of true issuance date of the loans.

To find the optimal network structure of the NH-DBN, the process is very similar to that of the Homogenous DBN. There are two levels of learning that must be conducted. The first level of learning is for the intra-slice edges. This process is done the same as was described for the static case of the BN. Specifically, a nonparametric bootstrap approach is taken on the nodes for each time step and the networks at each individual time step are created. The second level of learning is for the inter-slice edges, those edges which are meant to represent dependency over time between each of the BNs at each time step. The edges are learned using a nonparametric bootstrap approach as well. The edges that show up in more than 95% of the iterations are then used in final interslice structure. These edges are learned separately between each time step.

Note, practically, NH-DBNs do not practically provide much value in the context of credit risk modeling, mostly because they require that there is no assumed consistency in network structure. For example, a NH-DBN may find that GDP growth affects default directly from time step 1 to time step 2 but not at all from time step 2 to time step 3. One key aspect that drew attention to the use of Bayesian networks as a modeling technique for this setting is that the models have practical interpretations. However, this inconsistency in such models make interpretations difficult. This is on top of the fact that there is also an increase in the complexity of the conditional probability distributions with differing layers of understanding. For these reasons, there is only a study what such a model shows in terms of structure and general results.

Constructing the NH-DBN

For the purposes of this paper, two specific NH-DBN model is considered. This is based off the restrictions placed on a previously discussed model in the BN with structure learning section of this paper. Let the networks be represented by $\boldsymbol{NHDBN_{S1}} = (\boldsymbol{B_{S1}}, \boldsymbol{B_{\rightarrow}})$ and $\boldsymbol{NHDBN_{L1}} = (\boldsymbol{B_{L1}}, \boldsymbol{B_{\rightarrow}})$**.** Note here, $\boldsymbol{B_{S1}}$ is BN Strict Order 1 and $\boldsymbol{B_{L1}}$ is BN Loose Order 1 as previously defined in the section on BN models with structure learning. Note, using $\boldsymbol{B_{S1}}$ and $\boldsymbol{B_{L1}}$ is a strategic decision. Specifically, by taking advantage of the restriction on the edges going into default, there is some consistency across the BNs at multiple time slices, making this network, partially non-homogenous. Let $\boldsymbol{B_{\rightarrow}}$ have the same set of restriction placed on $\boldsymbol{DBN_{HS1}}$ so as not to assume that the lender has all the information of each time step. $\boldsymbol{NHDBN_{S1}}$ will be referred to as NH-DBN Strict Order while $\boldsymbol{NHDBN_{L1}}$ will be referred to as NH-DBN Loose Order.

As mentioned in the introduction of this subsection, a non-parametric bootstrap approach is applied to find the optimal network structure under these restrictions. The resulting example network structure for one NH-DBN for the intra-edges (time steps are chronologically ordered from left to right) and inter-edges is given as follows (BIC as standardized scoring function is used for structure learning):

Figure 20: Intra-Network Structure of NH-DBN Loose Order

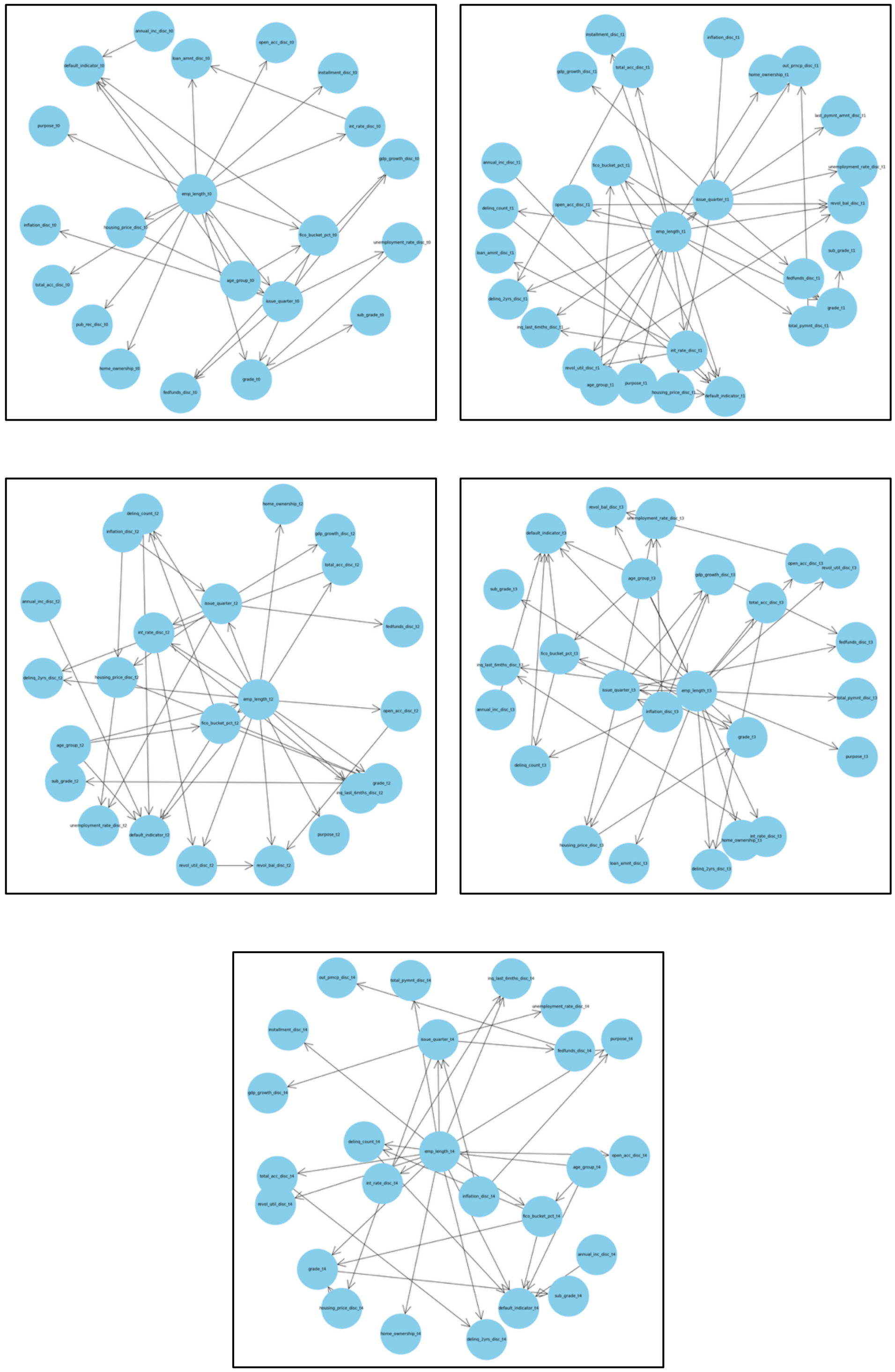
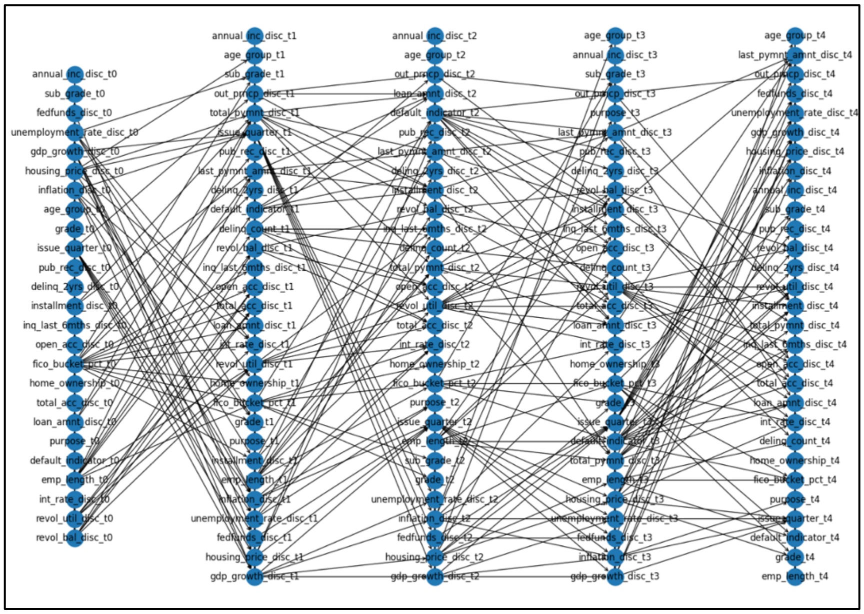


Figure 21: Inter-Network Structure of NH-DBN



In testing the performance of this models, because there is a need to avoid the false assumption that the BNPL-like lender has access to perfect information for the future of the loan payment and the credit of the borrower, there is the use of test data with the intermediate information removed. The only information assumed that the lender would have is access to the general outlook of the economy just as in the case of the homogenous DBN. This results in the following means and bootstrap confidence intervals of the classification metrics:

Table 20: Mean and 90% Intervals for Strict Order NH-DBN with BIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.5014 (0.4803–0.5174)	0.5294 (0.5071–0.5512)	0.3328 (0.2875-0.3774)
0.20	0.5874 (0.5576–0.6105)	0.3980 (0.3769–0.4342)	0.4855 (0.4511–0.5278)
0.25	0.7489 (0.7350–0.7655)	0.1550 (0.1392–0.1672)	0.7603 (0.7294–0.7916)
0.30	0.8184 (0.8116–0.8236)	0.0396 (0.0344–0.0429)	0.9469 (0.9191–0.9789)
0.35	0.8341 (0.8234–0.8437)	0.0157 (0.0116–0.0222)	0.9766 (0.9597–0.9904)
0.40	0.8361 (0.8301–0.8432)	0.0076 (0.0050–0.0104)	0.9885 (0.9870–1.0000)
0.45	0.8380 (0.8310–0.8401)	0.0024 (0.0004–0.0044)	0.9963 (0.9872–1.0000)
0.50	0.8412 (0.8380–0.8430)	0.0030 (0.0016–0.0076)	0.9988 (0.9952–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8840 (0.8702–0.8949)	0.4706 (0.4488–0.4929)	0.6141 (0.5908–0.6307)
0.20	0.8683 (0.8517–0.8796)	0.6010 (0.5658–0.6231)	0.7102 (0.6801–0.7286)
0.25	0.8548 (0.8461–0.8614)	0.8450 (0.8328–0.8608)	0.8499 (0.8399–0.8604)

0.30	0.8454 (0.8379–0.8517)	0.9604 (0.9571–0.9656)	0.8992 (0.8956–0.9026)
0.35	0.8447 (0.8346–0.8518)	0.9843 (0.9778–0.9884)	0.9092 (0.9029–0.9148)
0.40	0.8411 (0.8372–0.8450)	0.9924 (0.9876–0.9958)	0.9117 (0.9071–0.9147)
0.45	0.8422 (0.8350–0.8525)	0.9938 (0.9914–0.9966)	0.9117 (0.9066–0.9185)
0.50	0.8426 (0.8389–0.8474)	0.9963 (0.9941–0.9992)	0.9130 (0.9120–0.9176)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1896 (0.1763–0.2021)	0.6672 (0.6226–0.7125)	0.2953 (0.2750–0.3137)
0.20	0.1950 (0.1768–0.2112)	0.5145 (0.4722–0.5489)	0.2827 (0.2641–0.3061)
0.25	0.2265 (0.1961–0.2709)	0.2397 (0.2084–0.2706)	0.2328 (0.2021–0.2703)
0.30	0.1953 (0.1066–0.2681)	0.0531 (0.0211–0.0809)	0.0834 (0.0353–0.1241)
0.35	0.2165 (0.1108–0.3348)	0.0234 (0.0096–0.0403)	0.0421 (0.0175–0.0717)
0.40	0.2275 (0.0823–0.4022)	0.0116 (0.0043–0.0234)	0.0220 (0.0082–0.0440)
0.45	0.1494 (0.0004–0.4321)	0.0060 (0.0000–0.0150)	0.0115 (0.0000–0.0265)
0.50	0.1474 (0.0000–0.5000)	0.0032 (0.0000–0.0094)	0.0062 (0.0000–0.0184)

Table 21: Mean and 90% Intervals for Loose Order NH-DBN with BIC Score

Threshold	Accuracy	FPR	FNR
0.15	0.5004 (0.4868–0.5136)	0.5293 (0.5162–0.5440)	0.3397 (0.3118–0.3655)
0.20	0.5905 (0.5727–0.6089)	0.3960 (0.3777–0.4152)	0.4815 (0.4413–0.5387)
0.25	0.7541 (0.7441–0.7663)	0.1449 (0.1321–0.1554)	0.7865 (0.7484–0.8190)
0.30	0.8252 (0.8163–0.8342)	0.0294 (0.0245–0.0361)	0.9536 (0.9395–0.9725)
0.35	0.8358 (0.8248–0.8441)	0.0126 (0.0088–0.0165)	0.9819 (0.9889–1.0000)
0.40	0.8364 (0.8326–0.8432)	0.0072 (0.0042–0.0099)	0.9936 (0.9911–0.9996)
0.45	0.8411 (0.8395–0.8445)	0.0034 (0.0004–0.0069)	0.9983 (0.9966–1.0000)
0.50	0.8418 (0.8399–0.8430)	0.0023 (0.0008–0.0051)	0.9971 (0.9907–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8819 (0.8715–0.8945)	0.4707 (0.4560–0.4838)	0.6138 (0.6012–0.6273)
0.20	0.8704 (0.8586–0.8857)	0.6040 (0.5848–0.6223)	0.7131 (0.6969–0.7293)
0.25	0.8534 (0.8471–0.8609)	0.8551 (0.8446–0.8679)	0.8542 (0.8481–0.8618)
0.30	0.8450 (0.8356–0.8550)	0.9706 (0.9639–0.9755)	0.9034 (0.8979–0.9088)
0.35	0.8447 (0.8346–0.8518)	0.9874 (0.9779–0.9884)	0.9092 (0.9038–0.9148)
0.40	0.8431 (0.8360–0.8505)	0.9924 (0.9876–0.9958)	0.9108 (0.9071–0.9147)
0.45	0.8428 (0.8350–0.8525)	0.9960 (0.9931–0.9984)	0.9117 (0.9066–0.9185)
0.50	0.8423 (0.8380–0.8474)	0.9976 (0.9949–0.9992)	0.9137 (0.9107–0.9186)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1879 (0.1780–0.1963)	0.6603 (0.6345–0.6882)	0.2925 (0.2796–0.3061)
0.20	0.1964 (0.1785–0.2111)	0.5185 (0.4613–0.5587)	0.2848 (0.2567–0.3033)
0.25	0.2157 (0.1885–0.2512)	0.2135 (0.1810–0.2516)	0.2144 (0.1865–0.2468)
0.30	0.2258 (0.1427–0.2838)	0.0464 (0.0211–0.0809)	0.0767 (0.0459–0.0986)
0.35	0.2102 (0.1028–0.3348)	0.0234 (0.0096–0.0322)	0.0421 (0.0175–0.0717)
0.40	0.1928 (0.0594–0.3763)	0.0116 (0.0043–0.0234)	0.0196 (0.0082–0.0430)
0.45	0.2261 (0.0004–0.6833)	0.0065 (0.0000–0.0177)	0.0126 (0.0000–0.0284)

0.50	0.1611 (0.0000–0.5083)	0.0029 (0.0000–0.0094)	0.0066 (0.0008–0.0170)

EVALUATION

Evaluation of Theoretical Structural Results

Based on visualizations provided in the Analysis section of the various network structures, there are clear differences that exist between the structures of the networks which directly impact the prediction of the probability of default and thus the prediction of the default status of the BNPL-like loan. Each type of network focused on a different structure and those structures imply differing perspectives of what the network should look like. For the case of the naive BNs, the structures were widely taken as given. In the figures describing network structure in that section of the paper, it was clear that the focus was entirely on the prediction of default and the network is effectively useless in determining other characteristics. This is an important fundamental distinction between this type of network structure and all other structures that were considered. Notably, this structure is interesting as it effectively captures individual contributions of a node to the probability of default rather than some shared effect across multiple nodes. The effectiveness of the network swayed heavily with its structure because of this and specific examples of include the range of accuracy values seen between all threshold levels for the $\boldsymbol{B_{MI}}$ and $\boldsymbol{B_{M}}$ models. In contrast to this, when considering the network structure of various BN models with structural learning incorporated within them, while there was some influence on the network structure, the main results indicated that, even across scoring functions, the most connected nodes, or important pieces of information, would become employment length, a previous grading indicator, the FICO score buckets, and annual income. These nodes had the most outgoing directed edges into other nodes in the graph and likely because of this consistency, the results were relatively stable across multiple restrictions of learning and network assumptions. These pieces of information likely capture a lot of the dependency in the graph and the benefit of such a structure is that knowing these key pieces of information, even without all other smaller background credit information, can lead to a result better than guessing blindly when attempting to decide the final default indication status of a loan. Also, the results very slightly lean towards providing assumed temporal order to the development of the characteristics as being better than letting the network learning algorithm alone decide the intermediate structure. However, the complexity of the network is visually less sparse when doing so, and thus, there is a complexity-accuracy trade-off in learning the structure and practical use of such a network. This is an interesting contrast to what was seen in the case of the DBNs. Specifically, it was found that adding some restrictions to encourage sparsity within the inter-network structure resulted in meaningful conclusions. In using dynamic and static nodes as the framework for inter-slice restrictions, it was found that the accuracy seen in the BNs with structure learning case could effectively be replicated. However, the computational complexity of doing so increased heavily, especially when trying to use the entire network at once. The intra-slice structures were consistently formed as with the BN structures and the tradeoffs presented for that structure were consistent for the homogenous DBNs. However, when considering the NH-DBNs structure, learning the network structure was particularly interesting since variations in structure not only changed with the dynamic variables but also the underlying static variables. Overall, these network structures present some interesting takeaways about access to information and particularly about what a true BN and DBN profile may look like for a BNPL firm.

Evaluation of Implementation and Interpretability

Under the assumption that existing credit risk models can be used as a secondary model as an additional consideration to credit risk, the models have a variety of risk-reward tradeoffs being captured which are impacted by the network structure and assumptions. There may be multiple approaches that a BNPL-like credit firm may want to take with regards to the use of such a model. For example, it may be the case that the firm is offering loans in a sector with a very low overall write-off cost in the losses on such loans, such as providing loans on school supplies. In such a case, it may very well be more profitable to take on a higher risk of default while pocketing the immediate merchant fees and discounted selling price from the merchant for the credit providing firm and thus, they may care very little about stopping potential default in which case a BN with structure learning may be of the best option for them. Alternatively, if it was the case that the credit firm was very risk averse because they were selling an object with a high cost of default, such as technology-related items, the firm would be much more risk averse. They may even prefer losing the majority of non-defaulting customers with the hope of avoiding as much loss as possible from defaulters. In doing so, it would possibly be best for that firm to use a naive mixed approach at which the lowest tested threshold captures more than 70% of the defaulters while doing better than just randomly guessing on an overall accuracy perspective. The risk-reward trade-off is slightly better too as it captures more going up in threshold value but suffers in terms of overall accuracy which the firm may find important.

Another key consideration on the effectiveness of the chosen network type and structure implementation is access to key information. If it is the case that the lender has access to more information about time step payment periods, especially that which can show a significant effect between the credit risk profile of a consumer over the time steps would indicate that the lender is much better off picking a homogeneous DBN or NH-DBN model depending on the computational constraints faced by the lender. This structure may still be ideal even if the data does not show a strong trend but there is a large quantity of time-dependent data which may be used to effectively gauge the default probability in a wider variety of scenarios in the complex network framework.

When it comes to interpretability, not only do each one of these models provide their benefits in terms of explaining why an individual was selected for default, but also why it may be the case that other unseen objects would still be relevant to default. For example, consider the case of an individual who would default whose only information available is that for the naive mixed BN. It is easy to see the individual contribution of each of the available characteristics when considering the predicted default status of the individual at each threshold. This becomes slightly more complex in BNs and DBNs with a non-naive structure. However, even in these cases, as seen in the structures within the above sections, the use of the conditional probability distributions/tables can make it relatively straightforward to interpret the final probability of default in the context of known loan and credit characteristics. It may even be the case that with a subset of values, especially in learned-structure networks, that the resulting conclusion can be made more accurate with only a subset of the data but still be interpreted in the partial context of getting to that result. In essence, it may be the case that existing models of determining some input to the credit risk are inaccurate or the information fluctuates too much to be precise and thus, these models can still provide a clear link with information known. Overall, this presents a possible positive-sided argument that using such models increases the interpretability and the flexibility of using such models to logically and computationally determine the credit risk status of this loan-type. This is specifically because smaller loans, like those provided by BNPL companies, do not largely cause swings in consumers credit risk status. Those generally occur by larger debt products consumers take on and the models mentioned in this paper can additionally handle the flexibility of including or not including these characteristics depending on the preferences of the BNPL firm while not largely sacrificing the interpretability behind the computational decision made to reject or accept the loan.

CONCLUSION/DISCUSSION

In this paper, the consideration of structures of various structures of Bayesian Network models in the context of credit-risk modeling for BNPL-like firms revealed some meaningful indication of the use of such models. First, the analysis and results made clear that the models provide a clear interpretation of the reason why the default risk exists using the pre-defined discretized buckets. The cost-sensitive approach the paper took also considered the motivation behind the construction of such a model, whether it be limiting losses or providing a simple secondary check and confirmed that the use of such models can be flexible to the needs of the lender, particularly when the data is structured similarly to the proxy BNPL-like data used in the paper. The discrete BNs allow for the use of conditional probability tables to allow the lenders to adjust the classification thresholds to prioritize loss limitations or borrower acquisition. Additionally using a variety of characteristics, including that of the lagged macroeconomic variables, enables structures of BNs and DBNs for BNPL-like credit firms which are useful for inference and interpretability. The learned structure BNs reveal meaningful dependencies between borrower demographic, loan attribute, and macroeconomic factors which may be unknown to the lender. The temporal DBNs enhance the inference capabilities of the lender, capturing evolving borrower risk profiles over a short-term payment period. All the networks examined in this study remain lightweight enough to conduct quick inference, aligning with BNPL firms need for rapid decision making.

Note, an area where the analysis provides limited depth is in the assumption of a set number of maximum categories that any variable may fall in during the discretization process. While this may be a realistic assumption to make to an extent for the computational limits of this paper, depth of the analysis may be expanded by doing a sensitivity analysis of the current results to differences in the number of maximum discretized values. However, this would require a larger amount of data to effectively make meaningful conclusions and is thus left as a promising avenue of extension to the ideas presented in this paper

REFERENCES

Adhitama, R. P., and D. R. S. Saputro. 2022. Hill climbing algorithm for Bayesian network structure. *AIP Conference Proceedings* 2479(1): 020035.

Bhatore, S., L. Mohan, and Y. R. Reddy. 2020. Machine learning techniques for credit risk evaluation: A systematic literature review. *J Bank Financ Technol* 4: 111–138.

Borchani, H., A. M. Martínez, A. R. Masegosa, H. Langseth, T. D. Nielsen, A. Salmerón, A. Fernández, A. L. Madsen, and R. Sáez. 2013. Dynamic Bayesian modeling for risk prediction in credit operations. In *Frontiers in Artificial Intelligence and Applications*, 278: 17–26.

Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., and Siddique, A. 2016. Risk and risk management in the credit card industry. *Journal of Banking and Finance*, 72, 218–239.

Consumer Financial Protection Bureau (CFPB). 2022. Buy now, pay later: Market trends and consumer impacts. September. https://files.consumerfinance.gov/f/documents/cfpb_buy-now-pay-later-market-trends-consumer-impacts_report_2022-09.pdf.

Consumer Financial Protection Bureau (CFPB). 2023. Consumer use of buy now, pay later. March. https://files.consumerfinance.gov/f/documents/cfpb_consumer-use-of-buy-now-pay-later_2023-03.pdf.

Elkan, C. 2001. The foundations of cost-sensitive learning. In Proceedings of the Seventeenth *International Joint Conference on Artificial Intelligence*, Seattle, WA, August 4–10, 2001, 1.

Equifax. 2022. Buy now, pay later: Sustaining growth and building profitability. December 6, 2024. https://assets.equifax.com/marketing/US/assets/bnpl-sustaining-growth-and-building-profitability.pdf

Federal Reserve Bank of Philadelphia. 2023. 4 in 6 payment products: Buy now, pay later insights from new survey data. https://www.philadelphiafed.org/consumer-finance/4-in-6-payment-products-buy-now-pay-later-insights-from-new-survey-data.

Giudici, P., B. Hadji-Misheva, and A. Spelta. 2019. Network-based credit risk models. *Quality Engineering* 32 (2): 199–211.

Hardin, P., and R. Ingre. 2021. BNPL probability of default modeling including macroeconomic factors: A supervised learning approach (Dissertation).

Hintze, J. 2022. Buy-now-pay-later and alternative data bring disruption to retail credit. *GARP*, October 14. https://www.garp.org/risk-intelligence/credit/buy-now-pay-later-221014.

Hou, X. 2020. P2P borrower default identification and prediction based on RFE-multiple classification models. *Open Journal of Business and Management* 8: 866–880.

Kitson, N. K., and A. C. Constantinou. 2024. The impact of variable ordering on Bayesian network structure learning. *Data Mining and Knowledge Discovery* 38: 2545–2569.

Kowal, D. R. 2022. Bayesian subset selection and variable importance for interpretable prediction and classification. *Journal of Machine Learning Research* : JMLR, 23, 108.

Leão, T., S. C. Madeira, M. Gromicho, M. de Carvalho, and A. M. Carvalho. 2021. Learning dynamic Bayesian networks from time-dependent and time-independent data: Unraveling disease progression in Amyotrophic Lateral Sclerosis. *Journal of Biomedical Informatics* 117: 103730.

Leong, C. K. 2016. Credit risk scoring with Bayesian network models. *Comput Econ* 47: 423–446.

Li, X., D. Ergu, D. Zhang, D. Qiu, Y. Cai, and B. Ma. 2022. Prediction of loan default based on multi-model fusion. *Procedia Computer Science* 199: 757–764.

Liu, J., X. Zhang, and H. Xiong. 2024. Credit risk prediction based on causal machine learning: Bayesian network learning, default inference, and interpretation. *Journal of Forecasting* 43 (5): 1625–1660.

Liu, Z., B. Malone, and C. Yuan. 2012. Empirical evaluation of scoring functions for Bayesian network model selection. *BMC Bioinformatics* 13(Suppl 15): S14.

Lu, J., D. Wu, J. Dong, and A. Dolgui. 2023. A decision support method for credit risk based on the dynamic Bayesian network. *Industrial Management & Data Systems* 123.

Madaan, M., A. Kumar, C. Keshri, R. Jain, and P. Nagrath. 2021. Loan default prediction using decision trees and random forest: A comparative study. *IOP Conference Series: Materials Science and Engineering*.

Malik, M., and L. C. Thomas. 2010. Modelling credit risk of portfolio of consumer loans. *Journal of the Operational Research Society* 61 (3): 411–420.

Mandell, C. E., and Lawrence, M. J. 2023. Expanding access for the credit invisible with just four easy payments? The unregulated rise of buy now, pay later. *Loyola Consumer Law Review*, 35(275).

Martinez, E. 2021. How we investigated racial disparities in federal mortgage data. *The Markup*, August 25. https://themarkup.org/show-your-work/2021/08/25/how-we-investigated-racial-disparities-in-federal-mortgage-data.

Masmoudi, K., L. Abid, and A. Masmoudi. 2019. Credit risk modeling using Bayesian network with a lagged variable. *Expert Systems with Applications* 127: 157–166.

Mihajlovic, V., and M. Petkovic. 2001. Dynamic Bayesian networks: A state of the art. *CTIT Technical Report Series* 01-34. University of Twente.

Morales, E. A., B. M. Ramos, J. A. Aguirre, and D. M. Sanchez. 2019. Credit risk analysis model in microfinance institutions in Peru through the use of Bayesian networks. 2019 *Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*, Bogotá, Colombia, pp. 1–4.

Murphy, Kevin. 2002. Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D.

Nabende, P., S. Senfuma, and J. Nakatumba-Nabende. 2021. An evaluation of Bayesian network models for predicting credit risk on Ugandan credit contracts. In *Advances in Artificial Intelligence and Applied Cognitive Computing, Transactions on Computational Science and Computational Intelligence*, edited by H. R. Arabnia, K. Ferens, D. de la Fuente, E. B. Kozerenko, J. A. Olivas Varela, and F. G. Tinetti. Springer, Cham.

Ruffin, R., P. Ward, and A. Zhao. 2023. Understanding the psychological determinants of buy-now-pay-later (BNPL) in the UK. *International Journal of Bank Marketing*.

Shafiee Kamalabad, M., A. M. Heberle, K. Thedieck, and M. Grzegorczyk. 2019. Partially non-homogeneous dynamic Bayesian networks based on Bayesian regression models with partitioned design matrices. *Bioinformatics* 35 (12): 2108–2117. https://doi.org/10.1093/bioinformatics/bty917.

Serrano-Cinca, C., and B. Gutiérrez-Nieto. 2016. The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decision Support Systems* 89: 113–122.

Shiguihara, P., A. D. A. Lopes, and D. Mauricio. 2021. Dynamic Bayesian network modeling, learning, and inference: A survey. *IEEE Access* 9: 117639–117648.

Taş, C. 2023. Comparison of machine learning and standard credit risk models' performances in credit risk scoring of buy now pay later customers. M.S. Master of Science, Middle East Technical University.

Teles, G., J. J. P. C. Rodrigues, R. A. L. Rabê, and S. A. Kozlov. 2020. Artificial neural network and Bayesian network models for credit risk prediction. *Journal of Artificial Intelligence and Systems* 2: 118–132.

Triki, M. W., and Y. Boujelbene. 2017. Bank credit risk: Evidence from Tunisia using Bayesian networks. 93–107.

Wang, Y., Y. Zhang, Y. Lu, and X. Yu. 2020. A comparative assessment of credit risk model based on machine learning: A case study of bank loan data. *Procedia Computer Science* 174: 141–149.

APPENDIX

A.1: Relevant Data Formatting Aggregate Statistics from January 2025 BNPL Report

Table 1: BNPL Use in 2021 and 2022.

	2021	2022
Percent who borrowed with BNPL	17.6%	21.2%
Average transaction amount ($)	141	142
Median transaction amount ($)	110	108
Number of unique consumer observations	519,325	524,773

Table 2: Originations and Defaults by Credit Score Category, 2021–2022

Score Category	Share of Originations	Default Rate
No Score	3.9%	4.1%
Deep Subprime	45.0%	3.5%
Subprime	16.0%	1.1%
Near Prime	12.7%	0.8%
Prime	13.2%	0.7%
Super-prime	9.1%	0.8%
Observations: 892,668		

Table 3: BNPL Borrower Statistics for 2021 and 2022. Source: Consumer Financial Protection Bureau, January 2025.

	2021	2022
Average number of originations per consumer	8.5	9.5
Median number of originations by consumer	3	4
Average days between originations	42.4	44.8
Median days between originations	27.5	29.3
Average number of daily loans across borrowers	0.9	1.1
Median borrower's average number of daily loans	0.3	0.4
Observations	42,886	55,675

Table 4: BNPL Borrower Statistics: Heavy vs. Occasional Users for 2021 and 2022.

	2021		2022	
	Heavy Users	Occasional Users	Heavy Users	Occasional Users
Share of Users (%)	18.3	81.7	20.3	79.7
Median originations by consumer	22	2	23	2
Median days between originations	13.4	38.0	13.6	41.4
Median borrower's average daily loans	2.5	0.2	2.6	0.3
Observations	7,833	35,053	11,290	44,385

Table 5: BNPL Default Rates by Age Group, 2019–2022.

Year/Age Group	18–24	25–33	34–40	41–50	51–64	65+	Overall
Total	2.9%	2.4%	1.9%	1.5%	1.5%	1.4%	2.1%
2019	2.6%	1.9%	1.4%	1.9%	1.2%	4.8%	1.9%
2020	2.3%	1.7%	1.2%	0.9%	1.1%	1.0%	1.5%
2021	3.5%	2.9%	2.2%	1.8%	2.0%	1.3%	2.5%
2022	2.7%	2.2%	1.9%	1.5%	1.3%	1.4%	1.9%

Table 6: Credit Card Default Rates of BNPL Borrowers by Age Group, 2019–2022.

Year/Age Group	18–24	25–33	34–40	41–50	51–64	65+	Overall
Total	7.3%	10.4%	11.7%	11.0%	8.7%	6.3%	10.1%
2019	12.7%	16.0%	15.7%	14.7%	11.9%	11.5%	14.7%
2020	8.1%	12.4%	14.6%	13.5%	10.2%	7.0%	12.3%
2021	6.6%	10.0%	11.6%	11.1%	9.1%	6.7%	10.1%
2022	5.4%	8.8%	10.0%	9.6%	7.8%	6.0%	8.7%

Table 7: Share of Consumers with Open Unsecured Balances in 2022 by Borrower Group.

	Heavy BNPL Users	Occasional BNPL Users	No BNPL Use
Credit cards	89.2%	84.6%	81.5%
Personal loans	49.2%	38.4%	18.3%
Retail loans	54.9%	46.5%	36.6%
Alternative financial services	7.8%	3.9%	0.9%

Student loans	45.3%	39.7%	21.0%
Observations: Heavy (11,291), Occasional (44,384), No BNPL (469,098)			

Table 8: Conditional Correlations between BNPL Use and Non-BNPL Unsecured Debt

Outcome Variable	Mean	Coefficient	Std. Error	Magnitude (% of Mean)	Observations
Personal loan balances ($)	2,099	453.0	31.6	*** ***21.6%	6,124,561
Retail loan balances ($)	791	291.5	10.8	36.9%	6,124,561
Student loan balances ($)	15,800	5,733.6	201.4	36.3%	6,124,561
Credit card balances ($)	4,803	871.1	41.4	*** *** *** *** 18.1%	6,124,561
Payday balances ($)	0.66	0.37	0.1	55.9%	6,124,561
Other AFS balances ($)	49	28.6	2.0	58.6%	6,124,561

*** Significance at the 99% level.

Table 9: Raw Sample: BNPL Aggregate Statistics by Age Cohort, 2021–2022.

Age Group	Share of BNPL Borrowers	Unsecured Debt ($)	BNPL Purchases ($)	BNPL Share of Available Credit
18–24	37.1%	9,068	202	59.5
25–33	33.6%	20,486	242	46.9
34–40	28.7%	25,425	267	43.7
41–50	19.1%	28,991	260	40.1
51–64	9.9%	26,453	231	33.2
65+	5.9%	17,352	196	24.2
Average	21.2%	22,162	242	44.8

Table 10: Share of Credit Amount Charged Off among BNPL Borrowers

Year	BNPL Percent Charged Off	Credit Card Percent Charged Off
2020	1.9%	15.5%
2021	2.2%	14.3%
2022	1.4%	10.9%

A.2: Summary of Trends between Default Rate and Lagged Macroeconomic Indicators

Figure 1: Default Rate and Average Unemployment Rate by Issue Quarter (Time Slice 0)

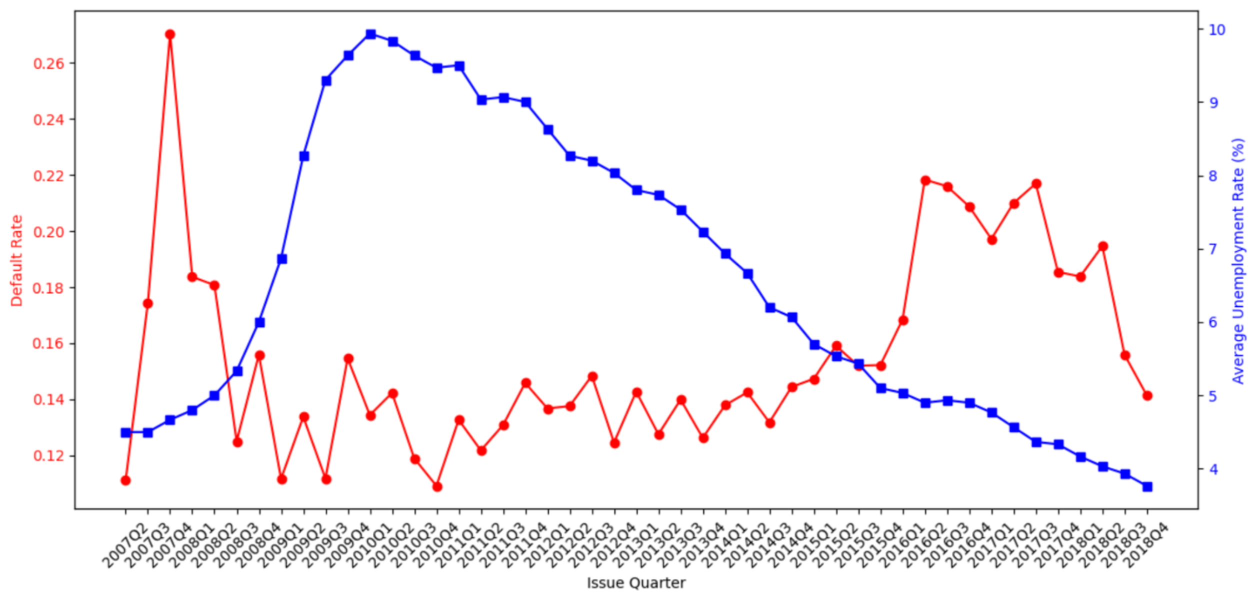


Figure 2: Default Rate and Average Housing Price by Issue Quarter (Time Slice 0)

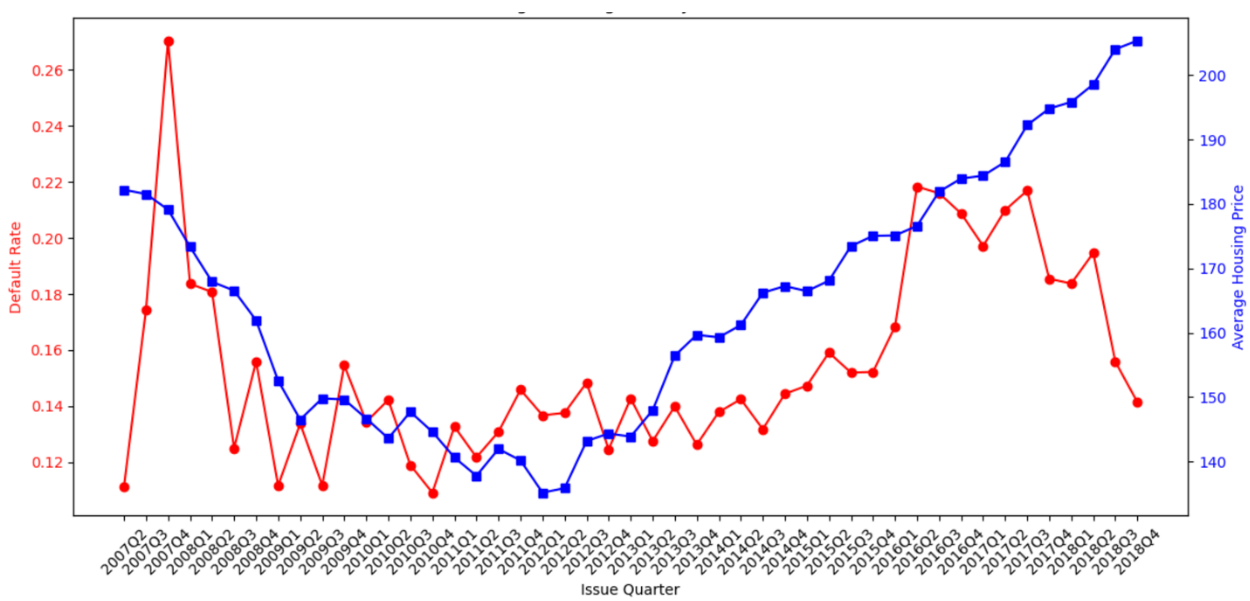


Figure 3: Default Rate and Average Inflation Rate by Issue Quarter (Time Slice 0)

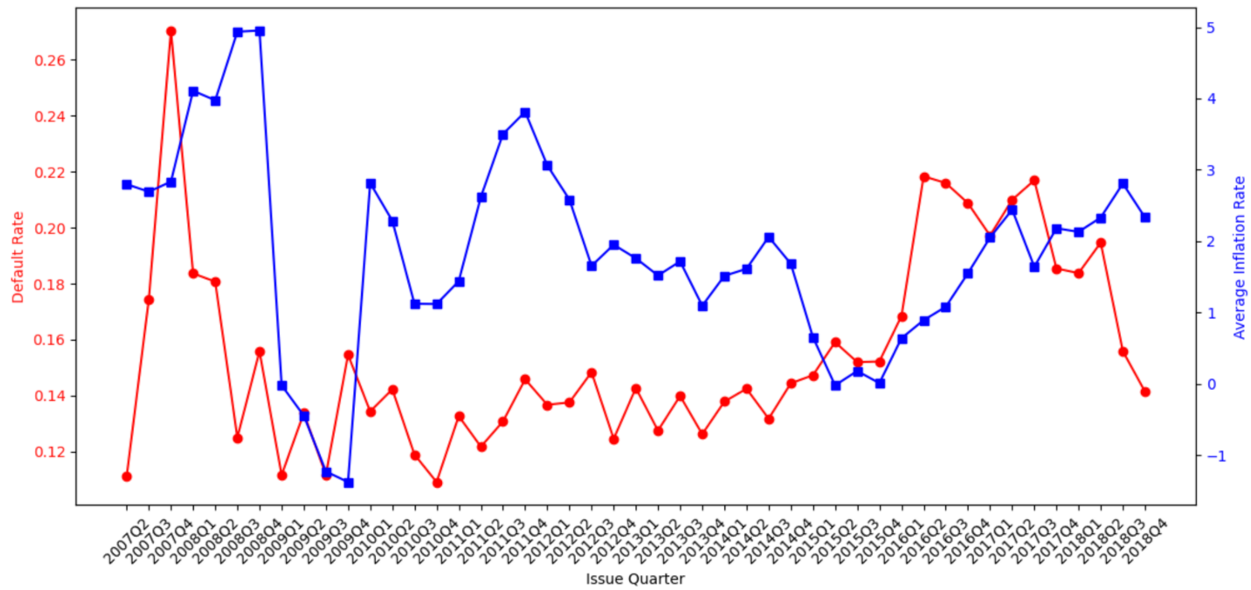
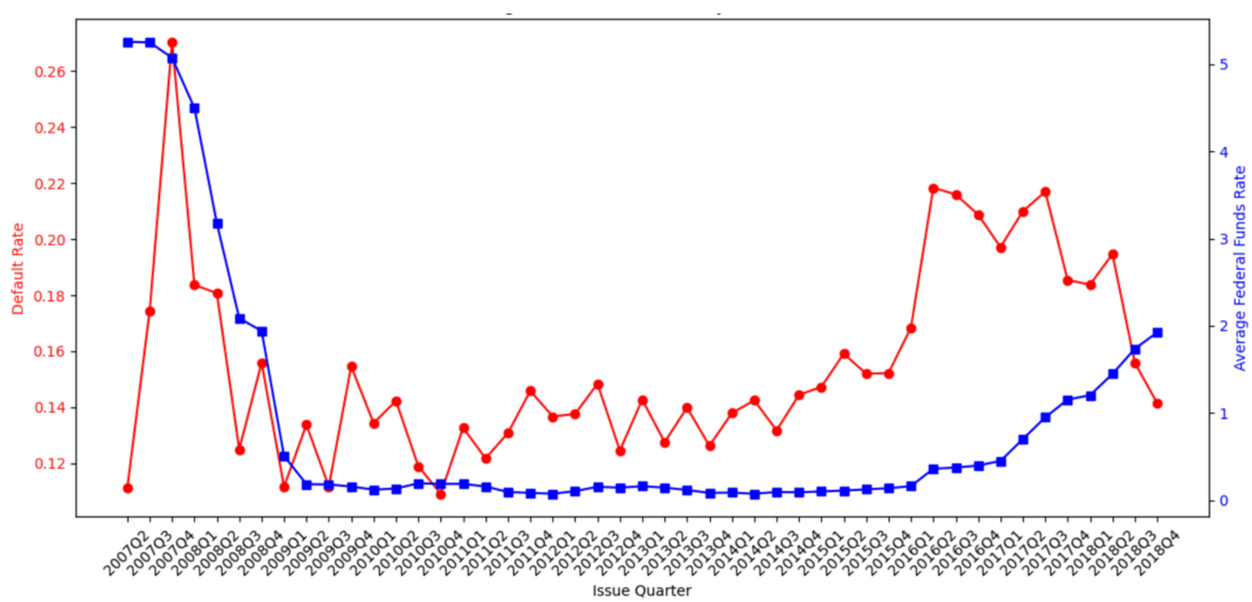


Figure 4: Default Rate and Average Federal Funds Rate by Issue Quarter (Time Slice 0)



A.3: Aggregate Statistics on Relevant Lending Club Loan Data

Table 1: Defaults by Age Group (Sample 1)

Age Group	Default	Count	Percentage
25–33	0	45,709	82.94%
25–33	1	9403	17.06%
34–40	0	30940	84.21%
34–40	1	5801	15.79%
18–24	0	45355	82.30%
18–24	1	9757	17.70%
51–64	0	15285	83.31%
51–64	1	3063	16.69%
41–50	0	15509	84.43%
41–50	1	2861	15.57%
65+	0	371	73.18%
65+	1	136	26.82%

Table 2: Defaults by Age Group (Sample 2)

Age Group	Default	Count	Percentage
25–33	0	53324	82.93%
25–33	1	10973	17.07%
34–40	0	38660	84.18%
34–40	1	7266	15.82%
18–24	0	37730	82.15%
18–24	1	8196	17.85%
41–50	0	15487	84.31%
41–50	1	2883	15.69%
51–64	0	7597	82.90%
51–64	1	1567	17.10%
65+	0	371	73.18%
65+	1	136	26.82%

Table 3: Counts of FICO Bucket by Age-Group (Sample 1)

Age Group	FICO Bucket	Count	Percentage
18–24	Deep Subprime	25639	46.52%
18–24	Near Prime	7022	12.74%

18–24	Prime	7712	13.99%
18–24	Subprime	10044	18.22%
18–24	Super-prime	4695	8.52%
25–33	Deep Subprime	26386	47.88%
25–33	Near Prime	6793	12.33%
25–33	Prime	6932	12.58%
25–33	Subprime	10110	18.34%
25–33	Super-prime	4891	8.87%
34–40	Deep Subprime	17447	47.49%
34–40	Near Prime	4507	12.27%
34–40	Prime	4727	12.87%
34–40	Subprime	6370	17.34%
34–40	Super-prime	3690	10.04%
41–50	Deep Subprime	8511	46.33%
41–50	Near Prime	2262	12.31%
41–50	Prime	2461	13.40%
41–50	Subprime	3022	16.45%
41–50	Super-prime	2114	11.51%
51–64	Deep Subprime	8467	46.15%
51–64	Near Prime	2204	12.01%
51–64	Prime	2466	13.44%
51–64	Subprime	3133	17.08%
51–64	Super-prime	2078	11.33%

A.4: Naive BN Additional Results

Table 1: Mean and 95% Intervals for Personal BN (Sample 1)

Threshold	Accuracy	FPR	FNR
0.15	0.5438 (0.5257–0.5581)	0.4724 (0.4568–0.4952)	0.3700 (0.3381–0.4076)
0.20	0.6635 (0.6504–0.6785)	0.2910 (0.2755–0.3084)	0.5784 (0.5342–0.6172)
0.25	0.7602 (0.7456–0.7724)	0.1379 (0.1271–0.1521)	0.7846 (0.7493–0.8156)
0.30	0.8027 (0.7888–0.8159)	0.0674 (0.0586–0.0780)	0.8893 (0.8656–0.9124)
0.35	0.8166 (0.8019–0.8291)	0.0423 (0.0358–0.0490)	0.9358 (0.9133–0.9594)
0.40	0.8223 (0.8112–0.8335)	0.0336 (0.0269–0.0405)	0.9450 (0.9275–0.9639)
0.45	0.8261 (0.8136–0.8392)	0.0273 (0.0213–0.0333)	0.9552 (0.9389–0.9738)
0.50	0.8372 (0.8238–0.8522)	0.0066 (0.0032–0.0108)	0.9885 (0.9796–0.9962)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8838 (0.8703–0.8998)	0.5276 (0.5048–0.5432)	0.6607 (0.6416–0.6727)
0.20	0.8669 (0.8553–0.8825)	0.7090 (0.6916–0.7245)	0.7800 (0.7690–0.7922)
0.25	0.8546 (0.8424–0.8682)	0.8621 (0.8479–0.8729)	0.8583 (0.8492–0.8667)
0.30	0.8482 (0.8364–0.8582)	0.9326 (0.9220–0.9414)	0.8884 (0.8797–0.8966)
0.35	0.8452 (0.8326–0.8578)	0.9577 (0.9510–0.9642)	0.8979 (0.8887–0.9056)
0.40	0.8449 (0.8345–0.8574)	0.9664 (0.9595–0.9731)	0.9016 (0.8946–0.9084)
0.45	0.8445 (0.8329–0.8582)	0.9727 (0.9667–0.9787)	0.9040 (0.8964–0.9117)
0.50	0.8416 (0.8283–0.8557)	0.9934 (0.9892–0.9968)	0.9112 (0.9032–0.9200)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2000 (0.1795–0.2158)	0.6300 (0.5924–0.6619)	0.3035 (0.2772–0.3227)
0.20	0.2142 (0.1855–0.2330)	0.4216 (0.3828–0.4658)	0.2840 (0.2514–0.3079)
0.25	0.2260 (0.1951–0.2582)	0.2154 (0.1844–0.2507)	0.2204 (0.1941–0.2486)
0.30	0.2358 (0.1841–0.2858)	0.1107 (0.0876–0.1344)	0.1505 (0.1206–0.1805)
0.35	0.2213 (0.1529–0.2784)	0.0642 (0.0406–0.0867)	0.0994 (0.0644–0.1306)
0.40	0.2349 (0.1652–0.2941)	0.0550 (0.0361–0.0725)	0.0890 (0.0592–0.1156)
0.45	0.2346 (0.1423–0.2979)	0.0448 (0.0262–0.0611)	0.0751 (0.0445–0.1011)
0.50	0.2516 (0.1068–0.4286)	0.0115 (0.0038–0.0204)	0.0219 (0.0073–0.0389)

Table 2: Mean and 95% Intervals for Personal BN (Sample 2)

Threshold	Accuracy	FPR	FNR
0.15	0.5458 (0.5288–0.5620)	0.4702 (0.4514–0.4888)	0.3687 (0.3287–0.4038)
0.20	0.6683 (0.6536–0.6840)	0.2855 (0.2715–0.3000)	0.5773 (0.5339–0.6198)
0.25	0.7597 (0.7444–0.7721)	0.1400 (0.1275–0.1540)	0.7768 (0.7391–0.8075)
0.30	0.8024 (0.7901–0.8153)	0.0663 (0.0585–0.0749)	0.8967 (0.8729–0.9277)
0.35	0.8176 (0.8021–0.8326)	0.0414 (0.0349–0.0511)	0.9343 (0.9097–0.9597)
0.40	0.8230 (0.8109–0.8357)	0.0327 (0.0262–0.0387)	0.9458 (0.9261–0.9631)
0.45	0.8267 (0.8132–0.8413)	0.0265 (0.0212–0.0330)	0.9559 (0.9378–0.9720)
0.50	0.8376 (0.8240–0.8526)	0.0063 (0.0033–0.0093)	0.9880 (0.9791–0.9964)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8846 (0.8715–0.9000)	0.5298 (0.5112–0.5486)	0.6626 (0.6475–0.6779)
0.20	0.8680 (0.8558–0.8814)	0.7145 (0.7000–0.7285)	0.7838 (0.7720–0.7959)
0.25	0.8555 (0.8434–0.8684)	0.8600 (0.8460–0.8725)	0.8578 (0.8478–0.8658)
0.30	0.8472 (0.8354–0.8573)	0.9337 (0.9251–0.9415)	0.8883 (0.8811–0.8962)
0.35	0.8455 (0.8330–0.8585)	0.9586 (0.9489–0.9651)	0.8985 (0.8889–0.9075)
0.40	0.8449 (0.8347–0.8571)	0.9673 (0.9613–0.9738)	0.9010 (0.8943–0.9095)
0.45	0.8445 (0.8324–0.8585)	0.9735 (0.9670–0.9788)	0.9044 (0.8961–0.9129)
0.50	0.8417 (0.8282–0.8559)	0.9937 (0.9907–0.9967)	0.9114 (0.9034–0.9203)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2011 (0.1784–0.2162)	0.6313 (0.5962–0.6713)	0.3049 (0.2744–0.3244)
0.20	0.2179 (0.1856–0.2383)	0.4227 (0.3802–0.4661)	0.2874 (0.2505–0.3097)
0.25	0.2298 (0.1880–0.2671)	0.2232 (0.1925–0.2609)	0.2262 (0.1921–0.2581)
0.30	0.2262 (0.1662–0.2705)	0.1033 (0.0723–0.1271)	0.1416 (0.1014–0.1708)
0.35	0.2294 (0.1474–0.2997)	0.0657 (0.0403–0.0903)	0.1020 (0.0646–0.1384)
0.40	0.2374 (0.1784–0.3051)	0.0542 (0.0369–0.0739)	0.0881 (0.0612–0.1168)
0.45	0.2379 (0.1429–0.3103)	0.0441 (0.0280–0.0622)	0.0743 (0.0466–0.1022)
0.50	0.2645 (0.1142–0.4039)	0.0120 (0.0036–0.0209)	0.0229 (0.0070–0.0395)

A.5: BN with Structure Learning Additional Results

Table 1: Mean and 95% Intervals - Strict Order BN 1 with BDeu Scoring

Threshold	Accuracy	FPR	FNR
0.15	0.4883 (0.4699–0.5015)	0.5469 (0.5305–0.5705)	0.3239 (0.2972–0.3618)
0.20	0.6958 (0.6830–0.7040)	0.2309 (0.2196–0.2466)	0.6956 (0.6694–0.7188)
0.25	0.8104 (0.8068–0.8127)	0.0491 (0.0448–0.0545)	0.8921 (0.8683–0.9142)
0.30	0.8289 (0.8239–0.8338)	0.0201 (0.0170–0.0226)	0.9744 (0.9620–0.9835)
0.35	0.8332 (0.8280–0.8376)	0.0122 (0.0087–0.0165)	0.9854 (0.9873–0.9943)
0.40	0.8367 (0.8322–0.8389)	0.0085 (0.0065–0.0109)	0.9959 (0.9913–1.0000)
0.45	0.8371 (0.8330–0.8401)	0.0069 (0.0042–0.0086)	0.9924 (0.9884–0.9963)
0.50	0.8406 (0.8373–0.8431)	0.0013 (0.0001–0.0024)	0.9987 (0.9961–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8819 (0.8716–0.8900)	0.4531 (0.4295–0.4695)	0.5985 (0.5772–0.6130)
0.20	0.8551 (0.8505–0.8601)	0.7691 (0.7534–0.7804)	0.8098 (0.8005–0.8143)
0.25	0.8438 (0.8411–0.8471)	0.9509 (0.9455–0.9552)	0.8942 (0.8919–0.8955)
0.30	0.8425 (0.8391–0.8467)	0.9875 (0.9835–0.9917)	0.9060 (0.9031–0.9088)
0.35	0.8415 (0.8383–0.8434)	0.9962 (0.9948–0.9981)	0.9088 (0.9057–0.9140)
0.40	0.8424 (0.8392–0.8444)	0.9989 (0.9978–1.0000)	0.9109 (0.9083–0.9121)
0.45	0.8418 (0.8386–0.8447)	0.9992 (0.9984–1.0000)	0.9112 (0.9088–0.9141)
0.50	0.8415 (0.8377–0.8439)	0.9995 (0.9990–1.0000)	0.9134 (0.9115–0.9161)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1881 (0.1788–0.1955)	0.6761 (0.6382–0.7028)	0.3239 (0.2972–0.3618)
0.20	0.1982 (0.1810–0.2399)	0.3044 (0.2812–0.3306)	0.2569 (0.2484–0.2639)
0.25	0.1878 (0.1597–0.2119)	0.0607 (0.0494–0.0715)	0.0917 (0.0754–0.1065)
0.30	0.1932 (0.1181–0.2887)	0.0256 (0.0165–0.0402)	0.0452 (0.0289–0.0670)
0.35	0.1903 (0.1001–0.3146)	0.0146 (0.0081–0.0210)	0.0271 (0.0149–0.0411)
0.40	0.2009 (0.1176–0.3405)	0.0032 (0.0000–0.0087)	0.0114 (0.0038–0.0250)
0.45	0.1732 (0.0909–0.2922)	0.0027 (0.0000–0.0073)	0.0053 (0.0000–0.0144)
0.50	0.1072 (0.0000–0.3663)	0.0003 (0.0000–0.0007)	0.0008 (0.0000–0.0010)

Table 2: Mean and 95% Intervals - Strict Order BN 2 with BDeu Scoring

Threshold	Accuracy	FPR	FNR
0.15	0.5389 (0.5317–0.5435)	0.4881 (0.4835–0.4971)	0.3172 (0.2855–0.3453)
0.20	0.7094 (0.6894–0.7193)	0.2266 (0.2138–0.2540)	0.6339 (0.6020–0.6658)
0.25	0.8136 (0.8084–0.8207)	0.0490 (0.0396–0.0571)	0.9079 (0.8833–0.9292)
0.30	0.8398 (0.8351–0.8441)	0.0051 (0.0029–0.0075)	0.9862 (0.9814–0.9902)

0.35	0.8428 (0.8400–0.8462)	0.0014 (0.0004–0.0072)	0.9961 (0.9928–0.9996)
0.40	0.8422 (0.8399–0.8458)	0.0006 (0.0001–0.0011)	0.9985 (0.9966–1.0000)
0.45	0.8426 (0.8407–0.8453)	0.0001 (0.0000–0.0004)	0.9999 (0.9981–1.0000)
0.50	0.8407 (0.8360–0.8433)	0.0001 (0.0000–0.0006)	1.0000 (1.0000–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8958 (0.8891–0.9048)	0.5119 (0.5029–0.5165)	0.6515 (0.6455–0.6546)
0.20	0.8675 (0.8601–0.8724)	0.7734 (0.7460–0.7862)	0.8177 (0.8023–0.8247)
0.25	0.8462 (0.8441–0.8496)	0.9510 (0.9429–0.9604)	0.8955 (0.8923–0.8995)
0.30	0.8431 (0.8387–0.8471)	0.9949 (0.9925–0.9971)	0.9127 (0.9100–0.9152)
0.35	0.8436 (0.8411–0.8471)	0.9986 (0.9977–0.9993)	0.9146 (0.9130–0.9166)
0.40	0.8426 (0.8401–0.8462)	0.9994 (0.9989–1.0000)	0.9143 (0.9130–0.9164)
0.45	0.8427 (0.8407–0.8456)	0.9999 (0.9996–1.0000)	0.9146 (0.9134–0.9170)
0.50	0.8408 (0.8362–0.8437)	0.9999 (0.9994–1.0000)	0.9134 (0.9115–0.9161)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2080 (0.1994–0.2145)	0.6828 (0.6547–0.7145)	0.3189 (0.3062–0.3271)
0.20	0.2315 (0.2149–0.2497)	0.3661 (0.3342–0.3980)	0.2834 (0.2629–0.3047)
0.25	0.2642 (0.2183–0.3307)	0.0921 (0.0708–0.1167)	0.1363 (0.1091–0.1722)
0.30	0.3525 (0.2324–0.4629)	0.0138 (0.0098–0.0186)	0.0265 (0.0189–0.0353)
0.35	0.3452 (0.0450–0.5775)	0.0039 (0.0004–0.0072)	0.0076 (0.0009–0.0141)
0.40	0.3167 (0.0000–0.8875)	0.0015 (0.0000–0.0034)	0.0031 (0.0001–0.0068)
0.45	0.3500 (0.0000–1.0000)	0.0008 (0.0000–0.0019)	0.0015 (0.0000–0.0038)
0.50	0.0000 (0.0000–0.0000)	0.0001 (0.0000–0.0006)	0.0003 (0.0000–0.0010)

Table 3: Mean and 95% Intervals - Strict Order BN 1 with BDs Scoring

Threshold	Accuracy	FPR	FNR
0.15	0.4878 (0.4792–0.4979)	0.5495 (0.5418–0.5604)	0.3155 (0.2833–0.3566)
0.20	0.6990 (0.6895–0.7087)	0.2279 (0.2155–0.2540)	0.6892 (0.6614–0.7207)
0.25	0.8127 (0.8037–0.8205)	0.0480 (0.0396–0.0571)	0.9079 (0.8833–0.9292)
0.30	0.8285 (0.8221–0.8375)	0.0195 (0.0146–0.0239)	0.9754 (0.9627–0.9839)
0.35	0.8343 (0.8299–0.8395)	0.0122 (0.0090–0.0151)	0.9856 (0.9813–0.9917)
0.40	0.8357 (0.8329–0.8389)	0.0085 (0.0063–0.0107)	0.9959 (0.9918–0.9974)
0.45	0.8369 (0.8337–0.8401)	0.0067 (0.0040–0.0086)	0.9933 (0.9917–0.9963)
0.50	0.8409 (0.8373–0.8434)	0.0014 (0.0004–0.0024)	0.9983 (0.9947–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8830 (0.8717–0.8947)	0.4505 (0.4396–0.4582)	0.5966 (0.5862–0.6055)
0.20	0.8561 (0.8477–0.8617)	0.7721 (0.7614–0.7845)	0.8119 (0.8050–0.8187)
0.25	0.8453 (0.8410–0.8497)	0.9520 (0.9442–0.9580)	0.8955 (0.8899–0.9002)
0.30	0.8417 (0.8375–0.8479)	0.9949 (0.9925–0.9971)	0.9058 (0.9019–0.9113)

0.35	0.8426 (0.8389–0.8462)	0.9986 (0.9977–0.9993)	0.9095 (0.9070–0.9166)
0.40	0.8414 (0.8391–0.8431)	0.9994 (0.9989–1.0000)	0.9103 (0.9087–0.9121)
0.45	0.8409 (0.8385–0.8447)	0.9999 (0.9996–1.0000)	0.9111 (0.9092–0.9153)
0.50	0.8408 (0.8380–0.8437)	1.0000 (0.9994–1.0000)	0.9135 (0.9115–0.9161)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.1908 (0.1813–0.1995)	0.6845 (0.6434–0.7167)	0.2984 (0.2829–0.3119)
0.20	0.2043 (0.1892–0.2188)	0.3108 (0.2793–0.3386)	0.2465 (0.2255–0.2609)
0.25	0.2052 (0.1763–0.2418)	0.0661 (0.0597–0.0734)	0.1000 (0.0911–0.1122)
0.30	0.1914 (0.1447–0.2677)	0.0246 (0.0161–0.0373)	0.0435 (0.0290–0.0653)
0.35	0.1626 (0.1112–0.2364)	0.0128 (0.0063–0.0188)	0.0237 (0.0119–0.0345)
0.40	0.2079 (0.1219–0.2700)	0.0119 (0.0062–0.0180)	0.0225 (0.0117–0.0336)
0.45	0.1392 (0.0173–0.3014)	0.0052 (0.0004–0.0096)	0.0099 (0.0008–0.0186)
0.50	0.0000 (0.0000–0.0000)	0.0001 (0.0000–0.0006)	0.0003 (0.0000–0.0010)

Table 4: Mean and 95% Intervals - Strict Order BN 2 with BDs Scoring

Threshold	Accuracy	FPR	FNR
0.15	0.5359 (0.5271–0.5455)	0.4948 (0.4845–0.5022)	0.3020 (0.2812–0.3330)
0.20	0.7163 (0.6998–0.7258)	0.2157 (0.2064–0.2292)	0.6456 (0.6016–0.6802)
0.25	0.8165 (0.8106–0.8243)	0.0478 (0.0390–0.0527)	0.9088 (0.8963–0.9265)
0.30	0.8388 (0.8361–0.8427)	0.0051 (0.0026–0.0077)	0.9863 (0.9760–0.9958)
0.35	0.8413 (0.8383–0.8443)	0.0014 (0.0004–0.0024)	0.9965 (0.9927–1.0000)
0.40	0.8406 (0.8375–0.8443)	0.0005 (0.0001–0.0016)	0.9985 (0.9947–1.0000)
0.45	0.8418 (0.8395–0.8445)	0.0001 (0.0000–0.0007)	0.9994 (0.9966–1.0000)
0.50	0.8416 (0.8390–0.8436)	0.0002 (0.0000–0.0010)	0.9998 (0.9985–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8985 (0.8887–0.9071)	0.5052 (0.4978–0.5155)	0.6467 (0.6404–0.6570)
0.20	0.8659 (0.8585–0.8733)	0.7843 (0.7708–0.7936)	0.8231 (0.8123–0.8299)
0.25	0.8486 (0.8436–0.8525)	0.9522 (0.9473–0.9610)	0.8974 (0.8937–0.9020)
0.30	0.8420 (0.8388–0.8460)	0.9951 (0.9923–0.9974)	0.9121 (0.9106–0.9145)
0.35	0.8422 (0.8390–0.8446)	0.9986 (0.9976–0.9996)	0.9138 (0.9120–0.9155)
0.40	0.8409 (0.8387–0.8445)	0.9995 (0.9984–1.0000)	0.9134 (0.9115–0.9156)
0.45	0.8418 (0.8395–0.8447)	0.9998 (0.9996–1.0000)	0.9146 (0.9130–0.9162)
0.50	0.8417 (0.8389–0.8450)	0.9998 (0.9993–1.0000)	0.9140 (0.9124–0.9171)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2105 (0.2001–0.2205)	0.6980 (0.6670–0.7188)	0.3235 (0.3091–0.3374)
0.20	0.2362 (0.2068–0.2550)	0.3544 (0.3198–0.3984)	0.2834 (0.2512–0.3097)
0.25	0.2639 (0.2097–0.3274)	0.0912 (0.0735–0.1037)	0.1355 (0.1088–0.1557)
0.30	0.3320 (0.1531–0.4617)	0.0137 (0.0042–0.0239)	0.0262 (0.0082–0.0450)

0.35	0.2917 (0.0000–0.6452)	0.0035 (0.0000–0.0073)	0.0068 (0.0000–0.0144)
0.40	0.2750 (0.0000–0.9250)	0.0013 (0.0000–0.0038)	0.0026 (0.0004–0.0076)
0.45	0.1167 (0.0000–0.6292)	0.0002 (0.0000–0.0015)	0.0006 (0.0000–0.0034)
0.50	0.1000 (0.0000–0.7750)	0.0002 (0.0000–0.0015)	0.0004 (0.0000–0.0029)

A.6: Homogeneous DBN with Structure Learning Additional Results

Table 1: Mean and 90% Intervals DBN Loose 1 with Looser Restrictions (Assuming that Lender has access to Revolving Utility and Revolving Balance Discretized Buckets)

Threshold	Accuracy	FPR	FNR
0.15	0.4543 (0.4401–0.4788)	0.5678 (0.5451–0.5895)	0.4270 (0.3709–0.4978)
0.20	0.7651 (0.7565–0.7752)	0.0815 (0.0654–0.0961)	0.8850 (0.8516–0.9210)
0.25	0.8169 (0.8074–0.8267)	0.0387 (0.0321–0.0445)	0.9542 (0.9348–0.9667)
0.30	0.8290 (0.8228–0.8373)	0.0207 (0.0151–0.0272)	0.9872 (0.9576–0.9958)
0.35	0.8342 (0.8271–0.8412)	0.0112 (0.0067–0.0170)	0.9946 (0.9927–1.0000)
0.40	0.8382 (0.8310–0.8440)	0.0086 (0.0050–0.0129)	0.9921 (0.9819–1.0000)
0.45	0.8394 (0.8307–0.8477)	0.0075 (0.0037–0.0139)	0.9983 (0.9861–1.0000)
0.50	0.8405 (0.8321–0.8497)	0.0054 (0.0004–0.0137)	0.9954 (0.9674–1.0000)

Threshold	Precision (0)	Recall (0)	F1(0)
0.15	0.8440 (0.8213–0.8656)	0.4322 (0.4105–0.4549)	0.5715 (0.5526–0.5942)
0.20	0.8434 (0.8383–0.8499)	0.8859 (0.8760–0.9035)	0.8641 (0.8574–0.8901)
0.25	0.8432 (0.8347–0.8523)	0.9613 (0.9555–0.9679)	0.8984 (0.8927–0.9041)
0.30	0.8431 (0.8381–0.8495)	0.9793 (0.9728–0.9849)	0.9061 (0.9024–0.9109)
0.35	0.8425 (0.8337–0.8518)	0.9888 (0.9830–0.9933)	0.9094 (0.9050–0.9136)
0.40	0.8442 (0.8385–0.8495)	0.9914 (0.9871–0.9947)	0.9119 (0.9077–0.9156)
0.45	0.8428 (0.8368–0.8459)	0.9996 (0.9987–1.0000)	0.9145 (0.9121–0.9162)
0.50	0.8444 (0.8390–0.8495)	0.9959 (0.9932–0.9984)	0.9134 (0.9110–0.9169)

Threshold	Precision (1)	Recall (1)	F1(1)
0.15	0.1587 (0.1441–0.1709)	0.5730 (0.5022–0.6291)	0.2485 (0.2240–0.2686)
0.20	0.1574 (0.1086–0.2038)	0.1150 (0.0790–0.1484)	0.1327 (0.0914–0.1668)
0.25	0.1818 (0.1294–0.2348)	0.0458 (0.0333–0.0652)	0.0731 (0.0535–0.1024)
0.30	0.1839 (0.1056–0.2782)	0.0250 (0.0127–0.0424)	0.0438 (0.0230–0.0731)
0.35	0.1754 (0.0585–0.3601)	0.0121 (0.0044–0.0223)	0.0187 (0.0082–0.0419)
0.40	0.1489 (0.0000–0.3389)	0.0079 (0.0000–0.0181)	0.0147 (0.0044–0.0170)
0.45	0.1576 (0.0000–0.3150)	0.0068 (0.0000–0.0138)	0.0129 (0.0000–0.0261)
0.50	0.1838 (0.0000–0.4050)	0.0046 (0.0000–0.0097)	0.0089 (0.0015–0.0189)

Table 2: Mean and 90% Intervals DBN Loose 1 (Assuming higher fluctuations of FICO score up to 100 points) with BIC Scoring

Threshold	Accuracy	FPR	FNR
0.15	0.5637 (0.5392–0.5897)	0.4450 (0.4229–0.4669)	0.3903 (0.3286–0.4507)
0.20	0.7984 (0.7861–0.8141)	0.0751 (0.0629–0.0867)	0.8790 (0.8491–0.9121)
0.25	0.8313 (0.8208–0.8377)	0.0191 (0.0127–0.0248)	0.9722 (0.9553–0.9909)
0.30	0.8357 (0.8289–0.8429)	0.0039 (0.0026–0.0061)	0.9872 (0.9832–0.9942)
0.35	0.8387 (0.8310–0.8458)	0.0012 (0.0004–0.0024)	0.9950 (0.9917–1.0000)
0.40	0.8406 (0.8369–0.8477)	0.0004 (0.0001–0.0016)	0.9989 (0.9962–1.0000)
0.45	0.8425 (0.8395–0.8459)	0.0002 (0.0000–0.0007)	0.9994 (0.9966–1.0000)
0.50	0.8412 (0.8380–0.8451)	0.0000 (0.0000–0.0000)	1.0000 (0.9994–1.0000)

Threshold	Precision (0)	Recall (0)	F1 (0)
0.15	0.8820 (0.8657–0.8963)	0.5550 (0.5330–0.5771)	0.6812 (0.6615–0.7030)
0.20	0.8493 (0.8425–0.8569)	0.9249 (0.9133–0.9371)	0.8855 (0.8784–0.8945)
0.25	0.8442 (0.8352–0.8510)	0.9809 (0.9752–0.9873)	0.9074 (0.9014–0.9115)
0.30	0.8422 (0.8332–0.8477)	0.9905 (0.9863–0.9942)	0.9103 (0.9064–0.9145)
0.35	0.8434 (0.8361–0.8511)	0.9932 (0.9899–0.9956)	0.9122 (0.9077–0.9163)
0.40	0.8432 (0.8356–0.8505)	0.9995 (0.9984–1.0000)	0.9140 (0.9115–0.9163)
0.45	0.8414 (0.8377–0.8447)	0.9998 (0.9996–1.0000)	0.9145 (0.9131–0.9165)
0.50	0.8419 (0.8402–0.8489)	0.9999 (0.9994–1.0000)	0.9137 (0.9120–0.9174)

Threshold	Precision (1)	Recall (1)	F1 (1)
0.15	0.2069 (0.1834–0.2296)	0.6097 (0.5493–0.6714)	0.3088 (0.2773–0.3381)
0.20	0.2334 (0.1662–0.3083)	0.1210 (0.0879–0.1509)	0.1590 (0.1139–0.1980)
0.25	0.2146 (0.0875–0.3208)	0.0278 (0.0091–0.0447)	0.0492 (0.0169–0.0760)
0.30	0.1791 (0.0633–0.3032)	0.0113 (0.0042–0.0246)	0.0212 (0.0078–0.0454)
0.35	0.1409 (0.0000–0.3636)	0.0064 (0.0000–0.0111)	0.0122 (0.0000–0.0280)
0.40	0.2667 (0.0000–0.9250)	0.0015 (0.0000–0.0038)	0.0034 (0.0000–0.0073)
0.45	0.3500 (0.0000–1.0000)	0.0006 (0.0000–0.0019)	0.0034 (0.0000–0.0076)
0.50	0.0333 (0.0000–0.7750)	0.0002 (0.0000–0.0019)	0.0004 (0.0000–0.0029)