2006

Using Neural Networks to Develop a New Model to Screen Applicants for Apartment Rentals

Michael T. Furick
Nova Southeastern University, mikefrk@bellsouth.net

Using Neural Networks to Develop a New Model to Screen Applicants for Apartment Rentals

by

Michael T. Furick

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Information Systems

Graduate School of Computer and Information Sciences
Nova Southeastern University

2006

We hereby certify that this dissertation, submitted by Michael T. Furick, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Maxine S. Cohen, Ph.D — Date
Chairperson of Dissertation Committee

Sumitra Mukherjee, Ph.D — Date
Dissertation Committee Member

Steven D. Zink, Ph.D — Date
Dissertation Committee Member

Approved:

Edward Lieblein, Ph.D. — Date
Dean, Graduate School of Computer and Information Sciences

Graduate School of Computer and Information Sciences
Nova Southeastern University

2006

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Using Neural Networks to Develop a New Model to Screen Applicants for Apartment Rentals

by
Michael T. Furick

June 2006

Credit scoring is a mathematical means of summarizing a consumer's credit and financial history into a three-digit number. This number provides an easy means of identifying and sorting consumer behavior into categories based on their financial history. To select applicants for loans and to set interest rates on loans, banks and financial institutions routinely use credit scoring. Auto insurance companies also use scoring to decide which consumers will be offered auto insurance and to set the price for auto insurance. Despite success in these two industries, scoring does not appear to be effective in the apartment rental industry in picking desirable applicants for apartment rental.

The first phase of this research analyzed the results of using six commercially available credit scores applied in one apartment complex to the task of selecting applicants. This part of the analysis answered the research question: How effective are commercially available credit scores in predicting applicant financial behavior when renting an apartment? This research determined that these six scores are not predictive and possible explanations are given.

Phase two of this research used neural networks to develop a new model using both credit data and other lifestyle data about the applicant. The hypothesis was that the addition of this lifestyle data would improve accuracy in selecting apartment rental applicants over currently available models based only on credit data. This part of the analysis answered the research question: How is the prediction accuracy of a new neural network based credit scoring model improved by adding lifestyle data to the credit report data? This research indicates that accuracy is greatly improved. Three variables were found to be most predictive for the apartment rental decision and these were a) percentage of satisfactory accounts in the applicant's credit file, b) total applicant income, and c) driving record of the applicant.

Four areas were suggested for future study and these are a) understanding the underlying human behavior differences that influence apartment financial decisions, b) addition of "fuzzy logic" techniques to the neural network, c) expanding the number of commercial credit models tested and size of the data set and d) effect of geography on model prediction accuracy. This dissertation also examined U.S. information policy and addressed consumer privacy considerations when using non-credit data to select applicants.

Acknowledgements

Conducting this research and writing this dissertation has been one of my most significant academic challenges. Without the support, patience, and guidance of a number of people, this study would not have been completed and it is to them that I owe my deepest gratitude.

First, let me thank Dr. Maxine Cohen of Nova Southeastern University who acted as my Dissertation Committee Chairperson despite her many other academic and professional commitments. Her wisdom, knowledge, and commitment to the highest standards inspired and motivated me. This dissertation simply would not have been possible without her help.

Secondly, I wish to thank Dr. Sumitra Mukherjee of Nova Southeastern University and Dr. Steven Zink of Nova Southeastern University for participating on my dissertation committee. Dr. Mukherjee's expert knowledge of artificial intelligence and neural networks steered me in the correct direction for applying these tools to this research. Dr. Zink's skills and questions helped me to produce a dissertation report that was clear, succinct, and relevant. The questions and encouragement of these three committee members were vital in helping me to complete this research.

Thirdly, I wish to thank the State Apartment Association and the employees and owners of the apartment complex who allowed me access to their tenant records. This research would not have been possible without this data and without their help in understanding the dynamics of the industry. Additionally, the local employees of Experian were particularly helpful in assisting me by providing information and an understanding of the scoring industry and the application of credit scores.

Lastly, I wish to thank the people who loved me throughout this five-year process particularly my children, Lynn, Ashley, and Steven.

This dissertation is dedicated to the memory of my wife, Susan, who through 21 years of marriage taught me to listen, and to always focus on the things that are important. "Just try to be better today than you were yesterday."

Table of Contents

Appendixes

List of Tables

Tables

List of Figures

Figures

Chapter 1

Introduction

Introduction and Background

The banking and financial services industry has used, for many years, credit report data and specifically, credit scoring as a means of determining the credit worthiness of consumers applying for loans. The intent is to weed out, or at least identify those applicants that will become questionable accounts while, at the same time, offer lower interest rates and better products to those applicants that are most desirable. Credit evaluation decisions are important for the financial institution involved due to the high risk and potential financial cost associated with a wrong decision (Piramuthu, 1998).

The advantages of credit scoring include reducing the cost of credit analysis, enabling faster credit decisions, closer monitoring of existing accounts, and prioritizing collections (Brill, 1998). Today, credit scoring is used by 97% of banks that approve credit card applications and by 82% of banks that determine whom to solicit for credit cards. Both the Federal Home Loan Mortgage Corporation and Federal National Mortgage Corporation are actively encouraging the use of credit scoring for all mortgage origination, and GE Capital Mortgage uses credit scoring for all mortgage insurance applications (Mester, 1997).

The credit scoring process generates a credit score, which is a three-digit number that predicts the likelihood that an applicant will repay a loan and repay it on time. This score is based on the data in a consumer's credit report and is the result of a process of modeling the variables important in the extension of credit. This modeling process is a statistical analysis of historical data for both good consumers and bad consumers, using certain financial variables that have been determined to be important in the evaluation of a consumer's financial strength and stability. These variables and the weighting of these variables change for each model and for differing industries. Typical variables used by the banking industry include the following (Leonard, 1996)

1. Number of bankruptcies.
2. Number of credit cards/trade line.
3. Percentage usage of these trade lines (percentage of credit limit).
4. Credit history and payment performance.
5. Length of employment.
6. Income.
7. Occupation.
8. Residential status.
9. Length of time at current address.

The analysis of these variables in the model produces coefficients that are translated into "weight scores." For example, if length of employment is longer than 10 years then add 50 points, if longer than 5 years add 25 points, otherwise add no points. Adding together these weight scores for each variable for each new loan applicant produces an overall score. The loan officer relies on this overall score in making the loan decision.

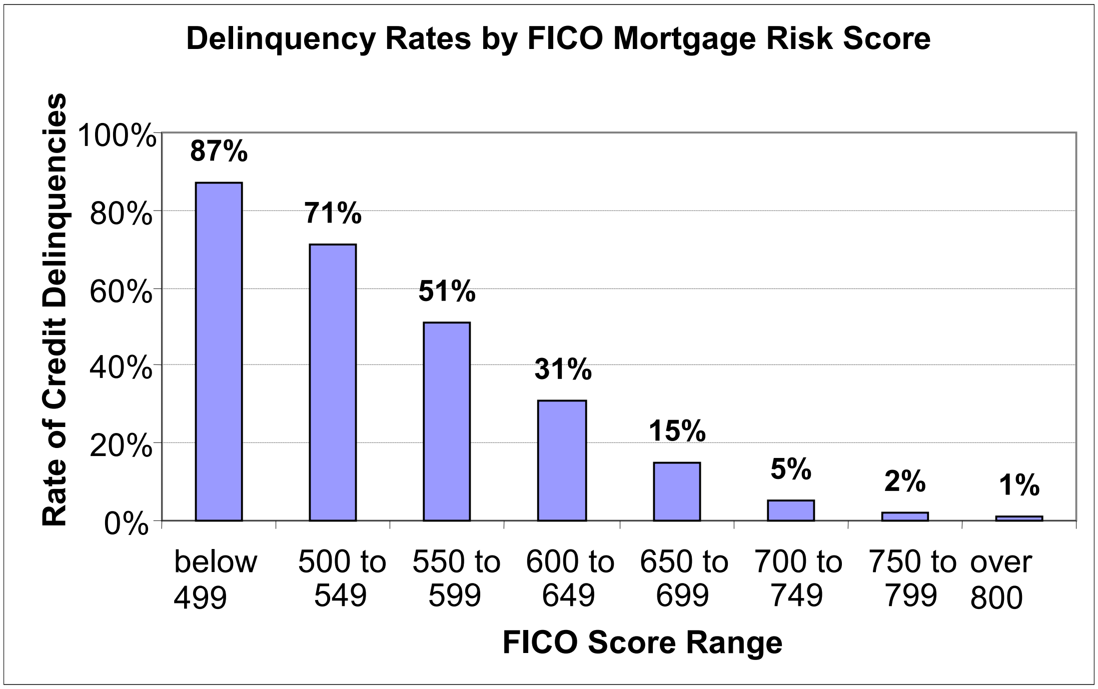
Table 1 Commercially available Credit Scoring Models

Reference Number	Name of Score
1	Crossview
2	FICO National Risk Score (used by the apartment complex in this research).
3	National Equivalency Score.
4	Old National Risk Score.
5	FICO Installment Score.
6	FICO Installment II Score.
7	FICO Automobile Score.
8	FICO Automobile II Score.
9	FICO Finance Score.
10	FICO Finance II Score.
11	FICO Bankcard Score.
12	FICO Bankcard II Score.
13	FICO Mortgage Risk Score (sold by Equifax as "Beacon"; also sold by the third credit bureau Transunion with the brand name "Empirica").
14	MDS Bankruptcy II Score.
15	Bankruptcy Watch.
16	Retail Risk Score.
17	TEC Risk Score.
18	Collection Score.
19	Collection Recovery Score (bankcard).
20	Collection Recovery Score (retail).
21	FICO Advanced Risk Score.
22	Fraud Shield.
23	Sureview Non Prime Score.
24	Automobile Risk Score.
25	Credit Union Risk Score.
26	Tella Risk Score.

Credit scoring has become widely used and accepted in the banking and financial services industries. Fair, Isaac Company (FICO) is the leading provider of these models and scores and has sold over 10 billion scores over the past 20 years. FICO estimates that over 75% of the mortgage decisions in the United States are based on one or more of its FICO credit scores (Angel, 2000). The FICO scores, and those from other companies, are available from the three U.S. credit bureaus when the loan officer orders a credit report. Each credit bureau offers a different set of scores as part of its product offering. Table 1 shows the 26 scores and models available from Experian, the second largest credit bureau (Equifax is the largest credit bureau). These scores range in price from about $0.25 to about $3.00 for each one obtained with the credit report.

Each of the scores listed in Table 1 has its own scale and direction of the scale. Some of the scores have a scale of 0 to 1000, while others have a scale from 300 to 850. Some of the scores are developed so that a higher number is better, but for other scores a lower number is better. For example, the FICO National Risk Score (number 2 in Table 1) uses a scale from 0 to 1000 and a lower score indicates a better applicant. This is opposite of the typical score where a higher score indicates a better applicant.

Figure 1 shows the typical statistics for the FICO Mortgage Risk Score that is number 13 on Table 1. This is the most widely used score in mortgage loan banking and this score has a range of 300 to 850 with a higher score indicating a better applicant. (Consumers without enough credit history to run the scoring model are given a “score” of zero.)



Source: Fair Isaac Corporation

Figure 1 Graph of delinquency rates for the FICO Mortgage Risk Score.

FICO defines the delinquency rate as the percentage of borrowers in a score range, who reach 90 days past due or worse (including bankruptcy or account charge-off) on any account on their credit report over a two year period (Fair, Issac & Company, 2002). The response of a lending institution to these scores in some cases is to deny a loan, but in more cases, their response is to adjust the interest rate on a loan to reflect their increased risk. As of January 2003, individuals with scores in the range of 700 to 719 were being quoted 5.94% for a 30-year mortgage, while those with a score of 620 to 674 were being quoted 7.63% for a 30-year mortgage (Chatzky, 2003).

This profusion in the use of credit scoring in financial transactions, particularly real estate/mortgage transactions is the result of several important advantages:

1. Eliminating subjectivity- numeric scoring eliminates much of the subjectivity associated with the credit approval process and eliminates the need for the loan officer's "gut feel," thus promoting a more consistent method of quantifying risk (Graves, 2000).

2. Reduced discrimination risk- quantifiable and consistent guidelines may eliminate discrimination in lending (Graves, 2000).

3. Faster response time to the consumer's demand for credit- the loan application process is significantly speeded up.

4. Accuracy- the use of credit scoring appears to have a high degree of accuracy in financial/mortgage transactions. A Dun & Bradstreet report determined that there is a 61% probability that applicants with a credit score in the low (bad) end of the score range will not repay a loan or will have serious late payment issues. This is compared to a 3% probability for applicants with a credit score in the high (good) end of the score range (Taylor, 2001).

The success of credit scoring in the banking industry has caused it to spread to other industries, most notably the auto insurance industry. A recent survey by Conning and Company determined that more than 90% of the insurance carriers surveyed claimed to use credit data and credit scoring, such as the FICO credit score, in their new business process for automobile coverage (Jones, 2001). This credit scoring is part of the process in determining who will get auto insurance and at what price the auto policy will be issued.

At a recent public hearing in Chicago, auto insurance representatives were repeatedly asked, "Why is there a relationship between a consumer's credit history and their auto loss ratios" (Mazer, 2001). The insurance industry representatives had no clear response as to *why* credit works, except to make the point that all the studies indicate that consumers with worse credit ratings will have more claims against their auto insurance policy than consumers with better credit ratings. A recent study (Monaghan, 2000) matched credit histories to 170,000 auto policies. Those with the best credit scores had a loss ratio of 74.1% while those with the worst credit scores had a loss ratio of 118.6%. (An auto insurance loss ratio is the amount paid out for claims on a policy divided by the premiums collected from the consumer on that same policy. So a loss ratio of 74.1% means the insurance company paid out 74.1 cents for every dollar in premiums collected, a very profitable account.) An additional study (Brockett, Shin & Kellison, 2003) compared 153,000 auto policies with their credit scores and tracked the claims in the following 12 months. The policies with the best scores averaged claims of $558 per policy, while those with the worst scores averaged $918 per policy.

Thirty-seven state governments have now enacted legislation to try to regulate the use of credit in the auto insurance underwriting process (Credit Infocenter, 2002). Since consumers in all 50 states are required to have auto insurance, these state governments think that the use of credit scores for auto insurance makes this insurance harder or more expensive to obtain (McDonald, 2003).

Statement of the problem, need for the study and research questions

The apartment complex that was studied is a 181 unit apartment complex in an older, slow growth southeastern U.S. city (name of the apartment complex is not used in the dissertation in order to protect privacy). This apartment complex has been using one of the FICO credit scores (number 2 on Table 1) as part of its new applicant process for apartment rentals since 1998. The management of the apartment complex has run the credit score on approximately 500 applicants for apartment rentals over the past five years. The opinion of the property manager is that,

> …the credit score is not very helpful in choosing applicants. It does not seem to accurately predict which applicants will honor their lease to the end. We seem to have just as many lease termination problems with people with good scores as we do with people with bad scores. (personal interview, January 12, 2003)

There does not appear to be a standard applicant selection process in the apartment rental market and credit scoring does not seem to be widely used. Seven other apartment complexes contacted have varying methods of choosing applicants with only two using any type of credit scoring (see Table 2). Possibly credit scoring is not used because the lack of success experienced by the subject complex has also been experienced by other complexes (it is not a goal of this research to investigate this). In these complexes consumer information is used as a barrier to entry, that is, credit and criminal information is used primarily to reject applicants.

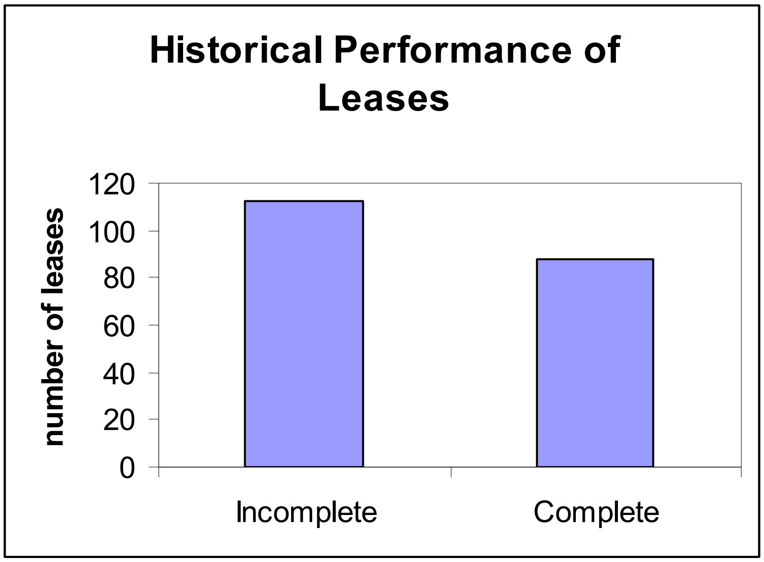
Table 2 Factors affecting Applicant Selection at Several Apartment Complexes

Apartment complex contacted	**Factors affecting applicant selection**
181 unit complex that is the subject of this research	• FICO National Risk Score used (number 2 on Table1) • Applicant rejected if previous landlord problems
96 unit complex in Baltimore, Maryland	• Applicant rejected if previous landlord problems • Applicant rejected if any criminal history • Credit scores not used
68 unit complex in Washington D.C.	• Applicant rejected if previous landlord problems • Applicant rejected if any criminal history • Credit score not used
395 unit complex in Chicago, Illinois	• FICO Advanced Risk Score used, number 13 on Table 1; minimum score must be 675 (see range on Figure 1) which is the best 15% delinquency rate of U.S. consumers • Applicant rejected if previous landlord problems • Applicant rejected if any criminal history
264 unit complex in the same city as subject complex	• Applicant rejected if previous landlord problems • Applicant rejected if any criminal history • Credit scores not used
210 unit complex in Athens, Georgia	• Applicant rejected if previous landlord problems or bankruptcy • Applicant rejected if any criminal history • Credit score not used • Income minimum three times rent • At least 80% satisfactory accounts
190 unit complex in Nashville, Tennessee	• Applicant rejected if previous landlord problems or bankruptcy • Applicant rejected if any criminal history • Credit score not used • Income minimum three times rent • At least 80% satisfactory accounts

An industry trade publication, *Rental Property Reporter*, recently conducted a survey of landlords, which showed that 33.8% of landlords in the survey ran criminal records searches, 62.6% ran credit checks, 65.5% called references, and none in the survey used credit scoring (Rental Property Reporter, 2005). A 1996 U.S. Census Bureau survey indicated that 50.6% used credit reports, 52.0% used employment/income verification, and 75.5% used personal interviews. Credit scoring was not specifically mentioned in the survey (U.S. Census Bureau, 1996).

Leases are vitally important to the success of an apartment complex because it is difficult for the management of the apartment complex to keep every apartment occupied at all times. If one tenant leaves, it takes a period of time before that vacancy can be filled and without the commitment of the lease, the managers would be constantly lining up new tenants. With the lease, however, the managers can assume that one apartment will stay occupied for a given period and focus on filling the others, thus maximizing their revenue. This expectation of the lease being fulfilled has been bundled into the price and is one of the reasons that apartment complexes charge less on a per diem basis than hotels for example, which do not have the expectation of a long stay for the tenant.

Because of the importance of leases, a sample of the past performance of applicants who moved into this apartment complex was taken by randomly selecting 200 tenants' billing records out of all the tenants in the specified population (i.e. 500 in the past 5 years). They were then divided into successes and failures based on the number on months with 12 or more months honored on their lease considered a success ("Complete" on Figure 2) while less than 12 months was considered a failure ("Incomplete" on Figure 2)



Source: Apartment complex historical data

Figure 2 Performance of 200 past tenants completing a 12 month lease.

This analysis is not comforting to the complex manager as it shows that historically about 55% of their tenants (112 in this sample) abandoned their apartments without completing their lease. As stated earlier, the purpose of a lease is to ensure a stable income on which the manager can rely when making decisions. If the majority of tenants do not honor their lease, as seen here, the apartment complex does not benefit from the lease. The 1996 U.S. Census Bureau survey found that 13% of large apartment properties have turnover exceeding 50% (another 21% have turnover between 20% and 49%).

In general, this apartment complex appears to have a history of selecting tenants who do not satisfy the terms of their lease despite using credit scoring. This lack of predictability of the credit score used at this apartment complex is in sharp contrast to the apparent success in the banking industry and the auto insurance industry. This lack of predictability has forced the management to rely on other factors in making the accept/reject decision on each applicant such as other financial ratios. These include the ratio of gross income earned to monthly rental amount, payment history at other rental properties, and other non-financial issues such as size of family, reputation at other apartment complexes and so forth. While banks have a highly predictive set of credit scoring models to help with decision making, apartments do not.

The first purpose of this study was to analyze the credit reports and credit scores of past applicants and compare these with the actual results of renting apartments to these applicants to determine if any commercially available scores are predictive of applicant/tenant behavior. The second purpose of this study was to identify other variables and factors related to the applicant that could be predictive of behavior and use these variables and factors in the development of a new more predictive credit scoring model. Seventy-six variables on each applicant were collected and these were simplified into 10 variables to be used in the building of the new model. The apartment complex managers contacted for Table 2 considered the following 10 variables to be important.

1. State of previous residence. (The managers thought that out-of-state tenants would have a higher tendency to honor the lease.)

2. Adult only, multiple adults or adult with children. (Multiple adults or adults with children would be less mobile and have a higher tendency to stay.)

3. Total applicant income. (cash available to pay debt)

4. Total *Blue Book* value of all vehicles. (High value vehicles would imply a tenant better able to handle financial obligations or conversely low value vehicles would be fully paid off thus freeing up cash for rent payments.)

5. Number of driving infractions. (background information)

6. Applicant has criminal background. (background information)

7. Total loan balance. (credit data- indication of debt load)

8. Total monthly payments. (credit data – an indication of other cash needs of tenant beside monthly rent)

9. Total credit file inquiries. (credit data – A high number of inquiries implies a tenant looking hard for credit, possibly due to financial problems not yet apparent.)

10. Percentage of total accounts that are satisfactory. (credit data - indication of tenant's tendency to reliably pay debts on time)

This application of a broader range of information to the problem of apartment applicant selection is an example of knowledge management. Enterprises are beginning to realize how important it is to "know what they know" and to be able to use this information and maximize use of the knowledge. This knowledge resides in many places, such as databases, knowledge bases, filing cabinets, and people's heads and is distributed around the enterprise. In the case of this apartment complex, management had been making tenant selection decisions based on credit score information that they believe are suspect, and other information that resides in management's head. All too often one part of an enterprise repeats the work of another part simply because it is impossible to keep track of and make use of, the knowledge in other parts or may not know the decision process of the rest of the enterprise. In this case one property manager may make decisions based on different criteria than another property manager. Therefore, enterprises need to know, a) what their knowledge assets are, and b) how to manage and make use of these assets to get maximum return.

The information and computer technology disciplines tend to focus on part a, that is, in storing the knowledge assets (such as the data provided by the credit bureaus). Knowledge assets however are broader and include the pieces of information regarding markets, products, technologies, and organizations that a business owns or knows which enable it to generate profits, add value, and succeed. The information technology disciplines alone cannot identify the key knowledge assets that need to be retrieved and stored.

Knowledge management (KM) tends to focus on part b, which is using and getting maximum return on these knowledge assets. This involves identification and analysis of the knowledge assets, and managing of the processes that act on these assets. Implementation usually involves a four-step process (Van Der Spek & Spijkervet, 1997).

1. Identifying what knowledge assets a company possesses or needs to possess. This is the feedback section for the information technology group and provides them with direction. Since there is a close working relationship, it is also the source of the confusion concerning KM as an object or a process. KM is involved in identifying and obtaining these knowledge assets (objects) but goes far beyond this.

2. Analyzing how knowledge can add value and where it can add value.

3. Specifying what actions are necessary to achieve better usability of the knowledge.

4. Reviewing the use of the knowledge to ensure that value was added In addition, to be of practical value, KM must influence what is done, how it is done, and how well it is done. Clearly then, one critical link between KM and business results is through business processes. The impact of KM on key business results might well be the greatest

through its potential for improving the performance of business processes. This is accomplished by identifying the knowledge needed to make the decisions, or take the actions that make up the process, as well as addressing the knowledge generated by those decisions and actions.

The model that was developed in this research fulfilled each of these four knowledge management steps by helping to determine what information variables are important and how these should be applied to the process of selecting applicants for apartment rentals.

Specifically the research questions that were addressed follow.

Research Questions and Goals

1. How effective are commercially available credit scores in predicting applicant financial behavior when renting an apartment?

2. How is the prediction accuracy of a new neural network based credit scoring model improved by adding qualitative lifestyle data to the credit report data?

Barriers and Issues

The goal of this research was to develop a new model for the apartment rental industry that was based on a combination of credit data and other available applicant data that would more accurately predict an applicant's performance in satisfying the apartment rental lease obligations. Developing a model of this type has been an elusive and difficult goal for several reasons.

First, credit data has been highly automated for the past 30 years but collecting data on an applicant beyond simple credit data has been difficult. It has only been in the last five years, for example, that states have begun automating their criminal records into searchable databases. Even now only 35 states (plus the District of Columbia) have databases of criminal records and only Virginia has a database that includes felonies, misdemeanors, and traffic violations. Most other states have only felony convictions or only convictions that involved jail time. Prior to even this modest automation, all criminal record searches literally involved a manual search through filing cabinets at the local or state courthouse. Most county criminal records searches are still conducted this way.

Furthermore, until recently, searching available databases required an individual search at each state (searching 50 states required 50 separate database searches). With the increased focus on terrorism since 2001, these databases are being further expanded and it is becoming easier to search all the state's databases "in mass." As a second example, obtaining information on driving history is available in an automated fashion but not vehicle ownership.

A second barrier is that it has not been clear what additional data beyond credit data will enable the model to be more predictive for a particular applicant. The working assumption of this research was that by adding lifestyle data (such as data from the criminal history, the driving record, the application and so forth) to the financial model, the financial decision accuracy would be enhanced.

Third, data beyond simple credit data is expensive to obtain. A credit report costs approximately $0.75 (or less if ordered in volume). However, other background reports are expensive (see Appendix A for typical pricing) and while credit data is available for all 50 states, the availability of other data varies by state:

1. automated national felony criminal history report costs about $11 per applicant ($4 plus $7 and available in 41 states).

2. manual state and county criminal searches costs up to $41 per applicant (usually about $17 for about 38 states and $24 per county usually from all 50 states).

3. automated driving record costs about $10 to $23 per applicant depending on the state ($3 for access plus state fees of usually $7 to $20 for all 50 states).

4. property ownership search $4.25 per applicant.

5. closed bank accounts $1.70.

6. vehicle ownership not yet available to the public.

Obtaining complete background information on a single applicant can quickly cost over $70. In addition, credit data is available from private companies that can make their data available to researchers at "no charge" if they chose to do this. FICO receives data from the three credit bureaus in this "no charge" manner and can tailor models to various industries without a cost for the raw data. Any revenue that FICO generates as a result is usually shared with the credit bureau that supplied the initial free raw data. State governments, on the other hand, control most of the other background data sources such as criminal records, driving records, and property records. State governments never (or almost never) make their data available to researchers at "no charge" because this would set a precedent in the public sector. Obtaining this data even for research purposes therefore is expensive.

Fourth, in most commercial credit scoring models, the group "most likely to be turned down for credit" has some of the following characteristics (Yin & Devaney, 1999)

1. who were renters,

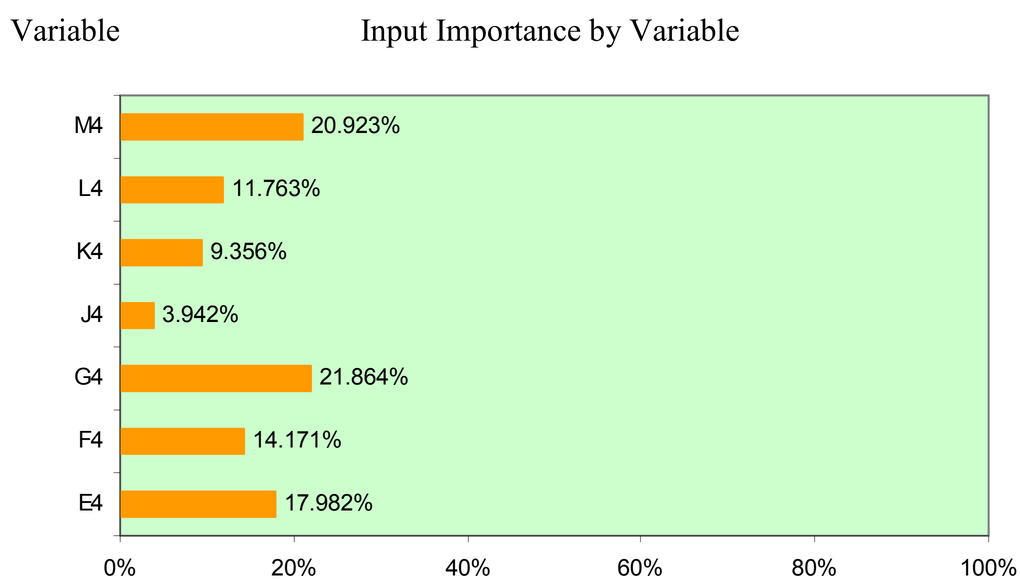
2. with less job tenure,
3. with older automobiles, and
4. with higher ratios for monthly debt payments to income.

For a banker or mortgage lender, filtering out this group would indicate a “good working” model. Unfortunately, the target applicant for many apartment complexes looks very similar to the group “most likely to be turned down for credit” with the existing credit scoring models. The challenge for the apartment complex management then is to use a scoring model to help pick out of the group “most likely turned down for credit” those applicants that are most desirable as renters (i.e. the best of the bunch)

Fifth, neural networks are a powerful tool for business decision-making (Walczak, 1999; Kim & McLeod, 1999). They have been successfully applied to solve a wide range of business applications and they work particularly well for problems involving classification and data fitting/function approximation. Neural networks often predict with higher accuracy than other statistical methods because of network capabilities of fitting any continuous function to what appears to be unrelated data. (Setiono, Leow, & Thong, 2000).

However, the main drawback of applying neural networks to solve problems of the type investigated in this research is the lack of explanation power due to the complex structure of the network and the hidden layers. In many applications, it is desirable to extract knowledge from trained neural networks in order for the user to gain better understanding of the problem at hand. Ideally, the knowledge would be expressed as symbolic rules of the form: **if** *condition*, **then** *consequence.*

This is difficult to accomplish with neural networks. For example, as shown in Figure 3, a neural network will provide information on the importance of each input variable (M4, L4, K4 etc.) in calculating the output.



Source: Forecaster XL software output

Figure 3 Importance of input variables in determining output for a neural network.

Without a rigorous program established, it is difficult to explain an acceptance or rejection to a consumer based only on these importance percentages. (The ideal explanation for a consumer would take the form "if you do this, then your score will be better and we can accept your application.") The Fair Credit Reporting Act requires that loan applicants be told why the application was declined. In addition, without a strong explanation and understanding of the internal working of the model, it is difficult at times to convince the user of the validity of the neural network. A statement such as "I can not really explain why this works, I just know that it does" does not provide a high level of confidence. As the usage of neural networks has expanded, the amount of work ongoing in this area has also expanded.

An additional area that is an issue for neural networks involves the key activity of the learning process. Human learning is composed of two parts: 1) the selection of an appropriate functional form or learning style and 2) the adjustment of parameters in the functional model to optimize some criterion or output. For most neural networks used today, the learning process consists of only number 2; that is, network architecture and learning style is usually fixed before learning begins (Nechyba & Yangsheng, 2000). Since there are hundreds of learning algorithms, this choice upfront can have unknown or possibly undesirable impacts on the network's performance. Most commercial neural network software (including the software used in this research) automatically selects the best learning algorithm based on the data.

Sixth and lastly, the period during which data was collected from the apartment complex is a time of uncertainty and economic hardship in the rental industry. The extreme drop in interest rates (Figure 4) has caused many would-be renters to purchase houses instead of renting and has reduced the demand for rental housing. The reduction in interest rates and thus reduction in rental demand has caused a corresponding reduction in occupancy levels as measured by the U.S. Census Bureau (Figure 5).

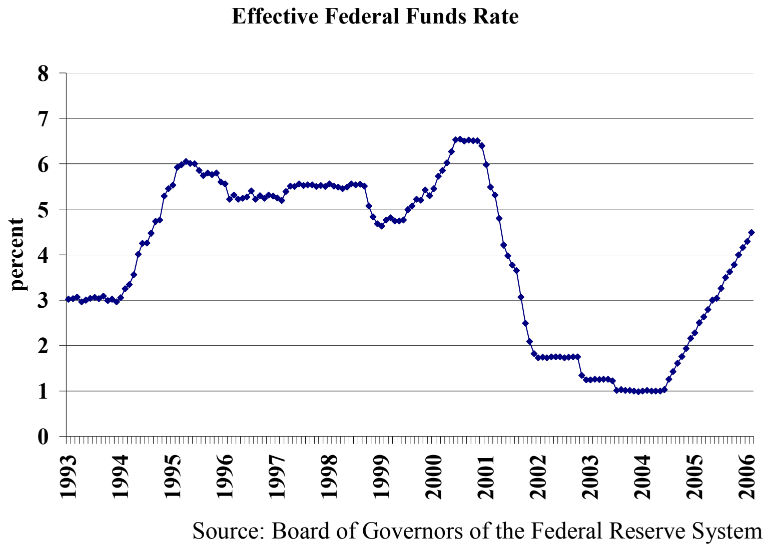


Figure 4 Federal funds interest rate.

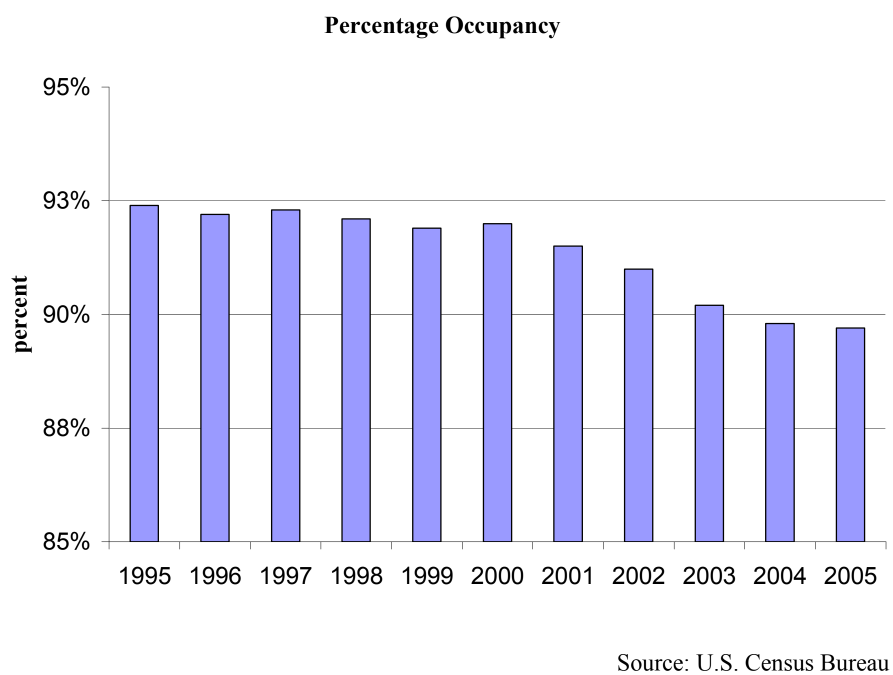


Figure 5 U.S. apartment occupancy trends.

Essentially the data that was used in the research was collected during an unusually bad economic time in the industry. This bad time may involve renters with inherently bad credit since these could be the persons unable to qualify to purchase a house and thus the renters still “left” in the rental market. The data may therefore be naturally biased toward riskier applicants. Since the goal of this research was to create a model to pick the best applicants from the available pool, this data bias would imply that there could be more undesirable applicants in the pool to be considered. Nonetheless, a working model will identify the most desirable applicants for apartment rental that is the previously mentioned “best of the bunch.”

Limitations

1) The data available only supported analysis of those applicants who were allowed to rent an apartment. There undoubtedly were applicants who were not approved for an apartment in this apartment complex, and presumably, these declined applicants would have gone on to rent an apartment somewhere else. No data is available to determine the eventual outcome of these initially declined applicants. This study therefore only analyzed the results of the applicants who received an initial positive approval and subsequently moved into the apartment complex. This is an example of a classic problem of “sample selection” and is a known problem in credit scoring (Greene, 1998). Essentially, the new model was constructed from a non-random sample, that is, only those applications that were accepted.

In this case, the ability to analyze the results of the declined applicants in addition to the accepted applicants would help determine the accuracy of the scope of the model. Specifically, was the new model selecting <u>all</u> the good applicants out of the potential pool of applicants or are some good applicants being declined here and then becoming good applicants at the next apartment complex. An analysis of this type would help determine if the screening of the new model was too tight, eliminating some good applicants. In general, since the market application of the new model was to screen applicants for entrance into an apartment complex, the limitation of using accepted applicants was not significant in this case.

2) This research analyzed the effectiveness of nationally available credit risk scores as applied to an apartment complex in one geographic area, specifically a southeastern U.S. city. However, would outcomes have changed and the model been weighted differently, if the city had been located in the northwest U.S. instead of the southeast U.S.? Since the credit scoring models used are national models, it is assumed that this impact was minimal on this research. However, some research has found that local economic factors show significant correlations with credit scores (Avery, Bostic, Calem & Canner, 2000).

The impact of local economic conditions is a concern when local banks and financial institutions use national credit scores. To address this concern, local banks and financial institutions usually adjust their procedures by changing the minimum acceptance levels for local conditions rather than trying to adjust a scoring model. For example, a Bank of American branch in Minneapolis may use a minimum score for loan approvals that is higher than a similar Bank of America branch in Dallas.

Experian, one of the three major credit reporting agencies recently released a study (Table 3) ranking cities according to credit scores (Experian, 2003). The average credit score for the U.S. was 678.

Table 3 Selected City Ranking by Credit Score

Metro Area	Credit Score for surveyed population
Minneapolis	707
Boston	705
Washington DC	693
Seattle	691
Cleveland	690
Philadelphia	688
New York	688
San Francisco	686
Chicago	680
Sacramento	676
Denver	675
Tampa	675
Detroit	675
Miami	672
Orlando	671
Atlanta	670
Los Angeles	667
Phoenix	660
Houston	655
Dallas	653

Scores for selected cities listed in Table 3 are based on the FICO Mortgage Risk Score, which is number 13 on Table 1 (also sold as Beacon and as Empirica). This score has a range from 300 to 850. About 11% of the surveyed population ranks above 800 with another 29% ranking between 750 and 799. Those with credit scores below 620 are considered “credit challenged” and pay significantly higher interest rates when borrowing money. It was not within the scope of this project to research the effect of geography on outcomes.

3) The apartment complex under study has a certain style and price range and attracts a certain type of tenant (specifically this complex was mostly blue collar, single person, or single parent with annual incomes in the $18,000 to $29,000 range). Other more expensive or less expensive apartment complexes, or those with more or fewer amenities would likely attract different types of tenants and this may change the outcomes of the research or the model to be developed. Specifically, research in this area could find that multiple models are necessary based on socio-economic factors, status of the applicant, size of apartment and so forth. The scoring model for applicants for a $500 per month, 2-bedroom apartment may need to be different from the scoring model for applicants for a $1500 per month, 2-bedroom apartment. It was not within the scope of this project to research this impact, if any. Please note that in the mortgage banking industry, there is only one model used for all applicants for home purchases (such as Equifax's "Beacon" score) regardless of the value of the home. Since one model is used in mortgage banking across all socio-economic levels, it is reasonable to assume, therefore, that one model should work across all socio-economic levels in the apartment rental industry as well.

4) The new model that was developed by this research used the data from 60 new applicants to the apartment complex. While the data collected per applicant was extensive, the number of applicants (60) is considered low for the development of a commercial model. Unfortunately, 60 applicants was the maximum number available due to the expense and the extent of the involvement of the apartment complex and the credit bureau. Nonetheless, 60 applicants were a sufficient number to identify additional data characteristics and create a more predictive model as 60 applicants for this apartment complex represents about 50%-60% of their yearly applicants.

Furthermore, Jensen (1992) developed a multilayer neural network for credit scoring with three outcomes: obligation charged off (11.2%), obligation delinquent (9.6%) and obligation paid off (79.2%). Jensen reported a correct classification result of 76-80% with a false positive rate (bad credit risk classified as good credit) of 16% and a false positive rate (good credit classified as bad credit) of 4%. Jensen concluded that the neural network had good potential for credit scoring with results developed on only 50 examples.

Summary

Credit scoring is widely used in a number of industries as an aid in helping managers to make financial decisions concerning the loans and leases made to consumers. In general, these scores are considered (and in many cases proven) to be accurate predictors of consumer performance in meeting financial obligations. However, the use of these general commercially available credit scores is not predictive when applied to consumer behavior in renting an apartment. This study analyzed the effectiveness of commercially available credit scores when applied to apartment rental decisions and developed a new model that used other data in addition to credit data to improve model predictability.

Chapter 2

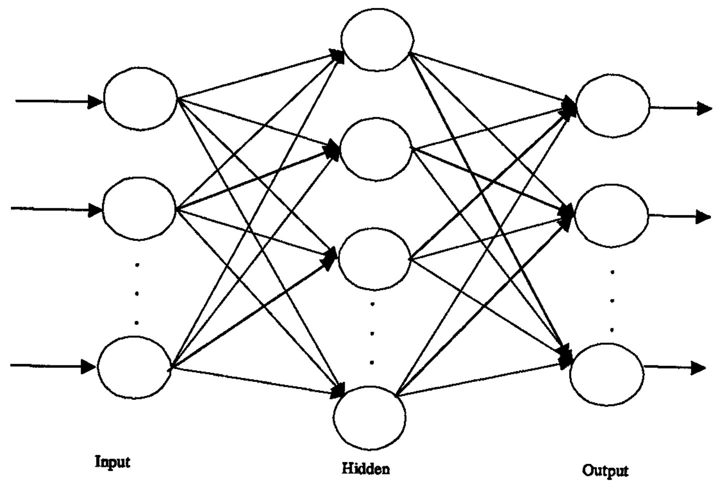
Review of the Literature

Historical Overview of the Research Literature

Credit scoring is a way of using advanced statistics to separate a population into groups based on differing risks and characteristics. Scoring can recognize the different groups in a population when the characteristics that separate the groups cannot be clearly identified (Bugera, Konno & Uryasev, 2002). Fisher (1936) first introduced the concept of separating a population into subgroups using statistics. Durand (1941) was the first to recognize that the same statistical techniques could be used to distinguish between good and bad loans. Bill Fair and Earl Issac formed the first consulting firm to commercialize scoring techniques in San Francisco in the early 1950s. At the time, their clients were primarily finance houses, retailers, and mail order firms (Fair, Issac Company today is the largest provider of credit scoring products in the U.S. and is known as FICO Inc.). The arrival of credit cards in the late 1960s made the banks and other credit card issuers realize the usefulness of scores since scoring resulted in a 50% or more drop in their loan default rates (Churchhill, Nevin, & Watson, 1977). The success of scoring for credit cards meant that banks started using scoring in the 1980s for other products such as personal mortgage loans. Retailers in the 1990s started using scoring to increase the response rate of advertising campaigns. Sears used scoring at that time to decide to whom to send its paper catalogs (Lewis, 1992).

During the 1980s and 1990s, logistic regression and linear programming were the main modeling techniques used to build scores. Today, with the improvement in computer technology and software, artificial intelligence techniques like expert systems and neural networks are used. Expert systems are knowledge-based systems that mimic the behavior of an expert and provide an aid to decision making. These are automated versions of rule based systems where the rules are derived from interviewing former "experts" on a subject such as loan approval (hence the name "expert system".) Two of the more well known expert systems used in banking are/were MARBLE (Managing and Recommending Business Loan Evaluation) (Shaw & Gentry, 1998) and CLASS (Commercial Loan Analysis Support System) (Duchessi, Shawky, & Seagle, 1988). However, expert systems lack robustness and flexibility and are difficult to create and modify. Their key advantage is that the expert system can clearly identify to the consumer, the reasons that a decision is made. As a technology, expert systems are being or have been replaced by neural network scoring systems, which is the dominant technology used to build scoring products today.

A neural network is a computer-intensive, algorithmic procedure for transforming inputs into desired outputs using inter-connected networks of relatively simple processing elements (often termed neurons, units, or nodes). Neural networks are modeled following the neural activity in the human brain. The essential features of a neural network are the nodes, the network architecture describing the connections between the nodes, and the training algorithm used to find the values or weights of each node in a particular network. A simple representation of a neural network with one hidden layer can be shown as in Figure 6 (Rumelhart, Hinton, & Williams, 1986).



Source: Rumelhart, Hinton, & Williams

Figure 6 Representation of a one layer neural network.

The high degree of action and interaction between inputs, hidden layers, and outputs gives the neural network its ability to analyze large amounts of data to establish patterns and characteristics in situations where rules are unknown and where there is a high degree of interdependence among attributes and/or many hypotheses are to be pursued in parallel. However, because of this complexity, neural networks do not produce an explicit model and thus lack explanation capabilities (Turban & Aronson, 2001). Specifically what input needs to change and by how much in order to change an output? This is a serious issue when neural networks are used for credit scoring, as it is impossible to explain to a consumer with any accuracy, those items in their credit file that most influenced their credit score. To alleviate this difficulty, each credit score given to a consumer also includes the most heavily weighted factors that affected that score. Examples of these factors follow.

1. presence of derogatory public records information.

2. presence of non-satisfactory ratings on accounts or lack of open accounts.

3. non-satisfactory ratings on revolving bank accounts or lack of revolving bank accounts.

4. credit available on satisfactory revolving bank accounts or lack of satisfactory revolving bank accounts.

Neural networks have become widely used in financial analysis since the late 1980s and early 1990s. There is a substantial amount of literature examining credit scoring and mathematical methods in general financial situations. Tam and Kiang (1992) compare neural networks with older techniques such as logistic regression, linear classifier, knn models, and ID3 models to predict bank failures. They conclude that neural networks are more accurate, adaptive, and robust. Swales and Yoon (1992) apply neural networks to differentiate among stocks that perform poorly or perform well. Lacher, Coats, Sharma, and Fants (1995) use neural networks to predict the financial health of a corporation. Studies by Dutta and Shekhar (1998) and Surkan and Singleton (1991) illustrate the use of neural networks to generate improved risk ratings of bonds. Altman (1994) employs neural networks to predict corporate financial distress among 1,000 Italian companies.

There is also a large body of literature analyzing neural networks more specifically in the area of credit applications and credit scoring. The literature of the 1980s and early 1990s tended to focus on the mathematical and statistical basis for the use of neural networks as applied to individual credit. Reichart (1983) examined the conceptual issues involved in developing credit scoring models. Jensen (1992) examined, specifically, the use of neural networks for credit scoring applications. Henley (1995) looked at the statistical issues of credit scoring and Henley later (1996) compared it to a k-nearest neighbor classifier. Altman (1994) examined the specific performance differences between linear discriminant analysis and neural networks. Cheng (1994) performed a detailed review of neural networks from a statistical perspective.

Once the statistical underpinnings of neural networks had been adequately examined, the literature of the late 1990s and today tended to focus on the use, improvement, and expansion of credit scoring as a concept. Brill (1998) looked at the importance of credit scoring models in improving cash flow and collections. Mester (1997) of the Federal Reserve examined the financial situations when credit scoring can best be applied. Thomas, Hand, and Jacka (1998) recommended methods for classifying applicants using credit data and credit scoring. Platts and Howe (1997) looked at the development of a single European credit scoring system.

The latest literature seems to be beginning to focus on credit scoring as applied to specific applications or specific industries rather than just statistical analysis or applications that are more general. Desai, Convay, Crook, and Overstreet (1997) examined credit scoring models as used in the credit union environment. Edelman (1997) applied credit scoring for lending to small businesses. Marteli, Panichelli, Strauch, and Taylor-Schoff (1997) determined the effectiveness of credit scoring as applied to high minority area populations. Monaghan (2000) examined the use of credit scoring data in the process of underwriting and issuing auto insurance policies. Emel, Oral, Reisman, and Yolalan (2003) determined the effectiveness of credit scores used in the commercial banking sector. Banasik, Crook, and Thomas (2001) created scoring models to predict usage of a credit card, not just approval for a credit card offer. The research conducted here continued this latest trend in the literature as it examined credit scoring models as applied to one industry, specifically the apartment rental industry and more specifically to one aspect, that of selecting applicants. It also continued the trend of applying the latest development tool of neural networks to develop the model.

Background and Definition of Neural Networks

Long before computers, humankind had developed conventional problem solving methodologies in an attempt to quantify and automate the solving of complex problems. Statistical models such as regression or forecasting, management science models for inventory level determination and resource allocation, and financial models for make versus buy decisions and equipment allocation have provided good results with problems that can be clearly defined. These statistical methods have existed for decades. However, as computer technology has progressed, human ability to address problems of ever-increasing complexity has also progressed. Unfortunately, the existing conventional statistical problem solving methods could not provide adequate results.

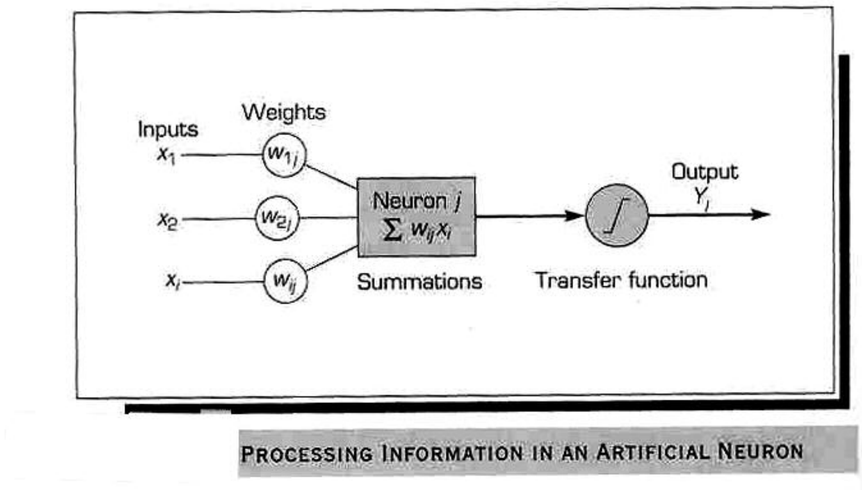
The solution has been the development of a group of computer-based problem solving methodologies usually known as machine learning or artificial intelligence. Machine learning refers to computer technologies that learn to refine their knowledge capabilities and accuracy from experience with historical cases. It is an attempt to teach machines to solve problems by showing them historical cases. Unlike traditional software programs that once programmed do not change, these machine-learning technologies learn from experience. The two most commonly used techniques are a) expert systems and b) neural networks. These have operating similarities but each of the problem solving methods addresses a different problem and these two techniques are not applicable to the same types of problems. In principle, expert systems represent a logical symbolic problem solving approach whereas neural networks are model based and use numeric and associative processing.

Expert systems are best applied to problems where inputs can be precise and these inputs lead to logical outputs. These outputs are determined by the system based on established facts and rules that have been developed through questions, conversations, and formalization of the job performance of expert persons in the field being studied (hence the name "expert system"). Expert systems are particularly useful for interacting with the user to define a problem and bringing in the facts and the solutions unique to the problem being solved. Decision tree logic is a form of a simple expert system as decision trees attempt to use defined rules and defined pathways to lead the user ("If this happens, then take that action").

A limitation in the application of expert systems arises in that the facts and rules must be gathered from experts in the field. Unfortunately, these experts do not always think of their problem solving ability in terms of rules. In addition, experts may not be able to explain their line of reasoning or they may explain it inaccurately. Thus with some problems and some experts, it is difficult to build an accurate knowledge base of facts and rules or it is simply too expensive to build this knowledge base. However, the major limitation in the use of expert systems is that the problem being solved must have clear inputs and by using definable rules, can produce acceptable outputs. Not all problems are able to be this clearly defined.

This limitation in the type of problem that could be solved with expert systems led to the development of neural networks. Neural networks are a problem solving methodology that attempts to mimic the functions of the brain. Learning is accomplished by analogy, by discovery, by observation, and by analyzing examples.

A network is composed of three main processing elements organized into units to form the network. These elements (sometime called layers) are the inputs, intermediary layers with transfer function, and outputs. Each unit (called a neuron and represented by a network node notation) represents an activity. Each of the neurons receives inputs, processes the inputs, and delivers a single output. The input can be raw data or the output of some other neuron. The output can be the final output or it can be used as the input into the next neuron as shown in Figure 7 (Turban & Aronson, 2001). Each unit of the network has associated software that performs an accounting of its inputs by computing a weighted sum. If the weighted sum exceeds a certain threshold value an output is generated otherwise the neuron continues calculating (Pfleeger, 2001). These internal layers of weights, summation function, and transfer function are usually hidden from the developer and user and, as such, are referred to as the "summations" or "hidden layers." This basic operation makes neural networks particularly effective when the relationship between inputs and output is unknown and/or the relationship between two or more inputs is unknown.

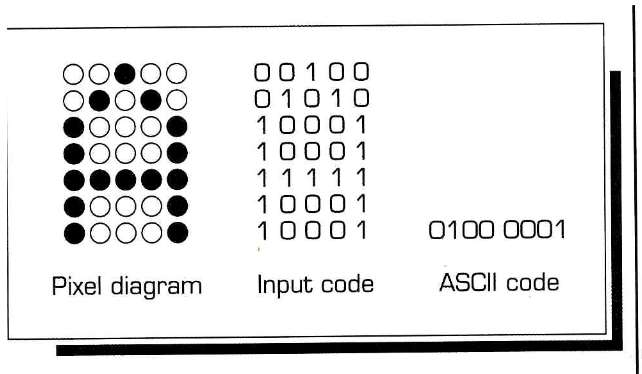


Source: Turban and Aronson

Figure 7 Neural network diagram with basic components.

Inputs

Each of the inputs corresponds to a single attribute. For example, in a loan application, each input could be a characteristic of the applicant such as income level, age, home ownership, and so forth. In determining the makeup of a batch of steel, the inputs could be the type or chemistry of the materials, amounts of materials, and temperatures of the process among others. Each attribute must be represented as a numeric value in order to be used as an input as neural networks can process only numbers. If a problem-solving attempt included qualitative data or pictures, these must be converted to a type of numeric scale. Qualitative data can be converted to numeric with questions such as "How strongly did the respondent feel about this subject on a scale from one to ten." An interesting problem arises when some of the neural network inputs are represented as pictures. Pictures must be converted to numeric data and a significant challenge is the design of a suitable coding system so that the data can be used. This could be accomplished for a black and white picture, for example, by converting each pixel to a one or zero. Using this type of coding system an "A" could be expressed in its ASCII format as in Figure 8 (Turban et al, 2001).



Source: Turban and Aronson

Figure 8 Pictorial conversion of qualitative data into quantitative data.

Transformation functions with weights and summations

Weights express the relative strengths of the input data (through mathematical values) and attempt to describe the connections between layers. The weights are a mathematical attempt to establish and identify the relative importance of each input in determining the output. Weights are crucial in that they store learned patterns of information through repeated adjustments. It is through the repeated adjustments that the network learns. The neural network is constantly changing and adjusting these weights as experience accumulates. The summation function computes the weighted sum of all the input elements entering the processing elements. This quantifies the impact that multiple neurons could have on a single processing element. The transformation function defines the relationships between the inputs (with their weights and summations) and the final output. This relationship can be linear or non-linear and the selection of this mathematical equation can have an impact on the accuracy of the network. The sigmoid transfer function has been shown to work reliably and is the standard used in neural networks, but other transformation functions have been developed for specialized applications.

Sigmoid transfer function: $Y^t = 1/(1+e^{-y})$

Learning

The neural network learning process is the process by which the software identifies patterns in the data that lead to certain outputs. The actual learning process starts with the setting of some values for the weights, either by some known rules or randomly. The software then begins to compare the output using the initial weights against the desired output for the given set of inputs. The objective is to minimize the difference between the produced output and the desired output by adjusting the weights on all the inputs. This learning process is usually accomplished on a set of data known as "training data." Training data is a collection of known inputs and known outputs that represent the correct solution to the problem. Several iterations of the complete training data are required to produce a consistent set of weights (Principe, Euliano & Lefebvre, 2000).

Having the neural network work with both known inputs and known outputs is referred to as supervised learning. However, one of the strengths of neural networks is its ability to do unsupervised learning. In unsupervised learning, only input data are shown to the network. The network becomes self-organizing in that it organizes itself internally so that each processing element is optimized and responds to different sets of inputs. No knowledge is supplied about which outputs are correct and those that the network derives may or may not have meaning. This process is useful for cluster analysis and to understand how various inputs may be affecting each other in addition to affecting the output. This is helpful as a first step, when very little or nothing is known about the solution to the problem. This research used supervised learning since outputs were known.

The actual technique of learning usually has the neurons look backward to see what has happened to other nodes. These are called backward propagation techniques and are the most widely used learning algorithms (Haykin, 1999). This technique requires training data and the network learns in a supervised manner. Additionally, most neural networks used today are feed forward networks, which mean that there are no interconnections between the output of a processing element and the input of a node in the same layer or in a preceding layer. Essentially, the calculations always go forward. Feed forward provides faster calculations in determining the weights and was the technique used with the software in this research.

Research Literature Specific to Neural Networks and Credit Applications

Neural networks are powerful forecasting tools that can be trained to map past and future values of time series data and thereby extract hidden structures and relationships that govern the data. They have been used for analyzing relations among economic and financial phenomena, forecasting, data filtration, generating time-series, and optimization (Hawley, Johnson, & Raina, 1990; White, 1988; Terna, 1997; Cogger, Koch, & Lander. 1997; Cheh, Weinberg, & Yook, 1999; Cooper, 1999; Hu & Tsoukalas, 1999; Moshiri, Cameron, & Scuse, 1999; Shtub & Versano, 1999; Garcia & Gencay, 2000; and Hamm & Brorsen, 2000.) Hsieh (1993) stated that the following potential corporate finance applications could be significantly improved with the adaptation to neural network technology:

1. Financial Simulation.
2. Predicting Investor's Behavior.
3. Evaluation.
4. Credit Approval.
5. Security and/or Asset Portfolio Management.
6. Pricing Initial Public Offerings.
7. Determining Optimal Capital Structure.

Trippi and Turban (1996) noted in the preface to their book that financial organizations are now second only to the U.S. Department of Defense in the sponsorship of research in neural network applications. Most of the major investment banks, such as Goldman Sachs and Morgan Stanley, have dedicated departments to the implementation of neural networks in analyzing financial and credit data.

There can be little doubt that the greatest challenge facing managers and researchers in the field of finance is the presence of uncertainty. Indeed risk, which arises from uncertainty, is fundamental to modern finance theory and, since its emergence as a separate discipline, much of the intellectual resources of the field have been devoted to risk analysis. The presence of risk, however, not only complicates financial decision-making, it creates opportunities for reward for those who can analyze and manage risk effectively. Dealing with uncertainty in finance primarily involves recognition of patterns in data and using these patterns to predict future events. Neural networks handle these problems better than other statistical techniques because they deal well with large noisy data sets, particularly where the relationship between variables is unknown. In traditional statistical analysis, the user is required to specify the precise relationship between inputs and outputs and any restrictions that may be implied by theory. Neural networks differ from conventional statistical techniques in that the analyst is not required to specify the nature of the relationships involved; the analyst simply identifies the inputs and the outputs. According to Sarle (1994), no knowledge of neural network training methods or statistics is required for successful use.

A growing body of literature is based on the comparison of neural network computing to traditional statistical methods of analysis. Hertz, Krogh, and Palmer (1991) offer a comprehensive view of neural networks and issues of their comparison to statistics. Hinton (1992) investigates the statistical aspects of neural networks. Weiss and Kulikowski (1991) offer an account of the classification methods of many different neural and statistical models. Yoon and Swales (1997) compare neural networks to discriminant analysis with respect to prediction of stock price performance and find that the neural network is superior to discriminant analysis in its predictions. Surkan and Singleton (1990) find that neural network models perform better than discriminant analysis in predicting future assignments of risk ratings to bonds. Trippi and DeSieno (1992) apply a neural network system to model the trading of Standard and Poor 500 index futures. They find that the neural network system outperforms passive investment in the index. Based on the empirical results, they favor the implementation of neural network systems into the mainstream of financial decision-making. According to Zahedi (1993), expert systems and neural networks offer qualitative methods for business and economic systems that traditional quantitative tools in statistics and econometrics cannot quantify due to the complexity in translating the systems into precise mathematical functions. Singleton and Surkan (1995) compared a neural network model with multiple discriminant analysis (MDA) and demonstrated that neural networks achieved better performance in predicting direction of a bond rating than discriminant analysis could. Kim (1993) compared the neural network approach with linear regression, discriminant analysis, logistic analysis, and a rule-based system for bond rating. He found that neural networks achieved better performance than other methods in terms of classification accuracy.

Other studies have reported inferior performance of neural networks compared to other models or found no significant advantage in credit related applications over traditional statistical analysis. Galindo and Tamayo (1997), in their empirical study, examined four different techniques: classification and regression trees (CART), neural network models, k-nearest neighbor, and the probi statistical method. Neural network models came second after CART in their experimental results. However, the difference in performance between them was small. Desai, Convay, Crook, and Overstree (1997) analyzed the work of Galindo and Tamayo and concluded that the neural network involved did not significantly outperform the conventional techniques in this case because the most appropriate variants of the techniques were not used. Yobas, Crook, and Ross (1997) came to a similar conclusion with respect to credit card applications. While empirical studies show that neural networks produce better results for many problems, results are not always uniformly superior (Quinlan, 1993; Altman, Marco, & Varetto, 1994; Boritz & Kennedy, 1995; Boritz, Kennedy, & Albuquerque, 1995). Although these studies suggest that neural networks may not always be the best possible tool for all credit related evaluations, they also reveal that it has never been more than marginally outperformed by other methods. As Wray, Palmer, and Bejou (1994) mention, the advantages of neural networks over statistical models are (1) neural networks requires no predefined knowledge of underlying relationships between input and output variables; (2) neural networks' associative ability make them robust enough to tolerate missing and inaccurate data; and (3) neural networks' performance does not diminish with multi-collinearity problems, violations of set assumptions, high influence points, and transformation problems encountered in regression analysis. In addition, according to Granger (1991) non-linear relationships in financial and economic data are more likely to occur than linear relationships. The non-linear properties of financial data provide many difficulties for traditional methods of analysis (or may make the use of these traditional techniques impossible) and a number of authors (Ormerod, Taylor, & Walker, 1991; Grudnitski & Osburn, 1993; Altman, Marco, & Varetto, 1994; Kaastra & Boyd, 1995; Witkowska, 1995) have examined this.

Widrow, Rumelhart, and Lehr (1993) demonstrate that most neural network applications fall into three categories:

1. Classification.
2. Time Series.
3. Optimization.

Classification problems involve either binary decisions or multiple-class identification in which observations are separated into categories according to specified characteristics. They typically use cross sectional data. Solving these problems entails "learning" patterns in a data set and constructing a model that can recognize these patterns. Commercial neural network applications of this nature include:

1. Credit card fraud detection (Bylinsky, 1993).

2. Optical character recognition (OCR) (Widrow, Rumelhart, & Lehr, 1994).

3. Cursive handwriting recognition (Bylinsky, 1993).

4. Cervical (Papanicolaou or 'Pap') smear screening (Schwartz, 1995; Boon & Kok, 1995).

5. Petroleum exploration to determine underground oil deposits (Widrow et al., 1993).

In time-series problems, the neural network is required to build a forecasting model from the historical data set to predict future data points. Consequently, they require relatively sophisticated neural network techniques since the sequence of the input data in this type of problem is important in determining the relationship of one pattern of data to the next.

Examples of time series problems include:

1. Chinese writing recognition (Hitheesing, 1996).

2. Foreign exchange trading systems (Penrose, 1993).

3. Portfolio selection and management (Bylinsky, 1993; Elgin, 1994).

4. Forecasting weather patterns (Takita, 1995).

5. Speech recognition (Nelson & Illingworth 1991; Illingworth, 1991).

6. Predicting heart attack, from electrocardiogram (ECG) (Bortolan & Willems, 1993; Baxt & Skora, 1996). Baxt and Skora (1996) reported in their study that the physicians had a diagnostic sensitivity and specificity for myocardial infarction of 73.3 and 81.1% respectively, while the neural network had a diagnostic sensitivity and

specificity of 96.0% and 96.0% respectively.

Optimization problems involve finding solutions for a set of very difficult problems known as Non-Polynomial (NP)-complete problems. Examples of problems of this type include the traveling salesman problem, job scheduling in manufacturing, and efficient routing problems involving vehicles or telecommunication. The neural networks used to solve such problems are conceptually different from the previous two categories (classification and time-series) in that they require unsupervised networks, whereby the neural network is not provided with any prior solutions and thus has to "learn" by itself without the benefit of known patterns. The intent is to discover the natural groupings of items or variables and search for good but not necessarily the best groupings. They are widely used in understanding the complex nature of multivariate relationships (Johnson & Wichern, 1988).

The research conducted is an example of the use of neural networks to solve a financial credit related classification problem. The specific classification problem was to discover the non-obvious relationships in the data about an applicant for apartment rental that influenced the decision to extend credit through the rent/not rent decision. Generally, there is widespread recognition that the capability of humans to judge the worthiness of a credit application is poor (Glorfeld, 1996). Some of the reasons are: a) a large gray area where the decision is up to the officers, b) humans are prone to bias and errors as a result of this bias, and c) it is likely that there is important knowledge hidden in the data which may be useful for assisting the decision making process. Unfortunately, the task of discovering useful relationships or patterns from data is difficult for humans because of the large volume of data to be examined in a reasonable time (Handzic, 2001). Neural networks are an example of knowledge discovery tools for discovering the non-obvious relationships in data, while ensuring those relationships discovered would generalize to the new/future data (Bigus, 1996; Marakas, 1999).

U.S. Information Policy Considerations and Impact of Consumer Privacy Concerns on this Research

Background

The current U.S. credit reporting system relies on routine collection and dissemination of consumer information to credit agencies and to financial and business institutions. The fact that collection is routine across society makes the information complete, and because it is complete, the information is likely to be reliable and accurate. If a consumer buys a car after spending only 30 minutes with the dealer's credit manager, becomes eligible for a credit card by signing a one-page form, or receives a department store one-day discount and credit card for opening an account at the store's cash register, that consumer has been a beneficiary of the credit reporting system (Soman, 2002). The credit reporting system has become such an important part of commerce in the U.S. that most consumers take this system for granted (Wallison, 2001)

Easy access to credit files has rewarded the consumer with convenience and the inexpensive availability of credit for all segments of the population as the following examples show (U.S. Chamber of Commerce, 2002)

1. Between 1970 and 2001, the overall share of families with general-purpose credit cards increased from 16 to 73 percent.

2. The percentage of households in the lowest income quintile with a credit card has increased from 2 percent in 1970 to 28 percent in 2001.

3. Increased use of credit scoring has reduced the consumer's price for credit, particularly credit card debt. The U.S. Chamber of Commerce estimates that increased competition has provided consumers with $30 billion per year savings on debt.

Congress has been concerned for many years with the need to balance the protection of consumer privacy with maintaining ready access to consumer private information and credit history information. In 1968, Congress began hearings to regulate the use of personal information in the analysis of personal credit. The result of these hearings was passage in 1970 of the Fair Credit Reporting Act (FCRA), which was the first and continues as the major governing privacy legislation. The purpose of the FCRA is to ensure accuracy and security of the information contained in credit reports. The philosophy was to establish reasonable procedures for meeting the needs of commerce in a manner that is fair and equitable to consumers with regard to the confidentiality, accuracy, relevancy, and proper utilization of information. This legislation imposed obligations on just two distinct classes of companies involved in consumer credit: the credit reporting agencies (currently Equifax, Experian, and Trans Union) and the users of consumer reports.

With the increase of the Internet as a form of commerce, there have been a number of updates to the FCRA, most recently the 1996 amendments passed by the 104th Congress and the 1999 Gramm-Leach-Bliley Act. The legislative direction is to increase privacy of information whenever possible, without affecting the dissemination of this information for limited and legitimate purposes. The U.S. government has tended to stay close to its free market philosophy in its approach to consumer information policy and has adopted a minimalist approach (Internet Policy Institute, 2000).

Government studies in the United States and abroad have recognized certain core principles of fair information practice (Landesberg, Levin, Curtin, and Lev, 1998). These principles are widely accepted as essential to ensuring that the collection, use, and dissemination of personal information are conducted fairly and in a manner consistent with consumer privacy interests. These core principles require:

1. that consumers be given *notice* of an entity's information practices.

2. that consumers be given *choice* with respect to the use and dissemination of information collected from or about them.

3. that consumers be given *access* to information about them collected and stored by an entity.

4. that the data collector takes appropriate steps to ensure the *security* and integrity of any information collected.

Moreover, it is widely recognized that fair information practice codes or guidelines should contain enforcement mechanisms to ensure compliance with these core principles. While most countries agree on the general objectives, policies to govern information privacy vary widely around the world.

Status of privacy policy legislation

The recent debate over privacy, and the role of law in protecting it, is unlike many other political debates for a variety of reasons. Privacy is an unusually broad term, encompassing both fundamental constitutional rights (such as freedom from government intrusions into homes and other forms of search and seizure, as well as the right of citizens to make decisions about marriage, health, contraception, and so forth) and less well-defined and arguably less critical issues (such as the desire to be free from direct marketing calls and mailings). Privacy is important for all individuals in a wide variety of settings because it involves restrictions on the information flows that are essential to consumer products and services, commerce, and government. The debate over how to protect privacy affects all citizens, consumers, most businesses, government agencies, and other institutions.

In practical terms, the U.S. government has tended to stay close to its free market philosophy in its approach to consumer privacy and information policy and has adopted a minimalist approach. Unfortunately, this legislative philosophy and the resulting information policy appear to be ineffective in controlling the spread of unauthorized uses of consumer information. Despite 34 years of enforcement of the FCRA as the primary information policy in the United States for protection of consumer information, identity theft is a growing problem. Identity theft accounted for 39% of the 635,000 total complaints received by the Federal Trade Commission in 2004 and is increasing (identity theft complaints were 161,000 in 2002, 215,000 in 2003 and 247,000 in 2004) (FTC, 2005). Many think that this is just a small fraction of the total number of actual victims with Synovate Research estimating that 9.9 million people per year are victims of identity theft (3.3 million through opening new fraudulent accounts and 6.6 million through fraudulent use of existing accounts) (Synovate Research, 2003). In 2002, Star Systems conducted a telephone survey they believe indicates that as many as 1 in 20 adults, or 11.8 million Americans, are victims of identity theft (Star Systems, 2002). According to a

May 2000 survey by CalPIRG and the Privacy Rights Clearinghouse, the average consumer victim spends 175 hours and $800 resolving identity theft problems, and it takes two to four years for victims to clear up all the resulting problems (Gayer, 2003). The Synovate Research report indicated that the average business loss of a single identity theft problem is $4,800 per victim.

The current system for gathering and disseminating private consumer information appears to “leak” private information routinely to those persons who should not have it. The existing pro-business focus of FCRA does not appear to offer effective incentives for business to control these information leaks. Under the 2003 amendments to the Fair Credit Reporting Act section 609(e), identity theft victims are entitled to get from businesses a copy of the application or other business transaction records relating to their identity theft free of charge. Businesses must provide these records within 30 days of receipt of the victim’s request and must provide these records to any law enforcement agency that the victim authorizes. However, this FCRA provision does not require a business to change its current information or record retention procedures. A business may even decline to provide the records if, in good faith, it determines that this FCRA provision does not require disclosure, the business entity does not have a high degree of confidence in knowing the true identity of the requester after reviewing the proof of identity provided by the requester, the requester has made a misrepresentation of fact relevant to the request, or the information requested is Internet navigational data or similar information about a person’s visit to a Website or online service. The burden is on the victim to prove that they need the information.

However, the Identity Theft and Assumption Deterrence Act, enacted by Congress in October 1998 (codified, in part, at 18 U.S.C. 1028(a) (7)) makes identity theft a federal crime. The Act makes it a federal crime when someone knowingly transfers or uses, without lawful authority, a means of identification of another person with the intent to commit, or to aid or abet, any unlawful activity that constitutes a violation of federal law, or that constitutes a felony under any applicable state or local law. Under the Act, a name or SSN is considered a "means of identification" and so is a credit card number, cellular telephone number, electronic serial number, or any other piece of information that may be used alone or in conjunction with other information to identify a specific individual. Violations of the Act are investigated by federal investigative agencies such as the U.S. Secret Service, the FBI, and the U.S. Postal Inspection Service, and prosecuted by the Department of Justice. In most instances, a conviction for identity theft carries a maximum penalty of 15 years imprisonment, a fine, and forfeiture of any personal property used or intended to be used to commit the crime. This act however makes no mention of the sources that provided the information used in the identity theft and provides no penalties for these sources of information.

Changing the information policy from the apparent pro-business focus to a pro-consumer focus, or identifying other workable solutions, is a complex problem for a number of reasons.

1. Easy access to credit is a bedrock principle of the U.S. economy. For business, consumer information is a valuable commodity that helps shape new products and reaches new potential customers. Consumer spending has been the one bright spot in an

otherwise sluggish economy, with consumer spending representing over two-thirds of the gross domestic product (GDP) of the U.S. in a typical year (Auten, 2000).

2. Powerful forces and persons hold strong opinions that make change difficult. The former Chairman of the Federal Reserve System, Alan Greenspan, made the following points before the U.S. House of Representatives Financial Services Committee, on April 30, 2003 in a hearing on U.S. monetary and public policy:

a. The complexity and sophistication of modern credit markets make it impossible for individual lenders to evaluate individual borrowers efficiently based on personal knowledge.

b. It is in consumer's interest to have consumer information and credit information freely flowing in order to reduce uncertainty and keep interest rates low.

c. Without the ability to rely on continuously updated credit evaluation systems based on shared information, it will be difficult to maintain current levels of credit availability.

Thomas Chapman, (2004, p. 3) President of Equifax, the second largest credit bureau, said in a speech that "further tightening of the FCRA would negatively impact his company's ability to disseminate credit information." Assistant Treasury Secretary Wayne Abernathy (2004, p.10) has noted, "The sharing of information, within secure parameters reinforced by uniform national standards, has increased the access of more consumers to a wider variety of financial services, at lower costs, than ever before."

3. States are prevented from enacting their own legislation. The FCRA defines a national credit system and prevents states from enacting their own legislation to tighten

consumer information policy laws. At least 39 states now have laws addressing identity theft, and these laws address the penalties for committing the theft (see Appendix B for recent laws enacted at the state level). None of these laws can address the issues of the "leaky" information sources used in the crime.

As states expand their own laws, business is no longer able to look to a single national standard regarding the handling and protection of consumer information. This produces the possibility of many costly compliance obligations for businesses in multiple states. While this topic is being currently discussed at the federal level, no clear direction has yet been established.

In spite of the uncertainty concerning a state's role, California has tightened its current state privacy laws and put some of the focus on business to protect consumer privacy. Two new laws that went into effect July 1, 2004 give Californians more information and, in some cases, more choices on how businesses use personal information. The California Online Privacy Protection Act requires a privacy policy to be posted on all commercial Web sites that collect personal information on California consumers. It also requires operators of commercial Web sites to comply with their posted policies. In other words, Web sites must say what they do and do what they say with Californians' personal information. The California Financial Information Privacy Act gives Californians more say in how their personal financial information is used. The law, which applies to banks, insurance companies, securities firms, and other financial service companies doing business in California, provides more consumer control than federal law. It also requires an easy-to-read, plain-language privacy notice. The Attorney General and state agencies that regulate financial institutions enforce the Financial

Information Privacy Act. A significant difference with this Act in comparison to other similar legislation is that penalties exist for business errors that include up to $2,500 per violation, with a maximum of $500,000, for negligent disclosure or sharing of nonpublic personal information. The penalty is also $2,500 per violation, with no maximum, for knowingly and willfully obtaining, disclosing, sharing, or using nonpublic personal information in violation of the statute. Penalties for the business are doubled if violation results in identity theft.

In general, federal information policy towards consumer information as defined by the FCRA and other legislation,

1. makes consumer credit widely and easily available.
2. assumes that business is handling consumer information properly and securely.
3. assumes that misuse of the information will be manageable and an issue that the nation can "live with."
4. assumes that legislation toward information policy can address specific problem "hot spots" as the need arises.

This federal information policy has not adequately protected consumers in the eyes of state government. As a result, state governments have enacted a long list of privacy laws in order to try to fill the void (Appendix B), with the California laws being the most recent and toughest. None of these patchworks of legislation fully addresses the problem because none appears to understand fully the consumer information collection system.

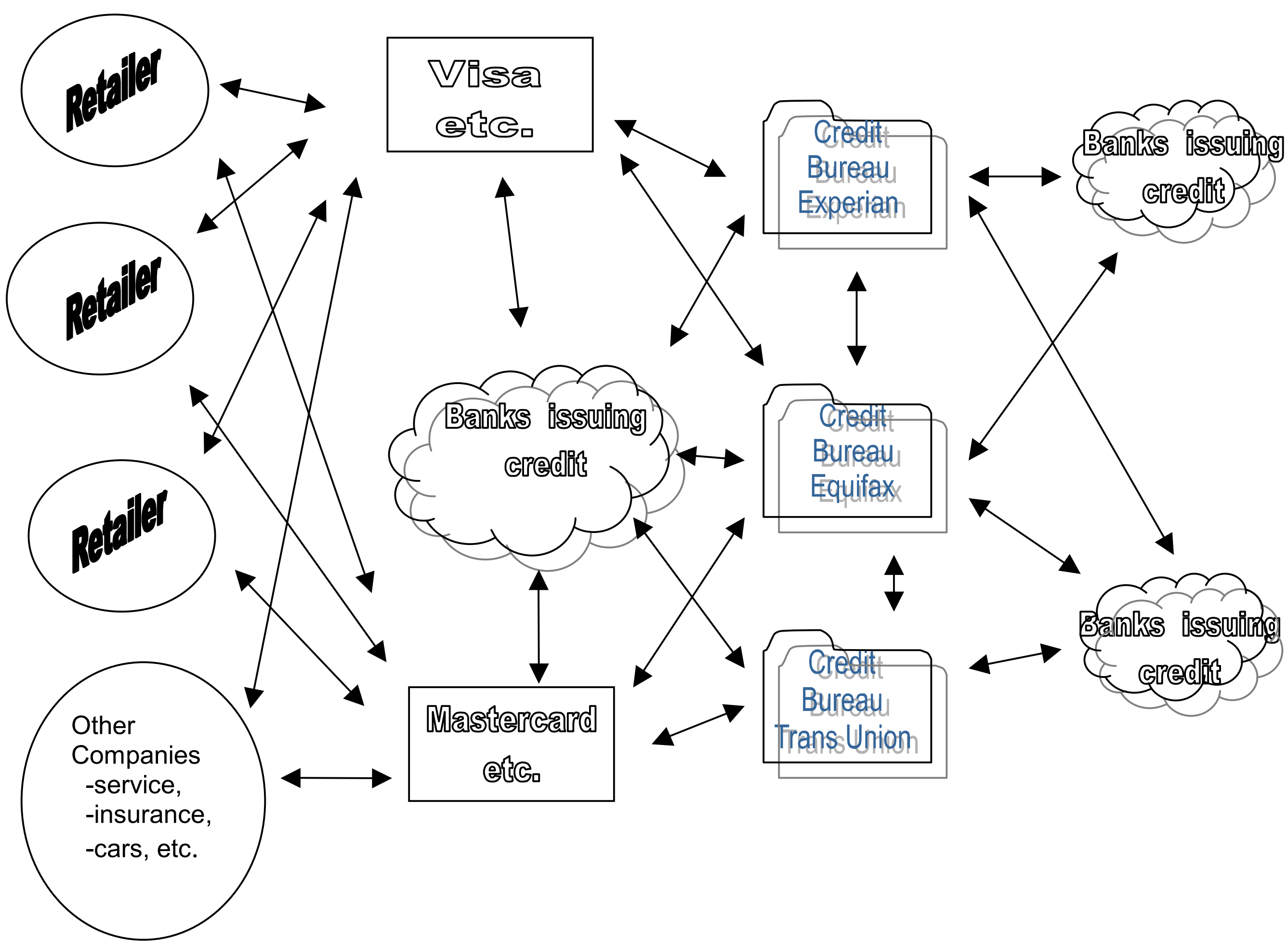
Historically, credit providers have faced three problems. First, they lacked inexpensive access to sufficient information about a potential borrower and the risk associated with that borrower. Second, they were often unable to sanction effectively those who violated their promise to repay a loan and to alert other credit providers to the risk. Third, they were unable to price loans via the interest rate, to reflect the degree of credibility of a borrower's promise to repay. To solve the first two problems of access to information and sharing of information, merchants banded together and began to form local repositories of credit information beginning in about the early 1900s. These tended to be local repositories that usually focused on one type of business (such as local banks pooling their information together or local retailers pooling their information) and often maintained unreliable or incomplete information (Furletti, 2002). This local approach and incomplete reporting complicated the ability of credit providers to charge higher rates to those with poor credit and provide better rates for better credit, especially as Americans became more mobile. Lenders needed this information because unlike collateralized loans, the promise to pay for most credit transactions is not backed by a particular asset. During the 1960s and 1970s, national repositories of information began to form from this local clutter with three emerging and operating today: Equifax, TransUnion, and Experian. These three credit bureaus continue to collect information from credit grantors and other companies, manage the data into individual credit files for each consumer in the U.S., and provide an overall system of assessing credit worthiness. The positive impact of this national system on the previously mentioned third problem of correctly setting the interest rate to reflect the risk of a borrower has been dramatic. Table 4 shows that interest rates today have become more widely dispersed with rates lower overall than they were in 1990. This is a direct result of risk pricing based on credit report data (Barron & Statton, 2003). This can be described as the evolution of a system of reducing the risk premium caused by hidden or unknown information.

Table 4 Percentage of Total U.S. Loan Balances Distributed by Interest Rate

Year	less than 5.5%	5.5 to 10.99%	11 to 16.49%	16.5 to 17.99%	over 18%
1990	0%	3%	3%	20%	73%
2002	15%	31%	25%	3%	26%

Note: Interest Rate on Loans (non real estate & auto)

The credit reporting system that collects the consumer privacy information that makes this possible is far-flung, loosely organized, and voluntary. The credit bureaus that hold national data enforce a “give-to-get” policy with a purpose of collecting as much consumer information from as many sources as possible. Essentially the “give-to-get” policy requires that a lender must supply weekly or monthly information about a customer to the credit bureau in order to use the credit bureaus files. In practice, this is usually only enforced with the larger users. However, as illustrated in Figure 9 (developed by researcher), there is a significant amount of consumer information that flows around the financial system and it is this wide-ranging flow that is the inherent source of the system’s strength and the inherent source of the “leakiness” related to consumer privacy.



Source: developed by researcher

Figure 9 Typical consumer information flows.

As an example, a consumer would typically use a MasterCard at a retailer to purchase an item. A record of this transaction and the consumer information is recorded with the retailer (for accounting purposes), also recorded with the bank that issued the MasterCard and is extending the credit, and with MasterCard central, which is acting as the clearinghouse between the retailer and the bank. If the transaction was the purchase of auto insurance then additional records are stored at the auto insurance company and at auto accident tracking companies. At some point, one or all of these players will send this consumer information to one or more of the credit bureaus. Since the system is voluntary, not all three credit bureaus are necessarily updated on every transaction. The recent California law attempts to protect consumer privacy by making the business responsible for an information leak also responsible financially for the error. In this example, if only the retailer is physically located in California, then only the retailer would be liable to California law for an information leak, although many others have the sensitive consumer data and could have been the source of the leak. It is easy to see that state level laws are probably not a reasonable approach to the information policy issues surrounding consumer information protection and privacy and probably will have limited impact on identity theft. This is a national problem requiring a federal government solution. State laws are, at best, only a stopgap measure (but good public relations and perhaps good politics).

Federal Information Policy Considerations for the Next Several Years

The past five years have witnessed an explosion in legislation, regulation, and litigation designed to protect the privacy of personal information. Congress alone has adopted comprehensive federal financial privacy legislation, online privacy protection for children, and the first federal prohibition on access to open public records without individual "opt-in" consent, among other privacy laws. These federal level laws have tended not to focus on preventing harmful uses of personal information or invasions of privacy by the government, but instead these laws have tried to grant individuals broad rights to control information about them that is used by the private sector. At the state level, legislators have considered hundreds of their own privacy bills in the past two years alone. State attorneys general have initiated aggressive privacy investigations and litigation. Outside of the United States, Europe has brought its sweeping data protection directive into force, while other industrialized countries either have adopted or are in the process of considering new privacy laws (Cate, 2002). While these foreign privacy policies are less desirable when applied in the U.S., in sum, there seems no shortage of sources to look for experience in enacting and enforcing federal privacy laws that improve the current situation.

The result so far has been a transformation of privacy law. Historically, U.S. privacy law has focused on two broad themes. The first and most visible was preventing intrusion by the government. Virtually all constitutional privacy rights reflect the reality that only the government exercises the power to compel disclosure of information and to impose civil and criminal penalties for noncompliance, and only the government collects and uses information free from market competition and consumer preferences. The second theme reflected in U.S. privacy law was preventing uses of information that *harm* consumers. When privacy laws addressed private-sector behavior, they were designed to prevent only specific, identified harms. For example, the Fair Credit Reporting Act, one of the earliest privacy laws applicable to the private sector, focuses primarily on correcting inaccuracies and assuring that credit information is not used in ways likely to harm consumers (Cate, 2002).

Increasingly, however, the dominant trend in recent and pending privacy legislation is to vest in consumers control over information in the marketplace—irrespective of whether the information is, or could be, used to cause harm. Alan Westin (1967, p. 7) describes this as “the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.” It appears that privacy is “an issue that will not go away until every single American has the right to control how their personal information is or isn’t used” (LaFalce, 2000, p. 4). This trend toward better control is reflected in the recently passed federal bill entitled the Fair and Accurate Credit Transactions Act (FACTA) of 2003. This federal law covers companies that hold consumer information or provides consumer information for business purposes such as determining a consumer’s eligibility for insurance or employment as well as credit. This bill is attempting to provide more safeguards and make consumers aware of the multiple places and systems where information is maintained:

1. Uniform credit standards: In 1996, Congress set uniform national standards on credit reporting. These standards set clear rules on what credit agencies could include in consumer credit reports. The new law made these standards permanent.

2. Safeguarding receipts: To help ward off identity theft, retailers must hide credit card and debit card information on customer receipts. Only the last five digits of a card number will be listed. As of January 1, 2005, all new cash registers and point-of-sale

terminals must print these safeguarded receipts. Merchants have until December 4, 2006, to phase out any existing registers or terminals that print full account numbers on receipts.

3. New opt-out rules: Consumers will have the right to "opt-out" and block solicitations from affiliates of companies with which they do business.

4. Disclosing bad credit news: A bank must notify a consumer if it reports any negative information to the credit bureaus. A bank will also have to alert the consumer if it grants credit at less favorable terms than those received by most other consumers. The actual details of this provision are still being negotiated (i.e. what is negative information: one late payment or two late payments and so forth).

5. Reporting of false credit news: Any debt collector that learns that information on a consumer's credit report is fraudulent must inform the creditor that the information is false. No retailer or creditor may report credit information to credit bureaus that is known or believed to stem from fraud.

6. More power for identity-theft victims: Identity-theft victims who file police reports will be able to block fraudulent information from appearing on their credit reports. In addition, fraud victims will also get more help from businesses in tracking down impostors. Under the new law, an identity-theft victim will be able to obtain copies of business records that list fraudulent transactions carried out by an identity thief.

7. Beefed-up fraud alerts: Consumers now have the right to place a fraud alert on the credit report. A fraud alert is a statement to alert creditors that private financial

information has been or may be compromised. Identity-theft victims would put fraud alerts on credit files to stop impostors from opening new accounts. Under the new law, once a credit bureau receives a fraud alert, it must take steps to ensure that the consumer and not the thief will be granted credit in the future. This extra step could be something as simple as calling the phone number listed in a consumer fraud alert whenever a new application for credit pops up.

8. Special alerts for the military: Americans in the armed forces will be able to place special alerts in their credit files while they are serving overseas to help minimize their chances of becoming victims of identity theft.

In order to allow consumers the ability to manage the information that is held, this bill grants free access to consumers concerning the reports in the following areas: (Weston, 2004)

1. Medical information. If a consumer has applied for life, health, disability, or long-term care policies, information about the consumer's health will usually have been reported to the Medical Information Bureau. This membership association of 600 companies is designed to help insurers detect fraud and deter applicants from lying on applications. This association is not affiliated with any healthcare organization.
2. Tenant history. No one company dominates this field, but some of the larger screening agencies include Registry Safe-Rent and U.D. Registry. These companies maintain information on consumer past living arrangements and rentals.
3. Auto and homeowners insurance claims. ChoicePoint's CLUE database and ISO A-PLUS database contains a record of every auto accident and traffic violation by consumer. These databases are used for auto insurance purposes. Additionally these

companies maintain records of claims against property insurance by consumer and by residence.

4. Check-writing history. ChexSystems is the largest player in this arena and maintains a database of consumers who have "mishandled" their bank accounts (typically by repeatedly bouncing checks).

5. Employment screeners. Companies that provide background checks to employers have to abide by other FACTA rules. They typically are not required to provide free reports to consumers because the typical background-checking firm does not maintain "permanent" files on consumers and instead puts together a one-time report for employers. Only companies that maintain databases of information on consumers must provide free reports. However, employers must get the applicant's written permission before a third party can run a background check. The consumer is not entitled to see the report unless the report is used to deny a job or promotion.

This bill is a continuation of the U.S. government's tendency to stay close to its free market philosophy in its approach to consumer privacy and information policy. Essentially the FACTA bill makes it the consumer's responsibility to check the information held by others; using all the free reports, consumers are to manage the accuracy of their own information. While this bill is a major step forward, it is in sharp contrast to the recent California information policy bills that begin making the businesses that hold the information responsible for the information, particularly when private consumer information leaks out. Neither approach alone is sufficient, although both approaches, if combined, would be an excellent start toward defining a workable and effective federal information policy since only consumers can verify the accuracy of information and only business can adequately safeguard the privacy of the information. As with most policy solutions in a democracy, the workable strategy is many small steps in many directions but all moving toward the information policy goal of a) greater consumer control of information b) supplemented with greater business responsibility toward protecting the privacy of that information.

Impact of Consumer Privacy Concerns on this Research

Privacy and protection of consumer information presents many complex issues to which there are no easy solutions. This is especially true in the U.S. where the availability and control of information inevitably and directly affects the efficiency, cost, and quality of the economic system. The important, but modest, steps taken so far by the federal and state governments demonstrate that the legislatures are beginning to understand the problem and issues. The federal government can continue to rely on its philosophy of using market forces, but, unfortunately, security rarely improves as a result of time, good intentions, or market forces. Laws must be enforced by penalties subsequent to and conditioned upon their violation because unless accompanied by some penalty for its violation, no act of a legislative body or sovereign prince can truly be considered a law (Mason & Lalor, 1877). Effective law enforcement needs to do three things in order to be effective: deter, capture, and prosecute malicious actors (Saloma, 1984). By establishing national regulations, enacting national penalties, and empowering and funding law enforcement, the federal government can provide direction to the courts and provide the incentive to business to protect privacy more aggressively than has been done to date.

The current privacy legislation guarantees a consumer's right to scrutinize their credit report but only covers the information actually held in the credit report file. When credit reporting bureaus develop a mathematical scoring model based on the credit file or sub-contract with a mathematical modeler (such as FICO and others) to obtain a credit score, the consumer protections related to the credit file do not apply. While a consumer may see the information in their credit report, a lender or credit user is not required to provide consumers with their credit score, nor with the calculations that led to that score (an exception is made if the score caused a consumer to be rejected for credit) (Cannon, 2000). Although the courts make decisions in this area very slowly, they have tended over the years to uphold this rather narrow definition and application of the privacy laws (Scranton, 2001). In addition, the mathematical algorithm that calculates the score is a competitive secret that may have been developed at great cost. The vendors of credit scores are reluctant, therefore, to reveal the internal mathematical makeup of their scoring algorithms. However, in 2001, under pressure from consumer groups and some of its customers, Fair, Issac & Company agreed to sell FICO scores directly to the public for $12.95 each. The credit bureaus are now also selling consumer scores on their Web sites (Kadet, 2003)

Technological advances, like computer credit scoring, have the potential to either support or erode society's values. While the future depends, to some measure, on technological capabilities, it depends even more on how technology is applied. As paper-based processes give way to IT-based processes, the fundamental challenges remain the same: how to promote values that sometimes are in conflict. The significant difference today is that enormous volumes of information can be collected, stored, used, combined, and shared instantly over long distances. While this new information capability can be used for dramatically more efficient, convenient, and sometimes life saving services, it can also be used in ways that challenge traditional assumptions about how to assess and balance different interest and values.

For example, such challenges are evident in the healthcare industry. New technologies can give health care workers timely access to patient records to improve service and possibly save lives. Hospitals and insurance companies can also use these records to speed treatment and process claims more efficiently. On the other hand, these same technologies can give employers inappropriate access to health records of prospective employees, or give marketers lists of potential customers. Electronic records are also vulnerable to destruction and misuse both inside and outside the healthcare industry.

In designing information systems for healthcare services, special care must be taken to balance the values and interests of various stakeholders. In some cases, privacy and security are clearly at odds. To provide a higher level of security, individual identities are authenticated, confirming, in advance, that these individuals are authorized to access records, and hold these individuals accountable. Unfortunately, these actions to protect security reduce the scope of anonymity that has traditionally been an important natural protector of privacy. As Justice Louis D. Brandeis of the U.S. Supreme Court stated in the dissenting opinion in the 1928 *Olmstead v. U.S.* case: “The makers of our Constitution…. conferred, as against the government, the right to be let alone -the most comprehensive of rights and the right most valued by civilized men.” Unfortunately, it is no longer possible to operate in an electronic world and be completely “let alone.”

Privacy and freedom of information have emerged as two of the most difficult information technology issues for several reasons:

1. Designers of systems have been focused on making systems efficient with high customer satisfaction and a free flow of information. Issues such as privacy and security have historically tended to be a secondary concern.

2. Stakeholders in the debate tend toward strongly held polarized positions making compromise difficult. Thomas Chapman's (the President of Equifax the second largest credit bureau) previously mentioned comments that "further tightening of (privacy legislation)…. would negatively impact his company's ability to disseminate credit information" (Chapman, 2004, p.3) is not a position tending toward compromise.

3. In a networked world, many third parties, both known and unknown, including telecommunications companies and public and private service delivery partners affect privacy and security.

Balancing the competing issues of privacy and freedom of information requires exactly that, a balance. Too narrow a focus on any one side or any one element is likely to lead to negative results, but so is sticking with the status quo. For example, service efficiency improves with information age healthcare. However if electronic services produce easy access to health information, patients may stop talking candidly to their doctors reducing available information and efficiency, and affecting privacy (CBC News, 2001). Good decisions will depend on good leadership in balancing the competing needs of the stakeholders.

Several suggestions can help accomplish this balancing act from an operational viewpoint (Mechling & Applegate, 2001).

1. Adopt existing standards where appropriate. The road to privacy with information availability, in many cases is well charted but not well traveled.

2. Educate and involve stakeholders early in the discussion.

3. Executive management, not IT, must be the creators of information policy.

4. Plan for privacy and security before collecting and using data. Retrofitting systems is expensive and difficult.

5. Consider IT an opportunity to enhance privacy not just maintain it. Aggressively develop new capabilities, and apply new technologies that enhance access and improve privacy, security and other values.

Maintaining the balance between information needs and privacy was an important concern as this research made use of credit report information, credit score information and other information for consumers who had applied for apartment rentals. In order to protect and address privacy concerns and provide the balance between information needs and privacy, no information on specific consumers has been provided in this final report. Specifically information was anonymous in two ways a) the name and location of the apartment complex was not revealed except to say that it is a mid-size apartment complex in a southeastern U.S. city and b) the information from the apartment complex had all references related to consumer names and social security numbers removed.

These measures safeguarded the privacy of the consumer information while not negatively affecting the research project. Furthermore, each applicant signed a statement during the application process, which gave the apartment complex permission to use credit data and other data as needed in making the apartment rental decision and this was essentially the consumer's "opt-in" permission.

Information Policy Summary

The needs of the government to maintain a safe society and battle terrorism has created a demand for the storage and easy access to large amounts of private data on U.S. individuals. This need is far surpassing the usual business needs for information on consumers, in both the amount of information and in the detail of the information collected and available. Conversely, at the same time, the rapid growth of business on the Internet, and the resultant expanding flow of consumer information, have created the potential for the widespread misuse of this information through identity theft. Maintaining a safe Internet therefore requires a more restrictive control of information. Balancing these competing demands for greater information availability with the need for greater information control is a legislative balancing act that the federal government and state governments are struggling to address. In practical terms, the U.S. government, to date, has tended to stay close to its free market philosophy in its approach to consumer privacy and information policy and has adopted a minimalist approach. Unfortunately, this legislative philosophy and the resulting information policy appear to be ineffective in controlling the spread of unauthorized uses of consumer information. It is likely that a far-reaching, complete, and clear policy toward privacy protection that relies heavily on regulations and bureaucracy would be naturally proposed and adopted. Unfortunately, the U.S. Department of Commerce has stated that a restrictive government policy toward information may be incompatible with the U.S. First Amendment and its specific limitations on the ability of government to control the free flow of information (Star Systems, 2001). Consequently, it appears likely that the U.S. is entering a time of rapidly changing and possibly conflicting information policies. This research recognized the need to protect consumer privacy and the data was blind without reference to individuals.

Summary of what is known and unknown

It is possible to draw the following conclusions from the literature review

1. Managing risk and uncertainty in financial transactions, such as granting of credit, is fundamental to modern financial theory. Dealing with uncertainty in finance primarily involves recognition of patterns in data and using patterns to predict future events.

2. Traditional statistical methods are available to manage data and identify patterns but these methods work most effectively when the dependent variable and independent variable occur in a linear relationship or operate in a known way. If the true relationship among variables is non-linear, then techniques, such as discriminant functions and logistic regressions, are inappropriate to develop knowledge from the data. In addition, these techniques ignore any possible interaction among variables in general.

3. Many authors think that non-linear relationships in financial and credit data are more likely to occur than linear relationships. As a result, advanced non-linear modeling techniques, such as expert systems and neural networks, are being applied to the finance and economics fields and written about in the literature with greater frequency. Many comparisons have been made between traditional statistical methods and neural network methods for solving the same problems.

4. While expert systems are a good approach to constructing a model, it is difficult to get the correct “knowledge base” and decide upon the relative importance of each rule.

Expert systems, as a tool, have thus declined in importance and are being replaced by neural networks to get the mapping between independent and dependent variables in non-linear problems.

5. Neural networks have been applied to a wide range of data types in the finance area and, in general, have had good predictive results. However, neural networks cannot explain the causal relationship among variables as related to the outcome (i.e. why did this variable input cause that outcome). This is a known problem inherent to the nature of neural networks. Experimentation is ongoing, using other techniques such as combinations of neural networks and fuzzy logic to attempt to overcome this.

6. Neural networks have been successfully applied to credit granting related problems, in general, and are beginning to be applied to more specific, individual problems. It is not known how size of the data set used to create the neural network for these individual problems affects accuracy of the model. Most of the models have focused on the use of financial data and financial ratios as the source of the data set. Some authors are beginning to experiment with the addition of qualitative variables related to management and other characteristics in evaluating the case. A tool capable of dealing with both quantitative and qualitative variables and their interrelations is needed (Khan, 2002).

7. Congress has been concerned for many years with the need to balance the protection of consumer privacy with maintaining ready access to consumer private information and credit history information. This research recognized the concerns of consumer privacy and complied with privacy legislation.

Contribution of this study

The apartment complex that was studied appeared to have a history of selecting tenants who did not satisfy the terms of their lease despite using credit scoring. This lack of predictability of the credit score used at this apartment complex is in sharp contrast to the apparent success in the banking industry and the auto insurance industry. This lack of predictability has forced the management to rely on other factors in making the accept/reject decision on each applicant, such as other financial ratios including the ratio of gross income earned to monthly rental amount, payment history at other landlords, and other non-financial issues such as size of family, reputation at other apartment complexes, management gut-feel, and so forth. While banks have a highly predictive set of credit scoring models to help with decision making, apartments do not.

This study analyzed the credit reports and credit scores of past applicants and compared these with the actual results of renting apartments to these applicants. This analysis was the basis for the identification of other variables and factors related to the applicant that appeared to be predictive of behavior. These variables and factors were used in the development of a new credit scoring type model. This research continued the latest trend in the literature as it examined credit scoring as applied to one industry, specifically the apartment rental industry. It furthered this trend by applying scoring to one segment of this industry, specifically the selection of applicants. Furthermore this research showed that general commercial based scoring models, currently in use or available for use, are not predictive in this industry.

Chapter 3

Methodology

Introduction

The first phase of this research analyzed the results of using six commercially available credit scores applied in one apartment complex to the task of selecting applicants. This analysis used linear regression and means testing using the t-test to determine the predictive accuracy of these models. This part of the analysis answered the research question: How effective are commercially available credit scores in predicting applicant financial behavior when renting an apartment?

Phase two of this research used neural networks to develop a new model using both credit data and other lifestyle data about the applicant. The hypothesis was that the addition of this lifestyle data would improve the predictive accuracy in selecting apartment rental applicants over currently available models based only on credit data. This part of the analysis answered the research question: How is the prediction accuracy of a new neural network based credit scoring model improved by adding lifestyle data to the credit report data?

Research methods employed

The analysis proceeded in two phases. Phase one analyzed historical data and compared the FICO credit score currently used by this apartment complex and five other commercially available scores with the actual results of renting apartments to these past applicants (The brand name of the score currently used is "FICO National Risk Score" and is number 2 on Table 1.) This verified the property manager's comments that the existing score was not predictive. The research also compared five other credit scores not currently used by the apartment complex (but suggested by the credit bureau Experian) with the actual results of renting apartments to these past applicants. This indicated whether other commercially available scores were more predictive than the credit score currently used by the apartment complex (Table 5).

Table 5 Possible Results of Phase One

Possible Results	**Possible Interpretation**	**Possible Effect on this Research**
None of the six scoring models (original plus five additional) tested are predictive	Commercially available models not tailored to apartment rental industry Credit scoring alone may not be predictive	Build new proposed model without existing credit scores as an input since none are predictive. (This was the eventual outcome.)
One or more of the six (original plus five additional) models tested are predictive	Model used by apartment complex should be changed to the scores that are predictive	Use result from most predictive model(s) as one of the input(s) in building the proposed new model
All of the six models (original plus five additional) tested are predictive	Model used by apartment complex should be changed to the scores that are most predictive	Test new model development with each predictive score individually and in combinations to try to improve accuracy of the new proposed model

The data was examined using standard regression analysis since the data should respond in a linear fashion as the credit score output is assumed to be a linear equation. Essentially as the applicant's credit score increases, implying better credit, the applicant's tendency to honor fully the lease to the end should also increase proportionally. The credit score was the independent variable and the length of lease was the dependent variable and the data was analyzed as a correlational study investigating the relationship between credit score (independent variable) and months of lease honored by the applicant (dependent variable).

Phase two used neural networks to analyze a combination of credit data and lifestyle data. Additional data on new applicants was collected and the apartment complex purchased additional data on each applicant such as driving records, vehicle ownership, and criminal background information to provide other variables for the new neural network model. A new scoring model was developed that combined credit data with the additional data obtained from the tenant applications and from the purchased data. These additional variables (Appendix C) were then simplified into 10 key variables (Appendix D) and the new model was developed using these 10 variables and neural network techniques. Apartment management suggested the following 10 variables as important in the decision process and these were used in the neural network model.

1. State of previous address.
2. Adult only, multiple adults, or adult with children.
3. Total applicant income.
4. Total *Blue Book* value of vehicles.
5. Number of driving infractions.

6. Applicant has criminal information.
7. Total loan balance.
8. Total monthly payments.
9. Total credit file inquiries.
10. Percent of satisfactory financial accounts.

Specific procedures employed

Phase One Overall

The same statistical analysis was completed on the National Risk Score currently used by the apartment complex and all of the five additional credit scores, that is:

1) a statistical regression analysis used all the data for each score in a single group. This regression should show a correlation between score and length of lease honored (i.e. high R Square factor) for each of the five scores analyzed.

2) a statistical regression analysis that examined the data subjects (applicants) when divided into two groups based on their fulfillment of the 12-month term of the lease agreement. Group one was those applicants that fulfilled the lease term and stayed for 12 months or longer and group two was those applicants who stayed less than 12 months. Group one represented the desirable applicants that the score should identify. With an accurate, predictive model, the hypothesis was that group one applicants would have a better mean credit score than group two applicants. If the mean scores of both groups (desirable applicants and less desirable applicants) were similar, this would imply little predictive value in using this scoring model. These two analyses were performed on each of the scores individually to determine the level of correlation and the predictive power.

3) to evaluate further, whether the difference between the means of the two groups is statistically significant, a t-test was run for each score. The t-test is the most commonly used method to evaluate statistically the difference in means between two groups (Hill & Lewicki, 2006).

Methodology for analysis of National Risk Score currently used by apartment complex

The study selected 50 applicants from those that applied for an apartment during the year 2000. Since this apartment complex leases to about 8 to 10 tenants per month, these 50 were a majority of the tenants over about a five-month period. These 50 were selected by the apartment complex management and the names and social security numbers of the applicants were removed.

Ideal Score

The scoring model used by this apartment complex is the National Risk Score provided by Experian (number 2 in Table 1), one of the three major credit bureaus in the U.S. This particular model creates a number score that directly corresponds to risk. Specifically, a score of 100 indicates that this applicant has a 10% probability that they will NOT fulfill their financial obligations. A score of 525 would indicate that this applicant has a 52.5% probability that they will NOT fulfill their financial obligations (essentially the higher the number for the National Risk Score, the higher the risk). This is opposite to the typical credit score that has a scale calibrated so that as the score number gets higher, the risk gets lower. The apartment management believes that scores of less than 200 for the National Risk Score model represent applicants that pose a reasonable business risk and should be the ideal candidate. However, applicants are routinely accepted with scores outside this ideal range based on other items in the credit file or on the application. These other items cause management to ignore the score. Examples include extraordinary medical expenses that are unpaid, bankruptcy due to a divorce, or credit problems due to loss of employment that is now corrected. Essentially, management was using its "gut feel" in selecting from this pool of applicants and these exceptions produced a data set that covered a broad cross section of applicants.

Ideal Tenant

A number of descriptors that describe an ideal tenant. Examples include honoring the 12-month term of the lease, paying rent on time, and social and living habits (i.e. problem neighbor?). Management's opinion is that honoring the 12-month lease is the most important descriptor, as they can manage most other issues. It seems reasonable that non-payment of rent or late payment of rent would be another important consideration. However, management said that late or non-payment would result in an eviction from the apartment complex and thus these payment issues would be recorded as the tenant having stayed for less than the 12 months of their lease. Additionally, 12 month or longer lease terms are desirable as operating expenses are lower as the term of the lease increases, because apartments do not have to be repainted, and re-cleaned as often. Consequently, phase one used the number of months that the tenant lived at the apartment complex and correlated this length of time versus score.

Methodology for analysis of five additional credit scores

The study selected 100 tenants at random from those that leased an apartment during the year 2002. Since the apartment complex leases to about 8 to 10 tenants per month these 100 were the majority of the tenants from about a 10-month period. These 100 were selected by the apartment complex management and the names and social security numbers of the applicants were removed. The credit bureau Experian generated five scores for each applicant for purposes of this research. The FCRA permits this because Experian owns the rights to the scores and can run them for test purposes without a permissible purpose. A permissible purpose would normally be required for direct access to a consumer's credit file but a score is not considered direct access because details of the credit file are not viewed.

The five additional scores that were analyzed for each applicant are as follows. The management of the credit bureau (Experian) suggested these five scores as possibly the best choices for use in the apartment rental industry.

1. Sureview Non Prime Score (number 23 on Table 1).
2. FICO Mortgage Risk Score (number 13 on Table 1).
3. Fair Issac Advanced Risk Score (number 21 on Table 1).
4. FICO Installment Loan Score (number 5 on Table 1).
5. Fair Issac Finance Score (number 9 on Table 1).

Phase Two Overall

Phase two analyzed 60 additional tenants from 2003 and 2004, identified characteristics of each applicant (Appendix C), and developed a new model using neural networks. Seventy-six variables were collected and simplified into 10 variables (Appendix D) for use by the neural network. This number of inputs produced a complex and unstructured problem and various machine learning methods have been shown to perform reasonably well (Piramuthu, 1998). Neural networks were used because in this problem of predicting applicant behavior, it was unknown how, or even which tenant characteristics (independent variable input) actually affect the predicted output of lease honored or not honored (dependent variable). Furthermore, it was also unknown how inputs were related to each other and thus affected output in combination. In the analysis performed in phase one of this research only one tenant characteristic, that of credit score, was used as an input so in that case, a traditional statistical technique of linear regression could be used. Neural networks are a problem solving methodology that can analyze large amounts of data, to establish patterns and characteristics, in situations where rules and relationships are not known (Turban, 2001) as was the case in this research. Creation of the neural network model required attention to four major areas of focus (Turban, 2001).

1. Data Collection and Preparation. Collect Data and separate into training data and test data
2. Prepare Network. Define a network structure and select a learning algorithm
3. Start Training and Test. Transform data if necessary to network inputs and train and determine weights
4. Implementation. Use the network with new data

Data Collection and Preparation

In general, the more data used with neural networks, the better the results. Larger data sets increase processing time during training but improve the accuracy of the training and often lead to faster convergence to a good set of weights. For a moderately sized data set, typically 80% of the data are randomly selected for training, and 20% for testing. For small data sets such as in this research, a slightly higher percentage is sometimes used for training and testing as 83% and 17% respectively. Freeman (1999) recommends that half the development time be spent in the data collection and preparation phase.

Prepare Network

The choice of network structure (in the form of the number of layers and nodes) and the choice of a learning algorithm are important and require careful consideration. Currently, however, there is no systematic set of rules for the determination of the optimal number of hidden layers or nodes for networks (Lee & Lam, 1995). Yen and Lu (2002) developed a hierarchical approach to this two-object optimization algorithm (number of layers and number of nodes) that proved promising. Fortunately, most neural network software packages provide guidance in these areas by making choices and presetting values that generally work well. For example, the number of nodes in a single hidden layer should be somewhere between ½ and 1½ times the total number of input and output nodes (1½ seems to be better). In addition, error tolerances of 10% and a learning rate of .1, with randomized weights, are a good starting point and these were used in the research. The setup of the network and these typical values has more of an impact on the time to train rather than on the accuracy of the output, although in some specialized applications both are affected equally. Since the data set in this research was small, preparation of the network was less significant because the time to train the neural network was short (a matter of seconds or, at most, minutes). In addition, the neural network software used was capable of selecting its own values.

Start Training and Test

The data was formatted as required by the neural network software system and when this was completed, the training phase began. The training phase consisted of presenting the training data to the network (80% to 83% of the data) so that the weights were adjusted to produce the desired outputs for each of the inputs. The software completed several iterations of the complete training set until a consistent set of weights was derived.

Once the training had been completed, the testing examined the performance of the network (using the derived weights) by measuring the ability of the network to classify the testing data correctly (using the remaining 17% to 20% of the data). The network was generally not expected to perform perfectly (a zero error is difficult if not impossible to obtain). In this research, a “1” meant a tenant satisfied the lease. For practical purposes, an output between 0.75 and 1.25 was considered to indicate a correct prediction. Similarly, a “0” meant a tenant did not satisfy the lease. For practical purposes, an output between -0.25 and 0.25 was considered to indicate a correct prediction. Since neural networks are usually an alternative to an existing, more labor-intensive process, it is usually possible to obtain benchmarks against which to test the system. For example, Ainscough and Aronson (1999) investigated the application of neural networks to the prediction of retail sales (with inputs such as price, promotions, and so forth). They compared their results to those of regression and improved the adjusted R Square from .5 to .7. In addition, they suggest that the weights be analyzed to look for unusually large values that may indicate problems, or overly small weights that may indicate irrelevant input factors and unnecessary nodes. Moreover, certain weights that represent major factors in the input can be selectively deactivated to make sure that outputs respond accordingly. The software provided feedback on the importance of each of the input variables and these were examined after each run on the neural network in order to select the most important variables.

Implementation

With a commercial model, the technology department would install the finished neural network model into the decision process in the working business environment. This step was not applicable here.

Methodology used for developing the new neural network model

Development of the neural network based credit scoring model followed a nine-step development process (Fensterstock, 2001).

1. Sample Data Selection. The sample of 60 applicants was selected and the additional variables were collected and purchased on each applicant.

2. Data Scaling. While not specifically necessary, training a neural network is most efficiently accomplished if all the inputs have a similar value range. To meet this

requirement, the raw data was scaled as needed to produce values in the same range. For example, total income was divided by 1000 to reduce the numeric size of this variable in line with other variables and data that was not numeric (adult only, multiple adults, or adult with children) was scaled as “1” for adult only, “2” for multiple adults, and “3” for adult with child(ren).

3. Data Splitting. The data was split into two data sets with each set consisting of the various types of applicants. One of the sets was used for training (about 49 applicants) and the other was used to validate the model (about 11 applicants). Using different sets for training and validation helped to ensure that the model’s performance was real and not just a result of memorizing the idiosyncrasies of the data set. The software split the data automatically and randomly.

4. Relationship Analysis. This analysis determined if any redundant variables existed, such as variables that are not needed because they correlate to a high degree with other variables in the data set. This information would normally be used to a) fine tune the model, b) reduce the number of inputs needed and c) identify those characteristics of a typical applicant that are most predictive. Because the data set was relatively small (60 applicants), when compared to the 76 variables that were collected, all the variables could not be used in the model creation at the same time. Therefore, the variables that naturally correlated with each other were simplified and combined into the following 10 variables for use in building the model (details in Appendix D). For example, an applicant with two cars had the *Blue Book* values for both vehicles added together, thus reducing six variables (make, model, and year twice) to one variable.

1. State of previous residence.
2. Adult only, multiple adults, or adult with children.
3. Total applicant income.
4. Total *Blue Book* value of all vehicles.
5. Number of driving infractions.
6. Applicant has criminal background.
7. Total loan balance.
8. Total monthly payments.
9. Total credit file inquiries.
10. Percentage of total accounts that are satisfactory.

With 60 data points, even this reduced set of 10 variables could not be used simultaneously and the variables were tested in smaller groups as described later.

5. Initial Model Creation. The initial model was developed by training the network using one of the data sets. Training was achieved by presenting to the neural network each data record with the inputs and the output values. For each record, the inputs were passed to the network's input nodes, and the network's outputs were compared to the actual outputs found in the data. The discrepancy between the predicted outputs and the actual outputs were used to adjust the weights within the model. The optimum training regimen involved passing the entire training set to the network until the model converged (i.e. no additional improvement occurs in the predictive accuracy of the model). The neural network software performs this process automatically.

6. Model Evaluation and Testing. The prediction accuracy of the neural network model was tested by using the portion of the data that had been reserved for this purpose. This step ran the model with the previously unused data in order to ensure that the performance of the model was not simply a result of memorizing the data characteristics.

7. Variable Selection. This step used the results from the model creation and test to determine which of the input variables were predictive. Finding predictive variables was a key part of this research. This was accomplished by testing the 10 variables in groups until a smaller set of most important variables was obtained.

8. Final Model Creation. In a commercial application, a final model is produced using the most predictive input variables selected in step seven and this model would be "frozen" in the development process. This step is not applicable to this research.

9. Implementation. With a commercial model, the technology department would install the finished neural network model into the decision process in the working business environment and provide end-user training. This step is not applicable here.

A rule of thumb is that the best models using neural networks are created when about 10 data points exist for each variable (Witten & Eibe, 2005). In this case, a data set of 60 applicants means that about six variables could be used at any one time to create the model. With the 10 variables used in this research, there were still more variables available than could be used in any one pass of the model creation software. To accommodate as many variables as possible, the neural network software was run multiple times with a different selection of input variables each time, to determine which combination of variables produced the best predictive values. Six variables were chosen from the list of available 10 variables and the neural network was rerun until all the variables had been tested at least once. Eighty-four combinations were each run three times using randomly selected data points for a total of 252 runs of the neural network software.

In general, this process of choosing the input variables used to create a model for making predictions and arriving at recommendations is an important decision. When building models with neural networks, it seems natural to assume that having more information is always better than having less, since the model-building tool should do no worse with additional input variables because no vital information has been removed. The reality of the situation is counter-intuitive. Adding inputs gives the model more things to consider, thus extra variables can confuse and dilute the outcomes. Since practical experience clearly shows that paring down the number of inputs often results in models that are more accurate or more robust, it is necessary to find sound ways to reduce the quantity of variables used as candidate inputs.

Several automated strategies have been developed to accomplish this type of input reduction and input management including, a)exhaustive search, b)ordered search, c)genetic search, and d)heuristic search (among others) (Dwinnell, 1998).

1. Exhaustive search is the only method guaranteed to find the optimal subset for an arbitrarily complex problem and this was the method used in this research. While in most commercial situations this method is too slow since all variables have to be tested in many combinations, if the number of inputs is reasonably small, as in this case, this is a viable option.

2. Ordered search involves systems like forward selection and backward elimination, which are often employed with multiple linear regressions. A forward search

starts by trying all possible models that use a single input. The best single input is retained and a search begins on the remaining candidate inputs to become the second input. The input that most improves the model is kept and so on. This process ends either when the model ceases to improve or when candidate inputs are exhausted. A backward search works exactly like a forward search, except that it moves in the other direction. Backward searching begins with a model that uses all the inputs and then removes input variables one at a time. Interestingly, forward and backward searches may not result in the same set of inputs. Many variations on these searches are available and this technique was used in this research to attempt to understand the predictive nature of each individual input.

3. Genetic search is a procedure driven by genetic algorithms (GA) -- powerful systems that are very good at handling difficult optimization problems. GAs cycle through many iterations and, within the context of input selection, require more stringent testing to ensure that they have not accidentally located a bad solution that merely looks good. This type of search required more data than was available in this research.

4. Heuristic search modeling systems perform their own input selection as part of the modeling process. Symbolic machine learning systems (like those that search for IF...THEN rules) do this implicitly in their selection of variables to be used in the IF side of conditional statements. Generally, these systems are used with large data sets to assist in the selection of input variables. The neural network software used in this research included its own version of heuristic preprocessing of the data although it did not choose its own input variables.

This research used the exhaustive search technique, essentially trying different combinations of variables until the most predictive variables and combinations were found. This would not be necessary if a larger data set was available (i.e. more than 60 applicants) which enabled testing of all the variables at once.

Neural networks are widely used to find solutions to complex problems where the relationship between inputs and output is not clear. The complexity and hidden layers in the operation of neural networks make it difficult to understand this relationship, even when the network is accurately predicting outputs. As a result, while the development of the system software can follow the usual formal development paths, its application and implementation requires a degree of "art" in addition to "science." This has caused the software development community to develop a set of "rules of thumb" to aid developers in applying and implementing neural networks and these were used in this research. Freeman (1999) has summarized these as follows:

1. Quality of data is an important determinant of neural network success and the developer must understand the data and its relationship to the problem. Of course, there must be a sufficient quantity of data to provide adequate training and testing of the neural network. If possible, the developer should attempt to obtain an additional final portion of the data for retesting prior to any commercial implementation. While a significant amount of time was spent to ensure data quality, retesting was not possible in this research as sufficient additional data was not available.

2. Make sure that the problem is sufficiently complex to require a neural network. Many problems can be adequately solved using standard statistical and regression methods as was used in phase one of this research.

3. Avoid trying to map multiple functions using a single neural network. While neural networks excel in handling complex problems involving many steps and unknowns, trying to explain the relationship between inputs and outputs becomes more complex as the internal workings of the network become more complex. Multiple networks, with each handling a single function or piece of the problem, are easier to maintain, debug, and explain but were not necessary in this research.

4. Use as few training passes as possible. Overtraining can cause the network to be highly accurate with the training data set but predict poor results in actual use.

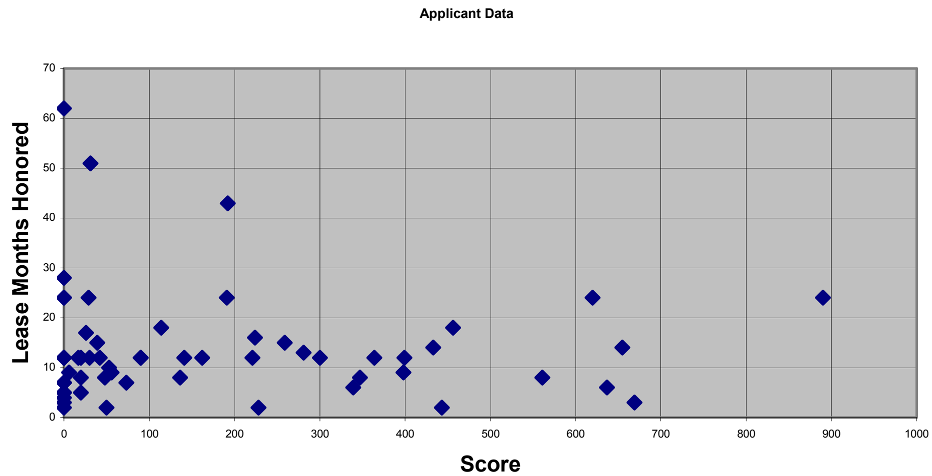
5. Use as few hidden layers as possible. Excess layers generate slow response and increase processing time. A neural network will tend to configure itself during the learning process such that excess hidden layers receive very small weights, making them non-participants in predicting outputs. Generally, begin with one hidden layer and try two later if results are not satisfactory. Skapura (1996) argues that a maximum of three hidden layers will solve virtually all neural network problems. This research used software that could modify layers to fit the data better and it did this automatically.

6. Involve domain experts, statisticians, and users from the early phases of development. This step was not applicable to this research.

Formats for presenting data

Phase One Analysis of Commercial Scores using Linear Regression

The format was a) a graph of the plot of the data followed by b) the results of the linear regression analysis performed for each set of data, followed by c) the results of testing the mean scores of each group using the t-test. An example follows (Figure 10 and Table 6) for a hypothetical set of applicants.



Source: developed by researcher

Figure 10 Sample hypothetical results for a non-predictive credit score.

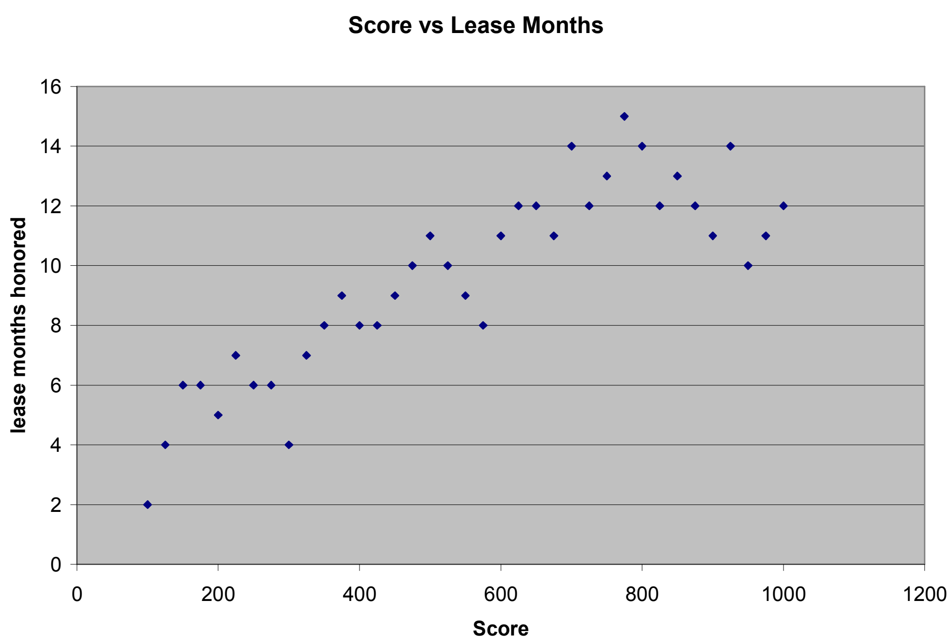
Table 6 Linear Regression Results for Hypothetical Non-Predictive Score

Summary Regression Statistics	
Multiple R	0.001997
R Square	3.99E-06
Adjusted R Square	-0.01666
Standard Error	11.23894
Observations	62

ANOVA	*Df*	*SS*	*MS*	*F*	*Signif. F*
Regression	1	0.03021	0.03021	0.00023	0.98771
Residual	60	7578.82	126.313		
Total	61	7578.85			

	Coefficients	*Std Error*	*T Stat*	*P-value*	*Lower 95%*
Intercept	12.93445	1.808151	7.15341	1.37E-09	9.31760
X Variable 1	0.000102	0.006617	0.01546	0.98771	-0.01313

In this hypothetical example, there is no relationship between the score and tenants' performance in honoring the 12-month term of their lease as can be observed visually in the graph. A regression of the data indicated an R Square approaching zero, which implies no correlation. Linear regression was used because as the score goes up, the risk goes up, and this relationship operates in a linear fashion. In addition, regression has been found to be the most accurate of the traditional methods applied to credit scoring problems (West, 2000). In this hypothetical example, the conclusion would be that this credit scoring model is not predictive of applicant behavior in honoring their lease. A credit scoring model that was predictive would tend to have the data points flowing toward the higher (better credit) end, which would mean that those applicants with better credit would be more likely to honor their lease to the end. An example of this graph is illustrated in Figure 11.

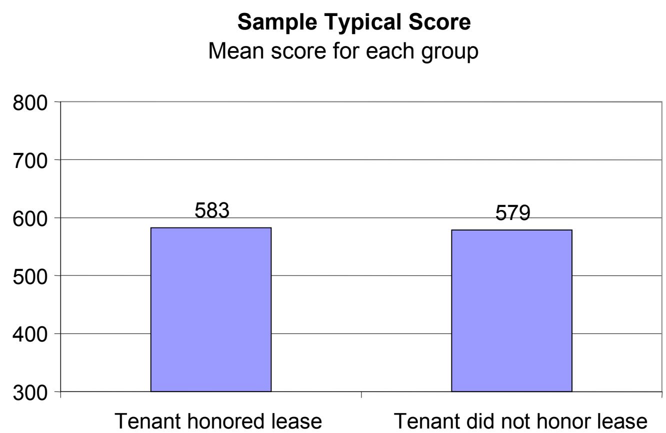


Source: developed by researcher

Figure 11 Sample hypothetical results for a predictive credit score.

A score that produced a graph similar to Figure 11 would be predictive and this would be proved or disproved statistically via the regression results. If one of the scores to be tested had been found to be predictive, this score would have been used as an input in the creation of the new model with the neural network software.

The means for group one and group two were displayed on bar charts with the t-test statistics listed (Figure 12).



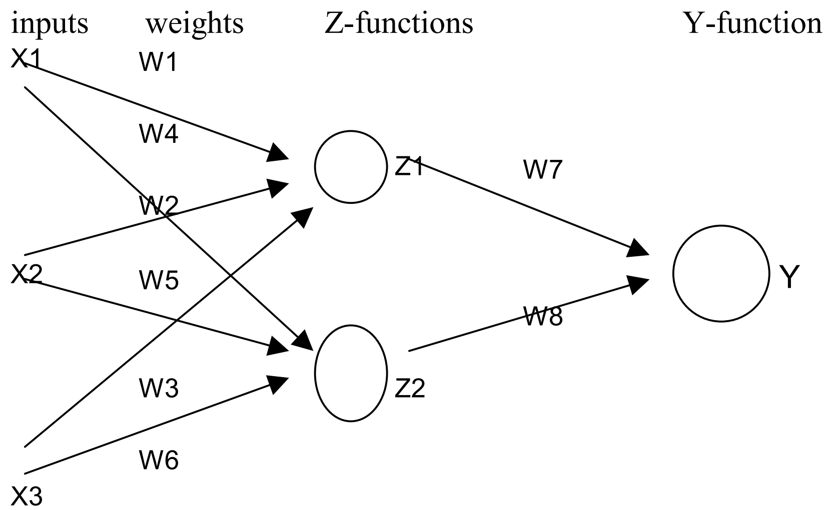
Std. Error	4.22	2.87	
R Square	.02125	.00203	
T-test Mean	531	531	p = .995

Figure 12 Sample analysis of means testing.

Phase Two Analysis for Creation of New Neural Based Scoring Model

Phase Two of the research used neural network modeling software to create a new scoring model based on credit financial data and other characteristics of the applicant. The intent was to improve the predictive capabilities of the model by combining financial and lifestyle information on the applicant. Figure 13 shows the typical information flow inside a neural network (Eberhart & Dobbins, 1990). The inputs in this case (shown as

X1, X2, and X3) were the applicant characteristics being tested, while the Y-function results were the expected outputs (i.e. lease honored). The model managed its own weights and z-function.



Source: Eberhart and Dobbins

Figure 13 Internal information flow of a neural network.

In this research, there were 10 inputs (such as X1, X2, X3, X4, X5, X6, X7, X8, X9, and X10) and one output variable (Y, that is lease honored). The neural network model was run many times with various combinations of inputs from these variables, in an attempt to create a model that accurately predicted the tendency of the applicant to honor or not honor the full term of the lease. Part of the output of the neural network is presented in Table 7 as summary information.

Table 7 Sample Neural Network Output

	Training set	Test set
# of rows:	46	11
CCR:	n/a	n/a
Average AE:	0.37278434	0.33837483
Average MSE:	0.20332339	0.15832975
Tolerance type:	Absolute	Absolute
Tolerance:	0.25	0.25
# of Good forecasts:	17 (37%)	4 (36%)
# of Bad forecasts:	29 (63%)	7 (64%)

R Square: 0.2193
Correlation: 0.4843

The neural network software to create the model was chosen from a lengthy list of available products (Appendix E). This research used neural network software called Forecaster XL provided by Alyuda Corporation. Forecaster XL was chosen because it does automatic neural network architecture and parameter selection based on the data**.** Additionally, it provides heuristic data preprocessing, algorithm selection, and neural network preparation. In essence, the software was doing much of the work needed to prepare and fine-tune the neural network automatically.

Resource Requirements

The first contingency for the successful completion of this research was the availability of data on applicants from the apartment complex. The management was interested and excited about using their data to develop a better understanding of applicants and to improve the selection process. Thus, the data was made available. A secondary contingency was the willingness of the credit bureau Experian to run the additional five scores on the applicants for phase one. This company also was interested in pursuing this research and provided the scores. Thirdly, neural network software was needed but this was commercially available. Lastly, the cost of the additional variables for phase two of the study was expensive but the apartment complex paid these costs.

Summary

The research proceeded in two phases. Phase one analyzed the current credit scoring model used by the apartment complex to validate the accuracy of this model in selecting applicants who will honor the lease. It then analyzed five other commercially available scoring models to determine if one of these models was more predictive of applicant behavior in honoring leases. Linear regression was used as the statistical analysis technique and the t-test was used for examining the statistical difference between the means of two groups. Phase two used neural networks to expand the analysis by including additional applicant information beyond financial credit data in an attempt to create a model that was more predictive of applicant behavior.

Chapter 4

Results

Introduction

The first phase of this research analyzed the results of using six commercially available credit scores applied in one apartment complex to the task of selecting applicants. This research determined that these six scores are not predictive and possible explanations are given. This part of the analysis answered the research question: How effective are commercially available credit scores in predicting applicant financial behavior when renting an apartment?

Phase two of this research developed a new model, using neural network techniques, which included both credit data and other lifestyle data about the applicant. The hypothesis was that adding this lifestyle data would improve the accuracy of the new model over currently available models based only on credit data. This research indicates that accuracy is greatly improved. This part of the analysis answered the research question: How is the prediction accuracy of a new neural network based credit scoring model improved by adding lifestyle data to the credit report data?

Data Analysis of Six Commercially Available Credit Scores

Analysis of National Risk Score currently used by apartment complex

The study selected 60 tenants at random from those that leased an apartment during the years 1999 and 2000. The scoring model used by this apartment complex is the National Risk Score provided by Experian, one of the three major U.S. credit bureaus. This particular model creates a number score that directly corresponds to risk. Specifically, a score of 100 indicates that this applicant has a 10.0% probability that they will NOT fulfill their financial obligations. A score of 525 would indicate that this applicant has a 52.5% probability that they will NOT fulfill their financial obligations (essentially the higher the number for the National Risk Score, the higher the risk). This is opposite to the typical credit score that has a range calibrated so that as the score number gets higher, the risk gets lower. Although the scoring model was run on every applicant, there were 16 applicants for which a score could not be created. This was due primarily to a lack of credit history and insufficient data in the credit file to run the model. These applicant files were not used in the linear regression analysis. Two analyses of the data were performed. The first examined the score using the data in its entirety and the linear regression results are displayed in Figure 14 and Table 8. There is no relationship between the score and the tenant's performance in honoring the 12-month term of their lease as the R Square approaches zero, which implies no correlation. Linear regression was used because as score goes up, the risk goes up, and this relationship operates in a linear fashion.

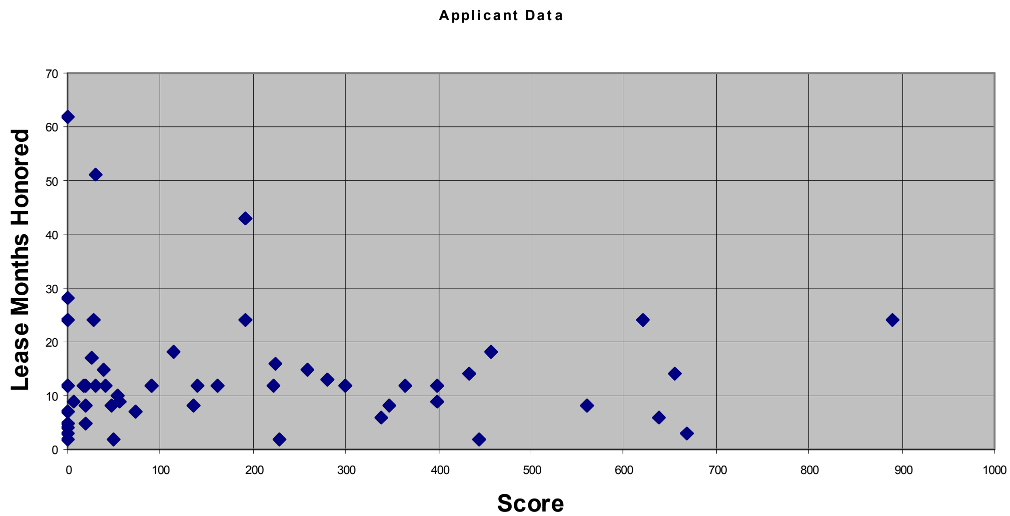


Figure 14 Linear regression result for National Risk Score using all the data.

Table 8 Linear regression result for National Risk Score using all the data

Summary Regression Statistics	
Multiple R	0.001997
R Square	3.99E-06
Adjusted R Square	-0.01666
Standard Error	11.23894
Observations	62

ANOVA	*Df*	*SS*	*MS*	*F*	*Signif. F*
Regression	1	0.030215	0.03021	0.000239	0.987712
Residual	60	7578.825	126.313		
Total	61	7578.855			

	Coefficients	*Std. Error*	*T Stat*	*P-value*	*Lower 95%*
Intercept	12.9344	1.80815	7.15341	1.37E-09	9.317606
X Variable 1	0.00010	0.00661	0.01546	0.98771	-0.01313

A second analysis of the data was performed that examined the data subjects (applicants) when divided into two groups based on their fulfillment of the 12-month term of the lease agreement. Group one contained those applicants that fulfilled the lease term and stayed for 12 months and the linear regression results are presented in Figure 15 and Table 9. Group two contained those applicants who stayed less than 12 months and the linear regression results are presented in Figure 16 and Table 10. Group one represents the desirable applicants that the score should identify. No correlation exists between the score and length of stay for either group one or group two. R Square for both groups is very low: .08 for group two and near zero for group one. This lack of correlation can also be seen when examining the percentage of applicants in the data set as shown in Table 11.

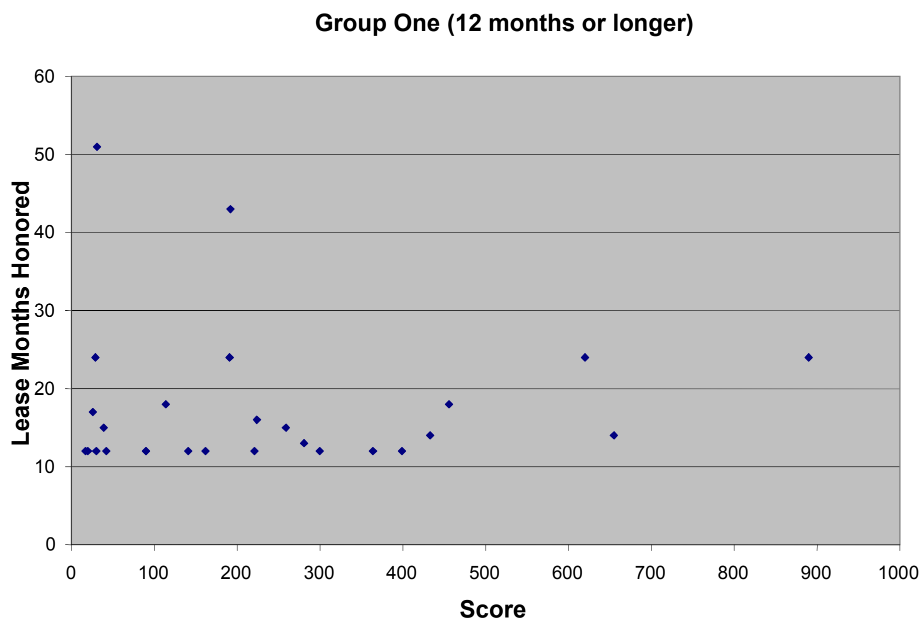


Figure 15 National Risk Score results for tenants who satisfied the lease.

Table 9 Linear Regression Results for National Risk Score for tenants who satisfied the lease

Group One (rented 12 months or longer)

Summary Regression Statistics	
Multiple R	0.011319561
R Square	0.000128132
Adjusted R	-0.039866742
Standard	9.724944578
Observations	27

ANOVA	*Df*	*SS*	*MS*	*F*	*Signi. F*
Regression	1	0.3029905	0.3029905	0.0032037	0.95531265
Residual	25	2364.3636	94.574547		
Total	26	2364.6666			

	Coefficients	*Std.Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	17.442892	2.7321563	6.3842953	1.1038E-06	11.8159155
X Variable 1	0.0004816	0.0085089	0.0566014	0.95531265	-

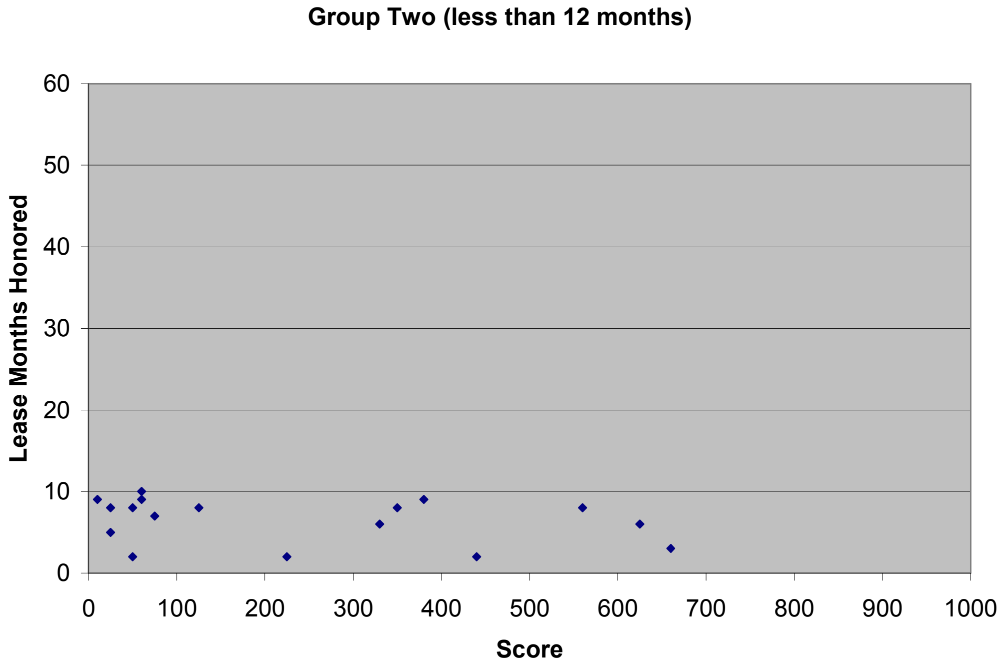


Figure 16 National Risk Score results for tenants who did not satisfy lease.

Table 10 Linear Regression Results for National Risk Score for tenants who did not satisfy the lease

Group Two (rented less than 12 months)

Summary Regression Statistics	
Multiple R	0.291214404
R Square	0.084805829
Adjusted R	0.023792884
Standard	2.685867358
Observations	17

ANOVA

	Df	*SS*	*MS*	*F*	*Signif. F*
Regression	1	10.027042	10.027042	1.3899645	0.25677452
Residual	15	108.2082	7.2138834		
Total	16	118.23529			

	Coefficients	*Std.Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	7.2888442	0.9518633	7.65744782	1.46987E-	5.25999420
X Variable 1	-0.00340606	0.0028890	-1.1789675	0.25677452	-

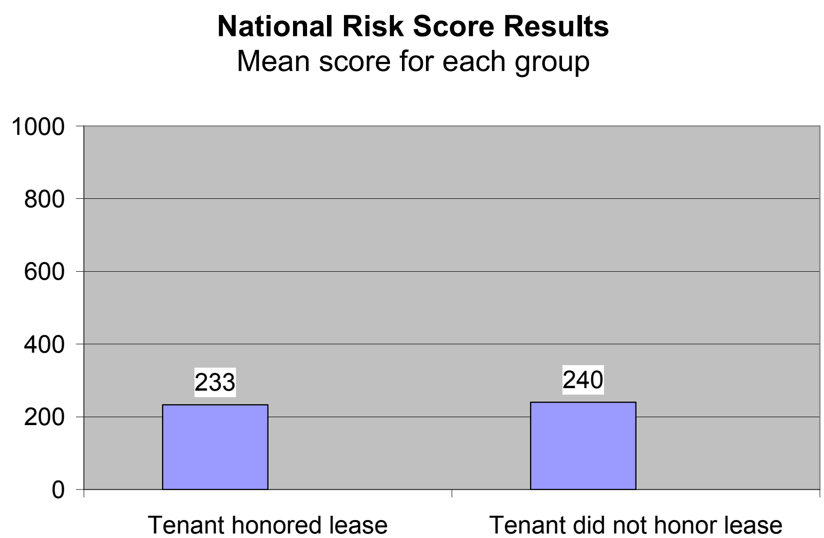
Table 11 Distribution of tenants tested with National Risk Score.

Score	Number of Applicants	Perceived Risk Level	Number who Rented 12 Months or more	Number who Rented less than 12 Months
None	16	Unknown	7 (43%)	9 (57%)
1 to 200	24	Low	15 (63%)	9 (37%)
210 to 500	15	Medium	9 (60%)	6 (40%)
Over 500	5	High	3 (60%)	2 (40%)

The number of applicants that rented for at least the 12-month term of their lease is about 60% whether the applicant score was in the preferred range of 1 to 200 or was above this range, as shown in Table 11. While there appears to be a slight increase in tenants who rented at least 12 months in the preferred range of 1 to 200 (i.e. 63% compared to 60% and 60% as the risk increases), this slight increase could be a sampling error as a change of only one applicant in the range 1 to 200 could change the percentage by up to 4 percentage points.

If this credit scoring model was accurately working as a predictor, the tenants in the preferred score range of 1 to 200 should have had significantly better results, which was not the case. One clear result from the data is that applicants without enough credit history to run a score represented the highest business risk since 57% of this group (9 of 16) stay for less than 12 months versus 38% (17 of 44) of those with a score.

With an accurate, predictive model, the hypothesis is that group one applicants (i.e. satisfied the lease) would have a better mean credit score than group two applicants. The mean scores of both groups one and two are outside the most desirable range of 1 to 200. Furthermore, the mean and median scores of both groups are similar, implying little predictive value in using this scoring model (Figure 17)



Std. Error	9.72	2.68	
R Square	0.0001	0.0848	
T-test Mean	234	240	p =.929

Figure 17 Mean score for tenants using the National Risk Score.

The National Risk Score, which has been used by this apartment complex for several years, is not helpful in choosing applicants for apartment rentals. There is no correlation between the credit score of an applicant and an applicant's honoring of their 12-month lease and no predictive value.

Analysis of five additional commercially available credit risk scores

This part of the study selected 111 tenants at random from those that leased an apartment during the years 2001 and 2002 and performed the same analysis as just discussed for the National Risk Score. These 111 were selected from a pool of about 300. Of these 111 files, 83 were used in the analysis. The remaining 28 could not be used for various reasons: 16 had multiple persons on the lease so the financial obligation and score were unclear, five had no scores or limited scores, and seven had never actually moved in or no move-in data could be found. Not all 83 of the useable files had all five scores because some of the scores could not be run for various applicants because of problems in the credit file. In general about 70 to 75 applicant files could be used for the analysis for each score (Appendix F).

The five additional scores analyzed for each applicant are as follows. The management of the credit bureau (Experian) suggested these five scores as possibly the best choices for use in the apartment rental industry.

1. Sureview Non Prime Score (number 23 on Table 1). This is a risk assessment tool developed by Experian specifically designed for non-prime bankcard issuers. It was developed to make predictions for five major classifications of consumers: 1) thin credit history and a limited number of derogatory trade accounts 2) young, full credit history and may have a limited number of derogatory trade accounts 3) mature, full credit history and may have a limited number of derogatory trade accounts 4) a high percentage of delinquencies or a bankruptcy on file 5) a high percentage of delinquencies or a

bankruptcy on file and at least one of the delinquencies is recent. These classes of consumers tend to be apartment rental applicants.

2. FICO Mortgage Risk Score (number 13 on Table 1). This model uses an in-depth review of the information in a consumer's credit file and attempts to identify customers most likely to result in serious delinquency, charge-offs and bankruptcy. This model is also sold by Equifax under the brand name "Beacon," and is the most widely used consumer credit score for mortgage loan applications.

3. Fair Issac Advanced Risk Score (number 21 on Table 1). This model helps determine which accounts are most likely to be profitable and which pose the greatest credit risk. It predicts the probability of serious derogatory credit behavior and indicates the likelihood that a customer will become seriously delinquent within the next 24 months (most apartment renters tend to a shorter term of 12 to 36 months).

4. FICO Installment Loan Score (number 5 on Table 1). This model predicts a consumer's performance on repaying short-term installment loans such as 36-month car loans or other leases. This type of financial transaction is similar to the apartment rental decision.

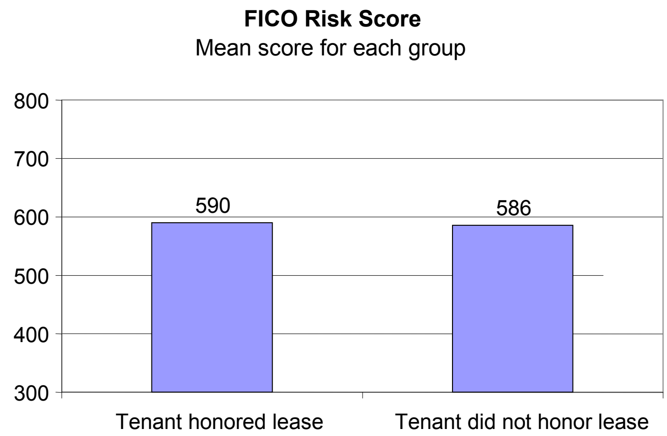
5. Fair Issac Finance Score (number 9 on Table 1). This model predicts a consumer's financial performance for loans originated at non-traditional finance companies, "cash your paycheck here" companies, or pawnshop-type lending businesses. These are short-term loans usually made to high-risk borrowers.

As previously described, the data for each score was examined twice. First, each score was compared to all the data then, second, the data was separated into two groups of applicants: those who satisfied the lease and those who did not satisfy the lease. The score results were first analyzed using linear regression using all the data. The results are illustrated in Table 12. The low R Square results indicated that there is no relationship between the score and the length of time that the tenant honored the lease for any of the five additional scores.

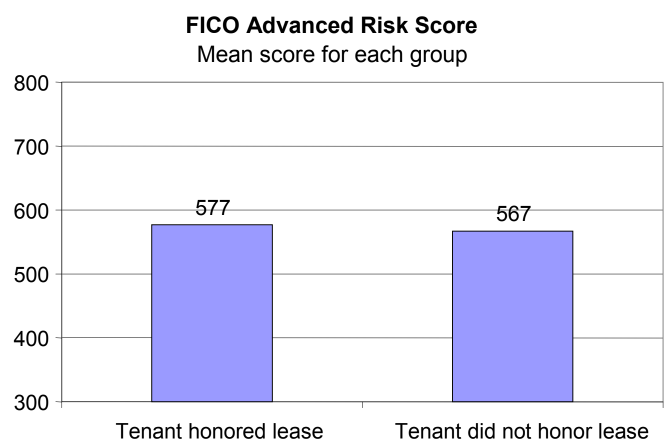
Table 12 Results examining each score using all the data

Score Name	R Square Results
FICO Risk Score	.007636
FICO Advanced Risk Score	.001535
FICO Installment Loan Score	.001535
FICO Finance Score	.007005
Experian Sureview Score	.0000332

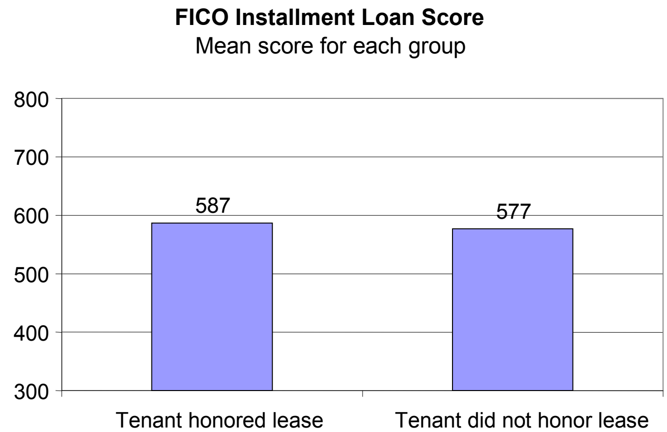
The second analysis divided the data into two groups: group one were those tenants who satisfied the term of the lease by staying for 12 months or longer while group two were those tenants who stayed less than 12 months. Linear regression was used on each score and both groups, and the mean score and t-test calculated. The results are illustrated in the five bar charts in Figure 18. The is little difference between the mean scores of a) the desirable group of applicants who honored their lease for 12 months or longer and b) the less desirable group that honored their lease for less than 12 months. The regression analysis indicated no correlation between score and lease months honored for either group and the t-test indicated that the means of each group are not statistically different. None of the five additional scores tested were predictive of applicant behavior in honoring their lease.



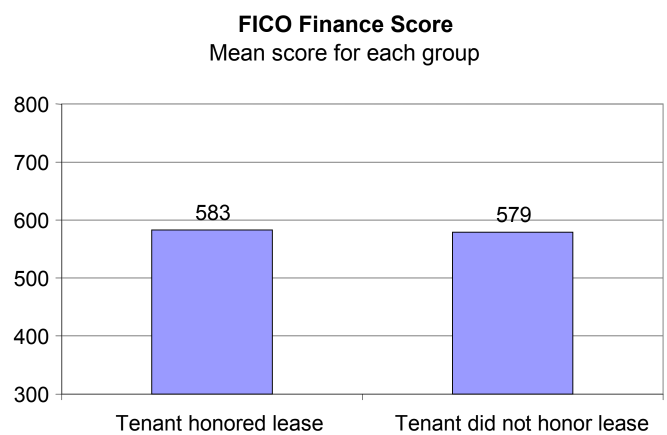
Std. Error	4.18	2.85	
R Square	.03859	.01573	
T-test Mean	508	495	p =.787



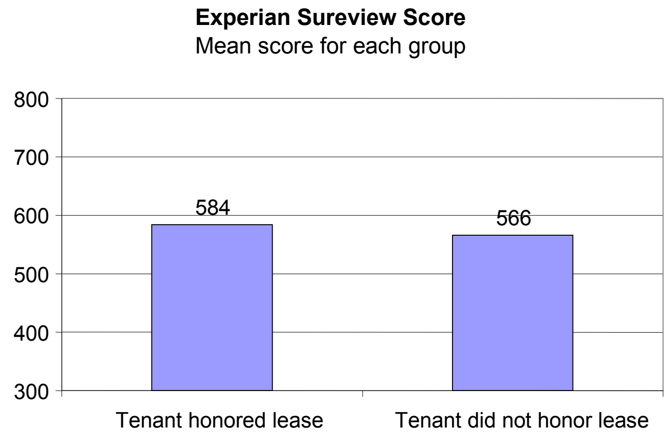
Std. Error	4.24	2.86	
R Square	.01334	.00542	
T-test Mean	498	478	p =.686



Std. Error	4.23	2.87	
R Square	.01886	.00305	
T-test Mean	534	529	p =.905



Std. Error	4.22	2.87	
R Square	.02125	.00203	
T-test Mean	531	531	p =.995



Std. Error	4.26	2.87	
R Square	.00080	.00497	
T-test Mean	538	494	p =.568

Figure 18 Mean for five scores.

Results of analysis using neural networks to create new models with expanded applicant data

The commercial credit scores previously examined were created using only credit report data and were not predictive. This portion of the research expands the data available on each applicant to include both credit data and lifestyle data for use by the neural networks. The data was collected from the years 2003 and 2004 on 60 tenants and included up to 76 variables on each. These variables then were simplified and combined into a set of 10 variables for use in the neural network model (Appendix D). For example, data was collected on vehicles owned by the applicant including year, make and model. If the applicant owned two vehicles, this produced six variables. These six variables were combined by obtaining the *Blue Book* value of each vehicle, adding the values together, and producing one variable of total *Blue Book* value for use in the neural network. A portion of the data collected on the 60 applicants is in Table 13.

Table 13 A portion of the data used in the neural network

App. Number	Lease Fulfill 0=no 1=yes	Moving from 0=local 1=other	1=one adult 2=many adult 3=adult +child	Total Applicant Income (year $000)	Vehicle Blue Book ($000)	Driving Infraction? 0=no 1=yes	Criminal Activity Reported 0=no 1=yes	Total Loans Balance ($000)	Total Monthly Payment ($000)	Number Of Credit Inquiries	Percentage Satisfactory Accounts
1	0	0	1	14.4	1	1	0	7.902	0.668	4	100.0%
2	1	0	3	20.064	0	0	0	28.253	0.743	10	66.7%
3	0	0	2	36.0	5	1	0	19.287	0.545	1	90.0%
4	1	0	2	18.0	2	0	0	0.315	0.028	1	66.7%
5	0	0	1	23.4	12.375	1	0	0	0		
6	1	0	1	37.224	1	0	1	13.663	0.057	1	85.0%
7	0	0	3	16.2	5.15	0	0	75.649	0.761	3	25.0%
8	0	0	2	20.0	2	0	0	0.6	0		100.0%
9	0	0	2	20.8	0	0	1	1.664	0	3	0.0%
10	1	0	1	39.0	0	0	0	38.78	0.868	3	81.8%
11	1	0	1	46.8	2	0	0	1.507	0.137	2	100.0%
12	1	0	3	54.0	12	0	0	105.789	1.606	1	100.0%
13	0	0	3	44.72	2	0	0	0.661	0.053	1	66.7%
14	0	0	2	22.2	1	0	0	2.658	0	1	0.0%
15	0	0	2	19.2	0.5	0	0	15.494	0.516	1	100.0%
16	1	0	2	99.9	4	1	0	21.969	0.418	11	78.9%
17	1	0	1	26.4	5	1	0	26.063	0.973	5	66.7%
Input Variable Identifier =		D	E	F	G	H	I	J	K	L	M

One of the variables collected (variable D) could not be used, that is the variable of "moving from" (state of previous residence). This variable could have provided a description of the applicant as moving from a local address or from an out-of-state address. It was possible that out-of-state applicants might tend to honor the lease. Unfortunately, 58 of the 60 applicants examined were local so there was not enough variation in this variable to make it meaningful. This variable was therefore dropped and nine data points were used in the neural network. The apartment complex purchased the data related to criminal history and driving record on each applicant. The 60 tenants included 30 who had satisfied the lease and 30 who had not. This 50%/50% split is similar to the apartment complex's actual experience of 45% satisfy lease and 55% do not satisfy the lease (Figure 2).

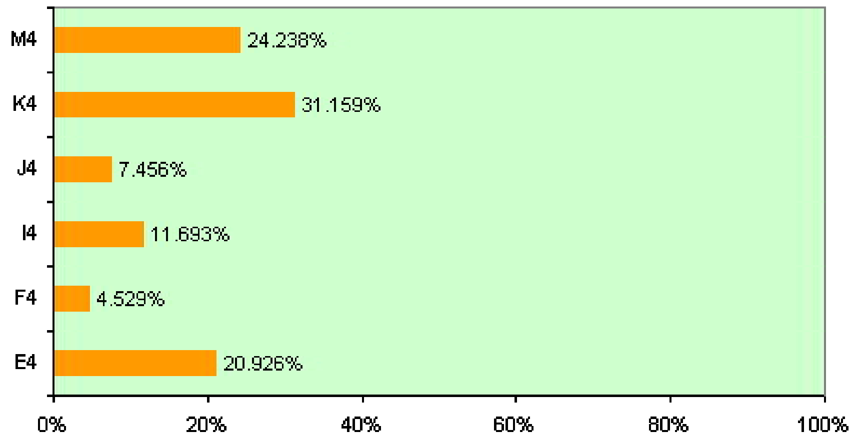
The neural network was first run using only the credit related data as input. These are the four variables of a) loan balance, b) total monthly payment, c) number of credit inquiries, and d) percent of satisfactory accounts. This provided a baseline for the later studies that add in lifestyle data and provided a comparison to the result from the tests of the commercial credit scores. Three runs produced R Square values of .2373, .2317, and .2800 with a mean of .2496. While these values are significantly better than those of the commercial credit scores, they are still low indicating a low prediction value. During each run, the software randomly chose data points for the training set and for the testing set. The neural network model was then run 100 times using only this credit data and the prediction accuracy was recorded. In the 100 runs, the neural network using only credit data correctly predicted 46.5% of the tenants in the test data. An accurate prediction was defined as in Table 14.

Table 14 Definition of accurate neural network prediction

Actual result	Neural network prediction	Neural network prediction
Actual value of 1 Tenant honored lease	Prediction in range of 0.75 to 1.25 considered accurate	
Actual value of 0 Tenant did not honor lease		Prediction in range of -0.25 to 0.25 considered accurate

In order to test all the variables (four credit plus five lifestyle variables), identify important variables, and possibly reduce the number of variables, the exhaustive search technique was used (Dwinnell, 1998). The 60 data points collected allowed six variables (out of the nine variables) to be tested simultaneously. There were therefore 84 possible combinations of variables and each combination was tested three times using the neural network software for 252 neural network runs. Each run of the software produced statistics indicating the performance of the network and the importance of each variable. Figure 19 illustrates some of these statistics and shows an R Square of .6054 with variable M (percentage satisfactory accounts) and K (total monthly payment) being the two input variables with the highest importance in this run. Results for the 252 runs of the neural network software are in Appendix G. The R Square for these 252 runs of the software had a range of .1614 to .6705, a mean of .3932, and a median of .3858. In each of the runs, the two most important variables were recorded (Table 15).

Input Importance



Summary

	Training set	Test set
# of rows:	47	9
CCR:	n/a	n/a
Average AE:	0.173055	0.13128627
Average MSE:	0.10938223	0.04187615
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	12 (26%)	5 (56%)
# of Bad forecasts:	35 (74%)	4 (44%)

R Square: 0.6054
Correlation: 0.8120

Figure 19 A portion of typical output results of the neural network.

Table 15 Occurrences of two most important input variables testing nine variables

Variable Name	Variable with Most Importance	Variable with Second most Importance	Total Occurrences
Adult, multiple adult, or adult/child (E)	0	39	39
Total Income (F)	9	51	51
Vehicle *Blue Book* (G)	8	40	40
Driving Infraction (H)	22	101	101
Criminal Activity (I)	2	9	9
Total loan balance (J)	4	4	4
Total monthly payment (K)	54	88	88
Credit Inquiries (L)	40	59	59
Percent satisfactory accounts (M)	113	0	113

Although no clear pattern of variable importance emerged, the criminal record variable and the total loan balance variable were of minor importance and these were dropped from the variable list leaving seven variables. The neural network was then run 100 times using these seven variables and the prediction accuracy recorded. The neural network using these seven variables accurately predicted 52.5% of the tenants in the test data. This prediction accuracy is higher than the 46.5% when using credit data only, indicating that lifestyle data is improving prediction accuracy (but still low).

The next analysis performed another exhaustive search on these seven variables selecting five variables from the remaining seven variables for each run of the neural network. There were possible combinations of 21 and each combination was run three times for a total of 63 runs. Results for the 63 runs of the neural network are in

Appendix H. The R Square for these 63 runs of the software had a range of .1668 to .5908, a mean of .369, and a median of .3516. In each of the runs, the two most important variables were recorded (Table 16).

Table 16 Occurrences of two most important variables testing seven variables

Variable Name	Variable with Most Importance	Variable with Second most Importance	Total Occurrences
Adult, multiple adult or adult/child (E)	2	9	11
Total Income (F)	5	14	19
Vehicle *Blue Book* (G)	3	5	8
Driving Infraction (H)	8	17	25
Total monthly payment (K)	15	4	19
Credit Inquiries (L)	6	10	16
Percent satisfactory accounts (M)	24	4	28

Although no clear pattern of variable importance emerged, the adult-adult/child variable and the vehicle *Blue Book* variable were of lesser importance and these were dropped from the variable list leaving five variables. The most important variables were a) total income, b) driving infractions, c) total monthly payment, d) credit inquiries, and e) percent satisfactory accounts. Since the number of variables has been reduced to five with 60 data points available for testing, these five variables can be tested in a number of different combinations.

The neural network was run 100 times first using all five variables and the prediction accuracy recorded. The neural network accurately predicted 55.8% of the tenants in the test data. The neural network was then run 100 additional times using the most important four variables of a) total income, b) driving infractions, c) total monthly payment, and d) percent satisfactory accounts (the variable of “credit inquiries” was removed for this test). The neural network accurately predicted 58.4% of the tenants in the test data. Two unique groupings of three variables were tested because the variables of total income and total monthly payment were equally important at 19 occurrences, and each of these two variables needed to be tested in combination with the other variables. This created two unique groupings of three variables. The neural network was then run 100 more times using the first grouping of the most important three variables of driving infractions, percent satisfactory accounts, and total monthly payments. The neural network accurately predicted 48.4% of the tenants in the test data. The neural network was then run 100 more times using the second grouping of the most important three variables of driving infractions, percent satisfactory accounts, and total income. The neural network accurately predicted 69.1% of the tenants in the test data. The neural network was then run 100 more times using only the two most important variables of driving infractions and percent satisfactory accounts. The neural network accurately predicted 41.4% of the tenants in the test data. The prediction results for all runs of the neural network, testing various combinations of variables, are summarized in Table 17.

Table 17 Prediction accuracy of neural networks with varying input variables

Variables Tested	Percent of tenants correctly predicted by neural network (Test Data)
4 credit variables only (baseline) loan balance (J), total monthly payment (K), credit inquiries (L), percent sat. accounts (M).	46.4 %
7 variables of 3 credit plus 4 lifestyle adult/child (E), total income (F), vehicle *Blue Book* (G), driving (H), total monthly payment (K), credit inquiries (L), percent sat. accounts (M)	52.5 %
5 variables of 3 credit and 2 lifestyle total income (F), driving (H), total monthly payment (K), credit inquiries (L), percent sat. accounts (M)	55.8 %
4 variables of 2 credit and 2 lifestyle total income (F), driving (H), total monthly payment (K), percent sat. accounts (M)	58.4 %
3 variables of 2 credit and 1 lifestyle driving (H), total monthly payment (K), percent sat. accounts (M)	48.4 %
3 variables of 1 credit and 2 lifestyle total income (F), driving (H), percent sat. accounts (M)	69.1 %
2 variables of 1 credit and 1 lifestyle driving (H), percent sat. accounts (M)	41.4 %

Summary

Six commercially available credit scores were tested and it was determined that these scores were not predictive of tenant behavior in honoring of the lease for 12 months. These scores were tested using all the data and also tested by dividing the data into two groups of tenants: those who honored the lease and those who did not honor the lease. In all cases, R Square was very low ranging from .00001 to .03859. There was little difference in the mean scores between tenants who honored the lease and those who did not, indicating no predictive value for any of the six scores tested (Table 18). These scores are created using credit data only.

Table 18 Summarized results of linear regression testing of six commercial scores

	All Applicants	Applicants that honored lease		Applicants that did not honor	
Score Name	R Square	Mean Score	R Square	Mean Score	R Square
National Risk Score	0.00001	233	0.00001	240	0.08480
FICO Risk Score	0.07636	590	0.03859	586	0.01573
FICO Advanced Risk	0.00153	577	0.01334	567	0.00542
FICO Installment Loan	0.00153	587	0.01886	577	0.00305
FICO Finance Score	0.00700	583	0.02125	579	0.00203
Experian Sureview	0.00003	584	0.00080	566	0.00497

In order to create a model that was more predictive, expanded data was collected on tenants that included both credit data and additional data on lifestyle. Seventy-six data points were collected on each of 60 tenants and this data was simplified into 10 variables for use by a neural network to create a new model. One of the 10 variables was dropped because it lacked enough variation to be relevant, thus leaving nine useful variables.

The neural network was run first using only the four variables of credit data. This provided a baseline for the study and a comparison to the test results of the commercial credit scores. In 100 runs using credit data, the neural network correctly predicted 46.4% of the tenants in the test data.

The neural networks were then run several hundred more times using various groupings of variables in order to identify those that were most predictive. The prediction accuracy of the neural network models improved to a high of 69.1% as it was focused on the most important variables, which in this case were the three variables of a) driving infractions, b) total income, and c) percent satisfactory accounts.

Chapter 5

Conclusion, Implications, Recommendations, and Summary

Conclusion

The first phase of this research analyzed the results of using six commercially available credit scores applied in one apartment complex to the task of selecting applicants. This part of the analysis answered the research question: How effective are commercially available credit scores in predicting applicant financial behavior when renting an apartment? Six commercially available credit scores were tested and it was determined that these scores were not predictive of tenant behavior in honoring of the lease. These scores were tested using all the data and tested by dividing the data into two groups of tenants: those who honored the lease and those who did not honor the lease. In all cases, R Square was very low and there was little difference in mean score between the two groups for all of the six scores tested, indicating no predictive value for any of the six scores tested. These scores are based on credit data alone. Only six of the 26 available scores listed in Table 1 were tested. However, the tested scores were chosen by the credit bureau, Experian, as those that should have been the most relevant to the tenant selection process. While it is unlikely, it is possible that one or some of the non-tested scores could be predictive.

Commercial scores for some of the tenant files could not be generated because of a lack of credit history and insufficient data to run the scores. One clear result from the data is that applicants with insufficient credit history to run a score represented the highest business risk since 57% of tenants without a score stay for less than 12 months versus 38% of tenants with a score. This implies that credit data is an important component of the decision process but the results from the testing of the commercial scores indicate that credit data alone is not a very strong predictor. This part of the analysis answered the research question: How effective are commercially available credit scores in predicting applicant financial behavior when renting an apartment? In this series of regression testing, the six commercial credit scoring models were found to be not predictive.

It was not a goal of this research to identify the reasons that existing commercial credit scores are not predictive. However, the problem with predictability of the six models may be based on the composition of the credit scoring model (statistical modeling issues). The reasoning is that these models were developed for other purposes such as home ownership and tend to filter out the typical tenants for apartment rentals (i.e. younger in age, less time on the job, lower paying job, and so forth). The weighting of the variables used in the model creation is targeted to answer or predict a different consumer behavior. Additionally, the problem with predictability of the six models tested seems to be centered on the use of credit data alone for the apartment rental application (data issues). The basic assumption with a credit data score is that there is a correlation between credit score and financial risk with an improving score indicating an improving financial risk. This correlation has been proven in the banking, credit card, and auto insurance industry but may not exist when picking applicants for the apartment rental market. The reason that credit data alone may not predict financial risk for apartment rental likely has to do with differing human behavior in transacting an apartment lease versus other financial transactions such as house purchases. It is beyond the scope of this research to discuss why human behavior in apartment rentals may be different from human behavior in house purchases or credit card usage. It is sufficient to note that apartment rental applicant performance cannot be predicted with credit data. However, it is reasonable to speculate that the applicant views an apartment rental as a short-term decision, similar to renting a car, while the existing credit scoring models predict behavior for longer-term decisions such as buying a house. This is not entirely accurate as at least one of the models tested (FICO Installment Loan Score number 5 on Table 1) attempted to predict consumer performance on loans similar to 36-month car leases and loans. The decision to abandon an apartment before the lease is up may be a decision likely based not entirely on credit matters, but instead based heavily on lifestyle issues (loss of job, change of school for children and so forth). A training program for the new tenant that explains the implication of breaking a lease may be a worthwhile program, as a lack of tenant understanding may be one of the root causes. The results of the testing of the six credit scores indicated that human behavior in the apartment rental market is different from other areas where credit scoring is used. Obtaining a better understanding of the underlying human nature elements that result in credit data being non-predictive in the apartment rental market, in contrast to other banking and credit card markets, could direct the researcher to become more specific in the choice of input variables other than the 10 variables suggested by apartment management. This is suggested as a possible fruitful area for future research.

Phase two of this research used neural networks to develop a new model using both credit data and other lifestyle data about the applicant. The hypothesis was that the addition of this lifestyle data would improve accuracy in selecting apartment rental applicants over currently available models based only on credit data. This part of the analysis answered the research question: How is the prediction accuracy of a new neural network based credit scoring model improved by adding lifestyle data to the credit report data? In order to create a model that was more predictive, expanded data was collected on tenants that included both credit data and additional data on lifestyle. Seventy-six data points were collected on each of 60 tenants and this data was simplified in 10 variables for use by a neural network to create a new model. The 10 variables developed for use in the neural network follow.

1. State of previous residence. (The managers thought that out-of-state tenants would have a higher tendency to honor the lease.)

2. Adult only, multiple adults or adult with children. (Multiple adults or adults with children would be less mobile and have a higher tendency to stay.)

3. Total applicant income. (cash available to pay debt)

4. Total *Blue Book* value of all vehicles. (High value vehicles would imply a tenant better able to handle financial obligations or conversely low value vehicles would be fully paid off thus freeing up cash for rent payments.)

5. Number of driving infractions. (background information)

6. Applicant has criminal background. (background information)

7. Total loan balance. (credit data-indication of debt load)

8. Total monthly payments. (credit data – an indication of other cash needs of tenant beside monthly rent)

9. Total credit file inquiries. (credit data – A high number of inquiries implies a tenant looking hard for credit, possibly due to financial problems not yet apparent.)

10. Percentage of total accounts that are satisfactory. (credit data - indication of tenant's tendency to reliably pay debts on time)

One of the variables collected could not be used, that is the variable of "moving from" (state of previous residence). This variable could have provided a description of the applicant as moving from a local address or from an out-of-state address. It was possible that out-of-state applicants might tend to honor the lease but this could not be tested because 58 of the 60 tenants in the data were local, so there was not enough variation in this variable to make it meaningful. This variable was therefore dropped and nine data points were used in the neural network. It is interesting to note that both of the tenants who moved in from out-of-state honored the lease for the full 12 months term. Two data points do not allow any conclusion to be drawn at it could be just coincidence. However, it is an indication that future testing of this type variable could be useful.

The neural network was run first using only the four variables of credit data. These variables were a) total loan balance, b) total monthly payment due, c) number of credit inquiries, and d) percentage of satisfactory accounts. There are more variables available from the credit file and some of these non-tested variables from the credit file could be predictive. However many of these other variables, such as oldest tradeline account, accounts that were delinquent but now current, and so forth are not clearly relevant. Testing these four variables alone provided a baseline for the study and allowed a comparison to the test results of the commercial credit scores. Three runs of the neural network software using these four variables produced R Square values of .2373, .2317, and .2800. While these R Square values are significantly better than those of the commercial credit scores, they are still very low indicating a low prediction value. In 100 runs using credit data, the neural network correctly predicted 46.4% of the tenants in the test data. This is an indication that credit data is an important component but not a strong predictor when used alone.

Since neural networks attempt to use all the input data, neural network performance can sometimes be improved by reducing the amount of input data and variables (Mozer & Smolensky, 1997). In an effort to identify the most important variables, all nine variables (four credit plus five lifestyle) were analyzed by running all the possible combinations through the neural network in a maximum size grouping of six (84 possible combinations). Six was the largest size grouping that could be tested each time because neural networks perform best when there are about 10 data points or more per variable (in this case 60 data points were available.) It is a limitation of the study imposed by the small size of the available data that larger variable combinations (i.e. seven, eight, or nine variable groupings) could not be tested. It is not known what effect this had on the research or if testing these larger groupings would or would not have improved the neural network performance. (However, the results indicate that the best predictive performance of the network occurred with three variables, so it is unlikely that adding more variables would have provided improved predictability.)

The first analysis ran all possible combinations of six variables from the original nine variables and identified the most important seven variables. The variables for criminal activity and total loan balance were of minor importance (Table 11) and these variables were dropped from the original nine variables leaving seven variables. It is not unexpected that the variable of total loans balance was a minor variable, since paying monthly rent is more of a cash flow issue for the tenant and is not directly impacted by total debt load. Although criminal history reports may provide an indication of the character of the tenant, criminal history as a variable was of minor importance for the neural network in predicting honoring of the lease. This conclusion is also supported by the raw data as 47% of the tenants with criminal activity did not honor the lease which is similar to (and slightly better than) the 55% without criminal history who did not honor the lease in the general tenant base. Criminal history is currently a key component of the decision process at six of the seven apartment complexes contacted for this research (Table 2). These remaining seven variables were used in 100 runs and created neural network models that correctly predicted 52.5% of the tenants in the test data. This prediction accuracy is a 5.9% percentage point improvement over the prediction accuracy using only credit data indicating that the addition of lifestyle data was improving the performance of the neural network.

The second analysis ran all possible combinations of five variables from the remaining seven variables (21 combinations) to identify the most important five variables. The variables for vehicle *Blue Book* value and for adult-adult/child were of lesser importance and these variables were dropped from the seven variables leaving five variables (Table 13). The original intent in including the variable for occupants of the apartments (i.e. adult adult-child variable) was that it was possible that multiple adults sharing an apartment, or an adult with child(ren) would tend to be less mobile and more likely to honor the lease. This variable was of lesser importance than expected. The effect on the neural network of *Blue Book* values of vehicles owned was unknown at the start of the research. The question revolved around whether higher value vehicles might imply a tenant better able to handle financial obligations, or conversely lower value vehicles may tend to be fully paid off thus allowing the tenant more monthly cash flow for rent payments. This is unanswered as this variable was of lesser importance. The remaining five variables with the highest importance were: a) tenant total income, b) tenant monthly payments, c) number of credit inquiries, d) percent of accounts that are satisfactory, and e) driving infraction record. In 100 runs, these five variables produced neural network models that correctly predicted 55.8% of the tenants in the test data, slightly better that in the previous analysis using seven variables.

Since five variables were identified and 60 data points were available, these five variables were tested in various combinations in order to identify a subset consisting of the most predictive variables. Five hundred runs of the neural network were used testing different combinations of these five variables in groupings of four, three, and two variables. The prediction accuracy of the neural network models improved as it was focused on the most important variables and performed best with a prediction accuracy of 69.1% when using only three of the variables: percent of satisfactory accounts, driving infractions, and total income. These three variables describe the tenant in the following way: percent satisfactory accounts indicate the tenant's tendency to pay their financial obligations reliably, total income indicated the tenant's cash available to pay off debt and pay rent, and the driving record could indicate a tenant's tendency to obey the law.

The apartment complexes contacted (Table 2) regularly use only one of these variables in their tenant decision process, that of tenant income (although a few also use percentage satisfactory accounts). Variables of this type are usually applied in the decision process using “if-then” type rules. For example, if tenant income is at least three times the monthly rent then the tenant is qualified to rent. (Those that also use percent satisfactory accounts may also have an “if…then” such as percent satisfactory account must be at least 70%). Two of the three variables (percent satisfactory accounts and driving record) found to be the most predictive represent potential new variables to add to the management decision process. Unfortunately, it will be difficult to incorporate driving records into the decision process (and to a lesser extent percent satisfactory accounts) in the usual way because this variable does not clearly fit into an “if-then” set of rules (if driving record is what? then do what?). This is particularly a problem since to the public there is not a clear connection between driving record and credit performance. This is the same problem faced by the auto insurance companies when they use credit report data to set auto insurance rates. However, in housing it is more important since federal regulations require an explanation for denial of housing (an auto insurance company could avoid a customer by just quoting a high price for the new auto policy). A neural network scoring solution is necessary in the apartment application process to order to incorporate all the elements into a decision in a consistent and defensible way.

Including the new variable of driving record into the decision process initially seemed the most unusual. However, as previously discussed, the auto insurance industry routinely makes use of credit/financial data to predict driving record (and thus set insurance rates.) This research may indicate that this process also works in reverse with driving records predicting credit/financial performance in the apartment rental market.

The best model performance correctly predicted the financial performance of 69.1% of the tenants. Although a desirable prediction range was not discussed with the apartment complex management, it is likely that this 69.1% would be considered too low to consider this model predictive. For example, the previously mentioned study that correlated auto insurance policies to credit scores had R Square correlations that exceeded 0.95 (Brockett et al., 2003). While R Square does not directly relate to prediction accuracy as discussed here, the high R Square values of a working auto insurance prediction model implies that a model must be statistically and substantially significant. However, since the apartment complex studied only accurately selects about 50% of their tenants, this 69.1% accuracy would still represent an improvement. The problem appeared to be caused by the small data set size of 60 data points. The range of R Square for this final model was 0.0771 to 0.7169 which was a wide range, and the predictive accuracy on each of the 100 runs varied widely from 36.4% correct to 90.9% correct. Essentially, as the model randomly picked training and test sets for each run, some of the sets provided better results and some much worse results. It is likely that the 60 data points included some data that was far outside the bounds of what would be considered "normal" (i.e. some bad data) and disturbed the model as it tried to fit all the data points. With a small data set, these bad data points could have had an unusually large effect on the model performance.

The R Square of the neural network models and the predictive accuracy in all the tests were significantly better than the commercially available credit scores and better than the neural network using only credit data. As the number of variables was reduced and as the variables used were focused on those that were most predictive, the performance of the neural network improved the most. This part of the analysis answered the research question: How is the prediction accuracy of a new neural network based credit scoring model improved by adding qualitative lifestyle data to the credit report data? In this series of tests, the answer was that the predictive accuracy of the neural network was greatly improved over the commercial credit scoring models through the addition of lifestyle data into the scoring process.

Limitations of this research

The data available only supported analysis of those applicants allowed to rent an apartment. There undoubtedly were applicants who were not approved for an apartment in this apartment complex, and presumably, these declined applicants would have gone on to rent an apartment somewhere else. No data is available to determine the eventual outcome of these initially declined applicants. This study therefore only analyzed the results of the applicants who received an initial positive approval and subsequently moved into the apartment complex. This is an example of a classic problem of “sample selection” and is a known problem in credit scoring (Greene, 1998). Essentially, the new model was constructed from a non-random sample, that is, only those applications that were accepted.

In this case, the ability to analyze the results of the declined applicants in addition to the accepted applicants would help determine the accuracy of the scope of the model. Specifically, was the new model selecting all the good applicants out of the potential pool of applicants or are some good applicants being declined at the apartment complex under study and then becoming good applicants at the next apartment complex. An analysis of this type would help determine if the screening of the new model was too tight, eliminating some good applicants. In general, since the market application of a model would be to screen applicants for entrance into an apartment complex, the limitation of using accepted applicants is not significant.

The small size of the data set (60 data points) limited the ability of the research to test larger combinations of variables such as seven, eight, or nine variables tested simultaneously. It is unknown what impact this had on the research. The small data set size also had the effect of making any bad data a higher percentage of all the data and thus more significant. Since the model results tended to have a wide range in the individual runs, it is likely that having more data available would have caused the range of the results to be reduced and this may have improved the predictiveness of the final model. Nonetheless, 60 tenants were a sufficient number as this represents about 50% to 60% of this apartment complex's yearly tenants (who moved in) and therefore should be a representative sample. Furthermore, Jensen (1992) developed a multilayer neural network for credit scoring and concluded that the neural network had good potential for credit decision and scoring applications with results developed on as few as 50 examples.

This research analyzed the effectiveness of nationally available credit risk scores, and developed new neural network models based on data collected from one apartment complex in one geographic area, specifically a southeastern U.S. city. However, would outcomes have changed and the final model been weighted differently, if the city had been located in the northwest U.S. instead of the southeast U.S? Since the credit scoring models used are national in scope, it is assumed that this impact is minimal on this research. However, some research has found that local economic factors show significant correlations with credit scores (Avery, Bostic, Calem, & Canner, 2000). The impact of local economic conditions is a concern when local banks and financial institutions use national credit scores. To address this concern, local banks and financial institutions usually adjust their procedures by changing the minimum acceptance levels for local conditions rather than trying to adjust a scoring model. It was not within the scope of this project to research the effect of geography on outcomes.

The data collected was from a single apartment complex and this apartment complex had a certain style and price range and attracted a certain type of tenant (specifically this complex was mostly blue collar, single person, or single parent with annual incomes in the $18,000 to $29,000 range). Other more expensive or less expensive apartment complexes, or those with more or fewer amenities would likely attract different types of tenants and using this data could result in a different model. It was not within the scope of this project to research this impact, if any. One should note that in the mortgage banking industry, there is only one model used for all applicants nationwide for home purchases (such as Equifax's "Beacon" score) regardless of the value of the home. Since one model is used in mortgage banking across all socio-economic levels, it is reasonable to assume therefore that one model should work across all socio-economic levels in the apartment rental industry. However, in order to develop a commercially viable neural network model for the apartment rental industry, a broader cross-section of socio economic data would need to be used, although this lack of broad cross-section was not a limitation in this research.

Implications and Recommendations

Credit scoring is widely used in a number of industries as an aid in helping managers make financial decisions concerning the loans and leases made to consumers. In general, these scores are considered (and in many cases proven) to be accurate predictors of consumer performance in meeting financial obligations. The purpose of this study was to a) analyze the results of six credit scores when used for rental decisions at an apartment complex and b) develop a new model that uses other data in addition to credit data to improve model predictability. This research indicated that the six commercially available credit scores are not predictive when applied to consumer behavior in renting an apartment. The apartment complex studied appeared to have a history of selecting many tenants who do not satisfy the terms of their lease despite using credit scoring. This lack of predictability of the credit score used at this apartment complex is in sharp contrast to its apparent success in the banking industry and in the auto insurance industry. Only two of the seven apartment complexes contacted for Table 2 use credit scoring and it is likely that the lack of predictability is the key reason. Not using credit scoring should continue to be the standard operating procedure. Although purchasing a credit score represents an insignificant expense, the two complexes currently using scoring are likely getting no value from this and these complexes can stop. This lack of predictability has forced the management to rely on other factors in making the accept/reject decision on each applicant, such as the ratio of gross income earned to monthly rental amount, payment history at other rentals, and other non-financial issues, such as size of family, reputation at other apartment complexes, and so forth. Until a more predictive model is available, this should continue as the standard operating procedure.

The new neural network models were developed using both credit data and other lifestyle and background data on each applicant. The additional data appeared to make the neural network models more predictive when the lifestyle data was added to the credit data. The model performance improved further when the number of variables was reduced and the model focused on only the variables determined to be most predictive. In this case, the most predictive variables were a mix of credit data and lifestyle data and it is likely that the best decision process in the apartment rental market would reflect this dual importance. The current decision process is based heavily on a) credit data, b) applicant income, and c) an applicant's criminal history with the criminal history usually used as a yes/no criteria (i.e. any criminal history causes an immediate denial of an apartment). This research showed that the criminal history variable has little importance on the financial consequences of renting. Therefore, the continued use of criminal history data in the rental decision process will need to be justified for other reasons than financial.

Unfortunately, while the predictive accuracy of the neural network model improved to 69.1%, this is still too low to provide a clear recommendation on the use of the additional variables in the decision process particularly, the variable of driving infractions. However, it is interesting to note that the apartment complex that provided the data currently has less than 50% of their tenants satisfy the lease. The use of this neural network (even at 69.1% accuracy) could help management to improve this low result. It is clear however, that adding lifestyle data improved the neural network prediction accuracy and this research raises a number of interesting questions that can be addressed in future research.

First, only six of the available 26 commercial credit scores available from Experian were tested. While these six were the ones most likely to be predictive in the apartment rental decision, it is possible that one or more of the non-tested scores could be predictive. These non-tested scores should be tested in a future study to confirm that commercial scores are not predictive.

Second, these commercial scores may not be predictive of a tenant honoring the lease possibly because of a fundamental human behavior difference in the apartment rental decision that makes this decision different from other financial decisions (particularly the home buying decision). Possible causes could be the perceived short term nature of an apartment rental, or a misunderstanding on the part of the tenant of the strength of the lease as a legal document (for example, renting a car is a legal transaction that incurs no penalty when returned early), or that the apartment decision is driven by lifestyle choices (children's school, loss of job, and so forth) rather than credit financial choices. In addition, applicants of different socio-economic backgrounds likely rent apartments for a wide variety of reasons and these reasons may vary by type and location of apartment complex. A better understanding of this underlying human behavior difference would further help to identify possible predictive variables.

Third, expand the data set size beyond the 60 data points used. While several authors in the literature indicate that working models can be created with as few as 50 data points, most commercial models are created using thousands of data points. Since the typical large apartment complex rents about 100 apartments per year and most of the records are in paper format, obtaining thousands of data points would represent a major data collection effort that would probably need to be funded by an industry group. On the other hand, having a model development program established that uses only a small data set (as was accomplished in this research) may allow researchers to tailor an individual model to the specific applicant profile at a specific apartment complex. This tailoring may be desirable from a commercial standpoint. First, however, the larger study should be completed to expand on the work of this research and clearly identify important variables. Additionally, the model developed in this research used data from one geographic region and one city. Expanding the data set to a wider geography would allow results to be proven more broadly. This would help determine if a national model can be built or if regional models are necessary. Most commercial credit scoring models are national due to the high expense of creating each model. However, it seems likely that models that are more granular could work best in predicting apartment rentals since the reasons for renting an apartment vary widely and apartment types available locally also vary widely.

Lastly, unlike expert systems, the neural network is unable to explain why a certain input is causing a certain output. This is an inherent part of the learning process of a neural network and unfortunately makes neural networks difficult to use in a consumer situation when the consumer must be told why the application was not approved. The best that can be accomplished is to identify the most important input variables as was done in this research. This is the technique currently used by the commercial credit scoring vendors. When a consumer buys or receives a score as part of a loan process, the consumer is given a listing of the most important variables that influenced the score. There is some experimentation beginning that combines neural networks with other "fuzzy logic" techniques to try to understand what is occurring inside the network. This could create a working model that has the predictive advantages of neural networks with the explanation advantages of expert systems. The type of model built in this research could be an ideal starting point for this type future research as the interactions and interplay of the qualitative data with the quantitative data may be explained using “fuzzy logic” techniques. Since federal law requires that a consumer have an explanation when turned down for credit (or housing), a tool capable of dealing with both quantitative and qualitative variables and their interrelations is needed (Khan, 2002) and important in creating a working commercial model. Until this happens, listing the most important variables used in the model appears to be a workable alternative.

Summary

The banking and financial services industry has used, for many years, credit report data and specifically, credit scoring, as a means of determining the credit worthiness of consumers applying for loans. The intent is to weed out, or at least identify those applicants that will become questionable accounts while, at the same time, offer lower interest rates and better products to those applicants that are most desirable. Credit risk evaluation decisions are important for the financial institution involved due to the high risk and potential financial cost associated with a wrong decision (Piramuthu, 1998).

The credit scoring process generates a credit score, which is a three-digit number that predicts the likelihood that an applicant will repay a loan and repay it on time. This score is based on the data in a consumer’s credit report and is the result of a process of modeling the variables important in the extension of credit. This modeling process is a statistical analysis of historical data for both good consumers and bad consumers, using certain financial variables such as (Leonard, 1996): a) number of bankruptcies, b) number of credit cards/trade line, c) length of employment, d) length of time at current address, and e) residential status. Today, credit scoring is used by 97% of banks that approve credit card applications and by virtually 100% of the banks that issue mortgage loans.

The success of credit scoring in the banking industry has caused it to spread to other industries, most notably the auto insurance industry. A recent survey by Conning and Company determined that more than 90% of the auto insurance carriers surveyed claimed to use credit data and credit scoring, such as the FICO credit score, in their new business process for automobile coverage (Jones, 2001). This credit scoring is part of the process in determining who will get auto insurance and at what price the auto policy will be issued.

This research analyzed the effectiveness of credit scoring when applied to the decision process for selecting tenants for apartment rental. The first phase of this research analyzed the results of using six commercially available credit scores applied in one apartment complex to the task of selecting applicants. Six commercially available credits scoring models were tested against the results of renting apartments. The results indicated that these six models were not effective in predicting the financial performance of the tenant in honoring the apartment lease. These scores are based on credit data alone. This part of the analysis answered the research question: How effective are commercially available credit scores in predicting applicant financial behavior when renting an apartment? In this testing, the six commercial credit scoring models were found not to be predictive.

It was not a part of this research to identify the reasons that existing scores are not predictive. However, the problem with predictability of the six models may be based on the composition of the credit scoring model (statistical modeling issues). The reasoning is that these models were developed for other purposes such as home ownership and tend to filter out the typical tenants for apartment rentals (i.e. younger in age, less time on the job, lower paying job, and so forth). In essence, the weighting of the variables used in the model creation is targeted to answer or predict a different consumer behavior. Additionally, the problem with predictability of the six models tested seems to be centered on the use of credit data alone for the apartment rental application (data issues). The basic assumption with a credit data score is that there is a correlation between credit score and financial risk with an improving score indicating an improving financial risk. This correlation has been proven in the banking, credit card, and auto insurance industry but may not exist when picking applicants for the apartment rental market. The reason may be differences in human behavior in transacting an apartment lease versus for other financial transactions such as house purchases. It is reasonable to speculate that the applicant views an apartment rental as a short-term decision, similar to renting a car, while the existing credit scoring models predict behavior for longer-term decisions such as buying a house. Thirdly, the decision to abandon an apartment before the lease is completed may be a decision likely based not entirely on credit matters, but instead based heavily on lifestyle issues (children in school, change of job, and so forth.)

Phase two of this research used neural networks to develop a new model using both credit data and other available lifestyle data about the tenant. The hypothesis was that the addition of this lifestyle data into the new neural network based model would make the new model more accurate in selecting apartment rental applicants than commercial credit scoring based only on credit data. Neural networks were used because these knowledge discovery tools are well-suited for discovering the non-obvious relationships in data (Bigus, 1996; Marakas, 1999). Additionally, in this problem of predicting applicant behavior, it was unknown how, or even which tenant characteristics (independent variable input) affect the predicted output of lease honored or not honored (dependent variable). Furthermore, it was also unknown how inputs were related to each other and thus affected output in combination. Neural networks have been shown to perform reasonably well in this type of complex and unstructured problem (Piramuthu, 1998).

In order to create a model that was more predictive, expanded data was collected on tenants that included both credit data and additional data on lifestyle. Seventy-six data points were collected on each of 60 tenants and this data was simplified into 10 variables for use by a neural network to create a new model. The variables used were:

1. State of previous residence. (The managers thought that out-of-state tenants would have a higher tendency to honor the lease.)

2. Adult only, multiple adults or adult with children. (Multiple adults or adults with children would be less mobile and have a higher tendency to stay.)

3. Total applicant income. (cash available to pay debt)

4. Total *Blue Book* value of all vehicles. (High value vehicles would imply a tenant better able to handle financial obligations or conversely low value vehicles would be fully paid off thus freeing up cash for rent payments.)

5. Number of driving infractions. (background information)

6. Applicant has criminal background. (background information)

7. Total loan balance. (credit data- indication of debt load)

8. Total monthly payments. (credit data – an indication of other cash needs of tenant beside monthly rent)

9. Total credit file inquiries. (credit data – A high number of inquiries implies a tenant looking hard for credit, possibly due to financial problems not yet apparent.)

10. Percentage of total accounts that are satisfactory. (credit data - indication of tenant's tendency to reliably pay debts on time)

One of the 10 variables (i.e. state of previous residence) was dropped because it lacked enough variation to be relevant, thus leaving nine useful variables. In order to test these variables (four credit, plus five lifestyle), identify important variables, and possibly reduce the number of variables, the exhaustive search technique was used (Dwinnell, 1998). Hundreds of runs of the neural network software were used to test all combinations of these nine variables. In each run, the importance of each input variable was recorded and five variables eventually emerged as most important. These were total income, driving infractions, total monthly payment, credit inquiries, and percent satisfactory accounts. The accuracy of these variables in predicting tenant honoring or not honoring the lease was then determined through several hundred additional runs of the neural network software. The predictive ability of the model improved as the number of variables was reduced and as the variables used were focused on those that were most important as listed in Table 19.

Table 19 Prediction accuracy of neural networks with varying input variables

Variables Tested	Percent of tenants correctly predicted by neural network (Test Data)
4 credit variables only (baseline) loan balance, total monthly payment, credit inquiries, percent sat. accounts.	46.4 %
7 variables of 3 credit plus 4 lifestyle adult/child, total income, vehicle *Blue Book*, driving, total monthly payment, credit inquiries, percent sat. accounts	52.5 %
5 variables of 3 credit and 2 lifestyle total income, driving, total monthly payment, credit inquiries, percent sat. accounts	55.8 %
4 variables of 2 credit and 2 lifestyle total income, driving, total monthly payment, percent sat. accounts	58.4 %
3 variables of 2 credit and 1 lifestyle driving, total monthly payment, percent sat. accounts	48.4 %
3 variables of 1 credit and 2 lifestyle total income, driving, percent sat. accounts	69.1 %
2 variables of 1 credit and 1 lifestyle driving, percent sat. accounts	41.4 %

Three variables were found to be most predictive for the apartment rental decision and these were a) percentage of satisfactory accounts, b) total tenant income, and c) driving record. The apartment complexes contacted currently only regularly use one of these variables in their tenant decision process, that of tenant income (although a few also use percentage satisfactory accounts). Variables of this type are usually applied in the decision process using “if-then” type rules. For example, if tenant income is at least three times the monthly rent then the tenant is qualified to rent. (Those that also use percent satisfactory accounts may also have an “if…then” such as percent satisfactory account must be at least 70%). The other two variables found to be important (i.e. monthly payments and number of credit inquiries) represent potential new variables to add to the management decision process. The model with the highest prediction accuracy used the variable of driving record. Unfortunately, it will be difficult to incorporate driving record (or either of the other two variables if used) into the decision process in the usual way because these variables do not clearly fit into an “if…then” set of rules (if driving record is what? then do what?). This is particularly a problem since to the public there is not a clear connection between driving record and credit performance. This is the same problem faced by the auto insurance companies when they use credit report data to set auto insurance rates. However, in housing it is more important since federal regulations require an explanation for denial of housing (an auto insurance company can avoid a customer by just quoting too much for the new policy). A neural network scoring solution is necessary in the apartment application process to order to incorporate all the elements into a decision in a consistent and defensible way.

This part of the analysis answered the research question: How is the prediction accuracy of a new neural network based credit scoring model improved by adding qualitative lifestyle data to the credit report data? In this series of tests, the answer was that the predictive accuracy of the neural network was greatly improved over the commercial credit scoring models, although the prediction accuracy did not reach a high enough value to be definitive.

Future research is suggested in four areas. First, only six of the available commercial credit scoring models were tested in this research. While these six represented the models most likely to be predictive, it is possible that one or more of the non-tested models may be predictive and these should be tested. Second, these commercial scores may not be predictive of a tenant honoring the lease possibly because of a fundamental human behavior difference in the apartment rental decision that makes this decision different from other financial decisions (particularly the home buying decision). Possible causes could be the perceived short-term nature of an apartment rental, or a misunderstanding on the part of the tenant of the strength of the lease as a legal document, or that the apartment decision is driven by lifestyle choices (children's school, loss of job and so forth) rather than credit financial choices. A better understanding of this underlying human behavior difference would further help to identify possible predictive variables. Third, while several authors in the literature indicate that working models can be created with as few as 50 data points, expanding the size of the data set beyond the 60 data points used in this research would enable more analysis and possibly a better understanding of the interrelations among the important input variables. Many commercial models are developed with data set sizes of several thousand data points. Lastly, unlike expert systems, the neural network is unable to explain why a certain input is causing a certain output. There is some experimentation beginning that combines neural networks with other "fuzzy logic" techniques to try to understand what is occurring inside the network. Since federal law requires that a consumer have an explanation when turned down for credit or housing, a tool capable of dealing with both quantitative and qualitative variables and explaining their interrelations is needed (Khan, 2002).

Appendix A

Typical Pricing of Additional Consumer Data

Appendix A

Typical Pricing of Additional Consumer Data from Experian Pricing Manual

CONSUMER & COMMERCIAL BUSINESS CREDIT / EMPLOYMENT SCREENING / VEHICLE HISTORY REPORTS			
Consumer		**Commercial**	
Rates are based on transactions per month			
100,000 or more transactions	$0.08	2,001 or more transactions	$0.25
50,001 - 100,000	0.09	1,001 – 2,000	0.32
20,001 - 50,000	0.10	251 - 1,000	0.37
10,001 - 20,000	0.13	51 – 250	0.42
5,001 - 10,000	0.15	1 – 50	0.49
1,001 - 5,000	0.17		
251 - 1,000	0.20		
1 - 250	0.25		

Additional Data Sources
Online Database Reports

Product	Cost	Court Fees	Coverage	Billable HIT	Billable NO HIT
People Search	$0.99	-	**50 states**	Yes	
Trace Detail	1.99	-	**50 states**	Yes	
Business Search	$1.99	-	**50 states**	Yes	
Reverse Phone Search	$0.15	-	**50 states**	Yes	
Evictions Report*	$3.99 NO HIT $0.99	-	**50 states**	Yes	Yes
Criminal National Search*	3.99	-	**41 states**	Yes	Yes
Criminal State Search*	3.99	-	**41 states**	Yes	Yes
NBD COPS Plus National Criminal Report*	7.00	-	**38 states**	Yes	Yes
NBD COPS Plus State Criminal Report*	6.00	-	**38 states**	Yes	Yes
Sex Offender*	7.00	-	**29 states**	Yes	Yes

Bankruptcy Search*	0.99	-	**50 states**	Yes	Yes
Property Search	3.99	-	**44 states**	Yes	
UCC Search	3.99	-	**48 states**	Yes	
Motor Vehicle Search	2.99	-	**16 states**	Yes	
Driver's License Search*	0.99	-	**16 states**	Yes	

Manual Reports

Product	Cost	Court Fees	Coverage	Billable HIT	Billable NO HIT
State Criminal plus state fees	$7.00	Yes	**38 states** Click here for coverage	Yes	Yes
State DOC plus state fees if any	$10.00	-	**43 states** Click here for coverage	Yes	Yes
County Criminal plus county fees	$10.00	Yes	**50 states** Click here for coverage	Yes	Yes
County Civil plus county fees	$14.00	Yes	**50 states** Click here for coverage	Yes	Yes
DMV Driving Records plus state fees	**$3.00**	**Yes**	**50 states**	**Yes**	**Yes**

Appendix B

2004 Enacted Identity Theft Legislation

Appendix B

2004 Enacted Identity Theft Legislation from National Conference of State Legislatures

from **http://www.ncsl.org/programs/lis/privacy/idt-01legis.htm**

States without a listing have no legislation as of this report

Status as of February 7, 2005

State:	**Bill Summary:**
Arizona	H.B. 2116 Signed by governor 4/19/04, Chapter 109 States that a person commits criminal possession of a forgery device if the person makes or possesses any material, good, property or supply designed or adapted for use in forging written instruments or with the intent to aid or permit another person to use it for the purpose of forgery. Expands the definition of taking the identity of another person to include purchasing, manufacturing, recording, or transmitting any personal identifying information to include entities and real or fictitious persons/entities. Requires a peace officer to take a report on the request of any person or entity whose identity has been taken. Allows prosecutors to file a complaint charging multiple identity theft violations in the county where the greatest number of violations is alleged to have occurred. States that it is unlawful for a person to intentionally or knowingly make or possess with the intent to commit fraud anything specifically designed or adapted for use as a scanning device or re-encoder. Adds to the definition of personal identifying information any written document or electronic data that provides information concerning a signature, electronic mail address or account, tax identification number, employment information, citizenship status, alien identification number, personal identification number, photograph, DNA or genetic information or other financial account number. Clarifies that beginning on January 1, 2005, it is illegal for a person or entity to print a number that is known to be an individual's Social Security number. States that if a number is received from a third party, there is no duty to determine if the number is an individual's Social Security number. The number may be printed on materials mailed to the individual, unless the person or entity mailing the number knows that it is the individual's Social Security number. States that beginning on January 1, 2009, no person or entity may knowingly print any sequence of numbers contained in an individual's Social Security numbers on any card required for the person to receive services, products, or materials that are mailed to the individual.

Colorado	H.B. 1134 Signed by governor 6/4/04, Chapter 365 Creates the Motor Vehicle Investigations Unit in the Department of Revenue to investigate and prevent the fraudulent issuance and use of driver's licenses, identification cards, motor vehicle titles and registrations, and other motor vehicle documents, and to assist victims of identity theft. Authorizes a criminal who wrongfully uses another's identify to be charged in the jurisdiction where a government agency issued identity documents. Sets standards and procedures for a court to determine that a victim's identity has been mistakenly associated with a crime.
	H.B. 1274 Signed by governor 4/26/04, Chapter 205 Requires a creditor or charge card company that offers credit or a charge card by mail, and that receives an acceptance of an offer that lists an address for the applicant that is different from the address where the offer of credit or a charge card was sent, to verify that the person accepting the offer is the person to whom the creditor or charge card company made the offer of credit or a charge card. Allows for a private right of action against a person who uses the personal identifying information of another to commit fraud-type crimes.
	H.B. 1311 Signed by governor 6/4/04, Chapter 393 Prohibits the display of a person's Social Security number on a license, pass, or certificate, issued by a public entity, unless it is necessary to further the purpose of the pass or required by state or federal law. Proscribes a public entity from requesting a person's Social Security number over the phone, via the Internet, or by mail unless federal law requires it or is essential to the public entity's service. Requires public and private entities to develop a policy for disposal of documents containing personal identifying information. Considers a public entity that is compliant with the state archives act to have met its policy development obligation. Exempts trash haulers from having to verify that documents have been destroyed or properly disposed. Allows an insured to require that an insurance company not display the insured's Social Security number on the insured's insurance identification card or proof of insurance card. Requires the insurer to reissue the card without the Social Security number, if the insured makes the request. Prohibits an insurance company, after January 1, 2006, from issuing an insurance identification card or proof of insurance card displaying the insured's Social Security number. Makes it a class 1 misdemeanor to possess another's personal identifying information with the intent to use the information, or to aid or permit another to use the information, to gain unlawfully a benefit or to injure or defraud another.
Connecticut	H.B. 5184 Signed by governor 5/21/04, Public Act 04-119 Concerns the nondisclosure of private tenant information in a sale of public

	housing to a private entity, including the tenant's Social Security number and bank account number.
Delaware	S.B. 233 Signed by governor 5/25/04, Chapter 248 Makes illegal (Class D felony) the possession with intent to defraud or the use with intent to defraud certain devices that facilitate the stealing and/or illegal use of credit card information.
Georgia	H.B. 656 Signed by governor 5/5/04, Act 451 Relates to unfair or deceptive practices in consumer transactions, so as to require that credit card issuers take steps to verify a consumer's change of address when a person responds by mail to an unsolicited application for credit and provides an address that is different from the address to which such solicitation was mailed.
Hawaii	H.B. 2674 Signed by governor 5/28/04, Act 92 Exempts disclosure of Social Security numbers from government payroll records that are public information; restricts retail merchant card issuers from requesting personal information except for credit purposes and from sharing cardholder information.
Indiana	H.B. 1197 Signed by governor 3/18/04, Public Law 43 Expands the class of criminal cases in which an individual's statement or videotape may be admissible to include certain crimes committed against an individual who is at least 18 years of age and considered a protected person because of the individual's incapacity to manage or direct the management of the individual's property or to provide or direct the provision of the individual's self care. Provides that a statement or videotape made by the protected person is admissible in certain criminal trials if: (1) the statement or videotape is reliable; and (2) the individual either testifies at trial or is unavailable.
Louisiana	H.B. 623 Signed by governor 7/6/04, Act 766 Provides for the imposition of a security freeze, by the consumer, on his credit report or score. Also provides for the methods of access after placement of a freeze and removal.
Maryland	H.B. 457 Vetoed by governor - cross-filed bill signed 5/26/04 S.B. 257 Signed by governor 4/27/04, Chapter 109 Authorizes a state's attorney or the attorney general to investigate and prosecute offenses relating to personal identifying information fraud; authorizes the attorney general to exercise all the powers and duties of a state's attorney to investigate and prosecute specified violations; and

	establishes that a prosecution for a violation of specified offenses relating to personal identifying information fraud or other crimes based on a violation may be commenced in a county in which an element of the crime occurred or in which the victim resides.
	H.B. 926 Vetoed by governor - cross-filed bill signed 5/27/04 S.B. 513 Signed by governor 4/27/04, Chapter 130 Establishes determinations as to the value of property or services involving specified theft crimes; establishes penalties for theft of property or services with a value of less than $100; establishes that action or prosecution for specified crimes must be commenced within two years.
Mississippi	S.B. 2957 Signed by governor 5/6/04, Chapter 526 Provides a lesser penalty for identity theft in cases involving a lesser amount of money, provides for aggregation of amounts in determining the amount of an offense, authorizes the attorney general to provide assistance to victims of identity theft in clearing their records, and clarifies that perpetrators of identity theft shall pay restitution to their victims; clarifies jurisdiction of offenses occurring in multiple jurisdictions; allows certain funds to be used for the purpose of consumer fraud education; authorizes a victim of identity theft to expunge his record of false charges accrued on account of activities of the perpetrator; authorizes the attorney general to issue "identity theft passports" under certain circumstances; defines identity theft; grants subpoena power to the attorney general in conducting investigations of identity theft; requires aggregation of amounts stolen from the same victim in determining the gravity of the offense of larceny.
Missouri	H.B. 916 Signed by governor 5/10/04 Makes it a class A misdemeanor when the identity theft results in the theft or appropriation of credit, money, goods, services, or other property valued at less than $500. Makes attempted identity theft a class B misdemeanor. Makes identity theft a class D felony when the value of the stolen property is more than $500 but does not exceed $1,000. Makes identity theft a class C felony when the value of the stolen property is more than $1,000 but does not exceed $10,000. Makes identity theft a class B felony when the value of the stolen property is more than $10,000 but does not exceed $100,000. Makes identity theft a class A felony when the value of the stolen property exceeds $100,000. Makes identity theft a class A felony when the identity theft is performed for committing a terrorist act. Makes identity theft a class C felony when the identity theft is performed for committing an election offense. Makes the identity thief liable to the victim for civil damages of up to $5,000 per incident or three times the amount of actual damages, whichever is greater. Allows the victim to seek a court order restraining the identity thief from future acts that would constitute identity theft. In these

	actions, the court may award reasonable attorney fees to the plaintiff. Clarifies that the estate of a deceased person may pursue civil remedies when the estate is a victim of identity theft. Sets a limitation on civil suits at five years and clarifies that a criminal conviction is not a prerequisite for a civil claim. Clarifies that identity theft does not include a minor's misrepresentation of age by using an adult person's identification. Clarifies that a criminal prosecution for identity theft may be conducted in any county where a victim or defendant resides, where the stolen property was located, or in any county where an element of the crime was committed. Makes a second offense of identity theft or attempted identity theft a class D felony when the value of the property is less than $500. Creates the crime of trafficking in stolen identities, a class B felony. The crime is committed when a person possesses or transfers any means of identification for committing identity theft. Unauthorized possession of a means of identification for five persons will be evidence of such intent. Expands the crime of false impersonation to include the providing of a false identity to a law enforcement officer upon arrest. If the false identity is not discovered until after the person is convicted, the prosecutor must file a motion to correct the arrest records and court records. Allows the court to order the expungement of the false arrest records for the person whose identity was used.
	H.B. 959 Signed by governor 6/14/04 Makes it a class A misdemeanor when the identity theft results in the theft or appropriation of credit, money, goods, services, or other property valued at less than $500. Makes attempted identity theft a class B misdemeanor. Makes identity theft a class D felony when the value of the stolen property is more than $500 but does not exceed $1,000. Makes identity theft a class C felony when the value of the stolen property is more than $1,000 but does not exceed $10,000. Makes identity theft a class B felony when the value of the stolen property is more than $10,000 but does not exceed $100,000. Makes identity theft a class A felony when the value of the stolen property exceeds $100,000. Makes identity theft a class A felony when the identity theft is performed for committing a terrorist act. Makes identity theft a class C felony when the identity theft is performed for committing an election offense. Makes the identity thief liable to the victim for civil damages of up to $5,000 per incident or three times the amount of actual damages, whichever is greater. Venue in this type of civil suit is proper in any county where any of the property stolen was located, where the defendant or victim resides, or in any county in which an element of a criminal charge of identity theft was committed. Allows the victim to seek a court order restraining the identity thief from future acts that would constitute identity theft. In these actions, the court may award reasonable attorney fees to the plaintiff. Clarifies that the estate of a deceased person may pursue civil remedies when the estate is a victim of identity theft. Establishes a limitation on civil suits at five years and clarifies that a criminal conviction

	is not a prerequisite for a civil claim. Clarifies that identity theft does not include a minor's misrepresentation of age by using an adult person's identification. Clarifies that a criminal prosecution for identity theft may be conducted in any county where a victim or defendant resides, where the stolen property was located, or in any county where an element of the crime was committed. Makes a second offense of identity theft or attempted identity theft a class D felony when the value of the property is less than $500. Creates the crime of trafficking in stolen identities, a class B felony, and is committed when a person possesses or transfers any means of identification for committing identity theft. Unauthorized possession of a means of identification for five persons will be evidence of the intent.
New Hampshire	S.B. 521 Signed by governor 6/11/04, Chapter 233 Increases the penalty for identity fraud to a class A felony in all cases.
Oklahoma	S.B. 1164 Signed by governor 6/3/04 Authorizes expungement of certain records related to crimes arising from identity theft, creates the Oklahoma Identity Theft Passport Program.
	S.B. 1168 Signed by governor 5/14/04, Chapter 279 Modifies the crime of identity theft.
	S.B. 1503 Signed by governor 5/12/04 Prohibits false or fraudulent statements to financial institutions to obtain certain information; prohibits false or fraudulent documents or documents without lawful authority to obtain certain information or to commit a crime; states penalty; and provides for restitution.
Tennessee	H.B. 3403 Signed by governor 6/8/04, Public Chapter 911 S.B. 3364 Creates Class C felony offense of identity theft trafficking; declares that victim of identity theft is also a crime victim; establishes method for law enforcement to obtain records from public or private entity in cases of identity theft; and establishes standards for destruction of records maintained by private entity that contains personal identifying information concerning a client.
Utah	H.B. 195 Signed by governor 3/15/04, Session Law Chapter 55 Deletes provisions that currently give the Division of Consumer Protection authority to regulate the misuse of personal identifying information.
	S.B. 16 Signed by governor 3/22/04, Session Law Chapter 227 Establishes that the residence of the victim of identity theft in this state is sufficient to establish jurisdiction in this state; permits the prosecution of an

	identity theft in the county where the identity was stolen or used, or where the victim resides; allows prosecution in any county where the identity was stolen, used, or where the victim resides when the offense occurs in multiple jurisdictions; and establishes that the unauthorized possession of another person's identifying documents is a crime.
Vermont	H.B. 327 Signed by governor 6/8/04, Act 155 Allows a consumer to request that a credit reporting agency place a security alert on the consumer's credit report if the consumer's identity might have been used to fraudulently obtain goods or services and to place a security freeze on the credit report if the consumer has a sworn complaint about the unlawful use of personal information. The consumer credit reporting agency would have to provide a written summary of the rights of the consumer. Establishes the crime of identity theft and penalties for violations.
Virginia	H.B. 872 Signed by governor 4/12/04, Chapter 450 Authorizes the attorney general, with the concurrence of the attorney for the Commonwealth, to assist in the prosecution of the crimes of identity theft (§18.2-186.3) and the use of a person's identity with the intent to intimidate, coerce, or harass (§18.2-186.4). Allows for a conviction under the identity theft statutes when the defendant uses a false or fictitious name. Requires DMV, upon notification from the attorney general that an Identity Theft Passport has been issued to a driver, to note the same on the driver's abstract. Directs child day programs that reproduce or retain documents of a child's proof of identity that are required upon the child's enrollment into the program to destroy them upon the conclusion of the requisite period of retention. The procedures for the disposal, physical destruction or other disposition of the proof of identity containing Social Security numbers shall include all reasonable steps to destroy such documents by (a) shredding, (b) erasing, or (c) otherwise modifying the Social Security numbers in those records to make them unreadable or indecipherable by any means.
West Virginia	H.B. 4104 Signed by governor 3/25/04, Chapter 79 Relates to creating the crimes of scanning device and re-encoder fraud; provides definitions; and establishes criminal penalties therefore.

Appendix C

Applicant Variables Collected before Combining Variables

Appendix C

Applicant Variables Collected before Combining Variables

Applicant Variables Collected before Combining Variables	Source Of Information	Comments
1. How did you hear about us	Welcome card	1=Apt locator 2=Referral 3=Newspaper ad 4=Sign 5=Brochure 6=Apt guide magazine 7=Yellow pages 8=other
2. Size apt preferred	Welcome card	0=Studio 1=One bedroom 2=Two bedroom 3=Three bedroom
3. Prepared to put down a deposit today	Welcome card	0=no 1=yes
4. Occupation	Welcome card	
5. Number of applicants	Application	1=one person as primary only 2=primary plus other applicants
6. Marital status of primary applicant	Application	0=single 1=married 2=other
7. Present address	Application	0=local 1=not local
8. Present landlord	Application	0=live with parents or none 1=apt in private home 2=other apt complex 3=moving from own home
9. Current monthly rent	Application	
10. Employer name	Application	
11. Size of employer	Chamber of Commerce	0=small 1=medium 2=large
12. Number of years with employer	Application	
13. Type of employer	Application	0=retail 1=restaurant 2=manufacturing 3=medical

			4=service 5=other
14.	Level of employee	Application	0=employee 1=manager
15.	Income	Application	
16.			
17.	Additional Income Amount	Application	
18.	Bankruptcy	Application	0=no 1=yes
19.	Nearest relative	Application	0=none 1=parent 2=spouse 3=relative 4=other
20.	Number of people to occupy apartment	Application	
21.	Relationship of 1st additional person occupying apartment	Application	0=child 1=spouse 2=roommate 3=other
22.	Relationship of 2nd additional person occupying apartment	Application	0=child 1=spouse 2=roommate 3=other
23.	Relationship of 3rd person additional occupying apartment	Application	0=child 1=spouse 2=roommate 3=other
24.	Age of applicant	Driver's license	
25.	Age of 1st additional person occupying apartment	Application	
26.	Age of 2nd additional person occupying apartment	Application	
27.	Age of 3rd additional person occupying apartment	Application	
28.	Gender of applicant	Application	0=female 1=male
29.	Gender of 1st additional person occupying apartment	Application	0=female 1=male
30.	Gender of 2nd additional person occupying apartment	Application	0=female 1=male
31.	Gender of 3rd additional person occupying apartment	Application	0=female 1=male
32.	Number of vehicles to be parked at apartment	Application	

33. Make of vehicle one	Application	
34. Model of vehicle one	Application	
35. Age of vehicle one	Application	
36. Make of vehicle two	Application	
37. Model of vehicle two	Application	
38. Age of vehicle two	Application	
39. Number of days from welcome card visit to actually applying for an apartment	Application date and welcome card date	
40. Number of IDs provided	ID card	
41. Type of ID one	ID card	0=driver license 1=state ID card 2=social security 3=passport 4=non U.S. ID card 5=other
42. Type of ID two	ID card	0=driver license 1=state ID card 2=social security 3=passport 4=non U.S. ID card 5=other
43. Number of public records	Credit Report	
44. Installment loan balance	Credit Report	
45. Real estate loan balance	Credit Report	
46. Total revolving loan balance	Credit Report	
47. Past due loan amount	Credit Report	
48. Estimated monthly installment loan payments	Credit Report	
49. Estimated real estate loan payments	Credit Report	
50. Total revolving loan available	Credit Report	
51. Number of inquiries to credit file in last 30 days	Credit Report	
52. Number of inquiries to credit file last 6 months	Credit Report	
53. Number of tradeline accounts	Credit Report	
54. Number of paid accounts	Credit Report	
55. Number of satisfactory accounts	Credit Report	
56. Number of delinquent accounts now	Credit Report	
57. Number of delinquent	Credit Report	

accounts in past 6 months		
58. Date of oldest trade account	Credit Report	
59. Number of different social security numbers on credit file	Credit Report	
60. FICO risk score 2	Credit Report	
61. FICO installment loan score	Credit Report	
62. FICO advanced risk score	Credit Report	
63. FICO finance score	Credit Report	
64. Sureview score	Credit Report	
65. Number of closed bank accounts in last year	Consumer Debit report	
66. Real estate owned	Property Search Report	0=none 1=commercial 2=residential
67. Value of real estate owned	Property Search Report	
68. Number of driving infractions	DMV report	
69. Type of 1st driving infraction	DMV report	1=speeding 2=DUI 3=moving violation 4=other
70. Type of 2nd driving infractions	DMV report	1=speeding 2=DUI 3=moving violation 4=other
71. Information found on national felony search	National criminal database	0=no 1=yes
72. Type of information in national database	National criminal database	0=drug 1=violence 2=other
73. Information found on county criminal search	County criminal search	0=no 1=yes
74. Type of information in county search	County criminal search	0=drug 1=violence 2=other
75. Information found on state criminal search	State criminal search	0=no 1=yes
76. Type of information in state database	State criminal search	0=drug 1=violence 2=other

Appendix D

Variables used in the Model

Appendix D

Variables used in the Model

Original Variables collected	**Original Variables Combined into Variables for Model**	**Ten Variables Actually Used in Neural Network Model (yes/no)**
• Present address • Present landlord	• State of previous address	• Yes (1)
• Level of employee	• Not used (data likely inconsistent: manager of pizza place different from manager of major company)	• No
• Name of employer	• Not used (cannot be quantified for use in model)	• No
• Income all applicants • Additional income all applicants	• Total applicant(s) income	• Yes (2)
• Number of people to occupy apartment • Relationship of 1st person to applicant • Relationship of 2nd person to applicant • Relationship of 3rd person to applicant • Age of 1st additional person in apartment • Age of 2nd additional person in apartment • Age of 3rd additional person in apartment • Gender of 1st additional person in apartment • Gender of 2nd additional person in apartment • Gender of 3rd additional person in apartment	• Adult only or adult with children 1= adult only 2= multiple adults 3= adult with children	• Yes (3)
• Number of vehicles to be parked at apartment • Make of vehicle one	• Total *Blue Book* value of all vehicles	• Yes (4)

• Make of vehicle two • Model of vehicle one • Model of vehicle two • Age of vehicle one • Age of vehicle two		
• Value of real estate owned • Real estate owned	• Not used (likely all data to be zero as these applicants are renting apartments, that is they do not own homes)	• No
• Number of driving infractions • Type of 1st driving infraction • Type of 2nd driving infraction	• Driving infraction yes=1/no=0	• Yes (5)
• Information on national criminal search • Type of information on national criminal search • Information on state criminal search • Type of information on state criminal search • Information on county criminal search • Type of information on county criminal search	• Applicant has criminal information yes=1/no=0	• Yes (6)
• Installment loan balance • Revolving loan balance	• Total loan balance	• Yes (7)
• Estimated monthly loan payment • Estimated monthly revolving loan payment	• Total monthly payment	• Yes (8)
• Number of inquiries to credit file past 30 days • Number of inquiries to credit file past 180 days	• Total credit file inquiries	• Yes (9)
• Number of tradeline accounts • Number of paid accounts • Number of satisfactory accounts • Number of delinquent accounts now • Number of delinquent accounts past 6 months	• Percentage of satisfactory accounts	• Yes (10)
• FICO risk score 2 • FICO installment loan score	• Not used since all scores tested are non predictive	• No

• FICO advanced risk score • FICO finance score • Sureview score		
• Number of Ids provided • Type of ID one • Type of ID two	• Not used; probably not helpful as most applicant's will use driving license	• No

Appendix E

Commercially Available Neural Network Software

Appendix E

Neural Network Software Generally Available in June 2005

Commercial Software

- TNs2Server
- CATPACK
- PolyAnalyst
- ECANSE - Environment for Computer Aided Neural Software Engineering
- DataEngine
- KnowMan Basic Suite
- Matlab: Neural Network Toolbox
- NeuroForecaster/GENETICA
- N-Net
- VBBackProp
- FCM (Fuzzy Control Manager)
- NeuroShell
- Neurogon
- Partek
- Domain Solutions' Neural Networks for Developers
- Neural Net Tutor
- Neural Parts
- Propagator
- Clementine
- FlexTools
- Neuframe
- BrainMaker
- Owl Neural Network
- NeuroLution simulation and development system
- Neural Bench
- Adaptive Logic Network
- NeuroLab
- Trajan
- Model 1
- Pattern Recognition Workbench - PRW
- NNMODEL
- NeuroModel®
- Neural Connection
- EXPO/NeuralNet
- Braincel
- NeuroSolutions
- NeuroGenetic Optimizer
- Saxon

- havFmNet++
- Attrasoft Boltzmann Machine (ABM)
- Thinks and Thinks Pro
- STATISTICA: Neural Networks
- SAS: Neural Network Add-On
- Attrasoft Predictor
- DataMining Workstation (DWM) and DWM/Marksman
- MacBrain
- BioNet Simulator
- Nestor Development System
- Neural Network Utility/2
- NeuralWorks
- Viscovery SOMine
- KnowMan Basic Suite
- WinBrain
- Process Insights
- havBpNet:J
- DynaMind Developer Pro
- havBpNet++
- NeuroClassifier
- NeuroWindows
- BrainSheet for Win95
- PathFinder
- PREVIA
- **Forecaster XL (used in this research)**
- NeuroCoM (Neuro Control Manager)

Freeware

- Net II
- SpiderWeb Neural Network Library
- tlearn
- NeuDL
- Mactivation
- Pittnet
- Binary Hopfield Net with free Java source
- NeuralShell
- PlaNet
- Valentino Computational Neuroscience Workbench
- Neural Simulation Language Version - NSL
- Neocognitron
- SOM Toolbox for Matlab
- Fuzzy ARTmap

- Xerion Simulator
- Rochester Connectionist Simulator (RCS)
- Aspirin/Migraines
- QwikNet
- PDP++ Software
- UCLA-SFINX
- FuNeGen
- Cascade Correlation Simulator
- SynWorks
- LVQ PAK
- Hyperplane Animator
- VFSR - Very Fast Simulated Reannealing
- Brain Neural Network Simulator
- SESAME - Software Environment for the Simulation of Adaptive Modular Systems
- NNCTRL
- Pygmalion
- NICO Artificial Neural Network Toolkit
- SOM PAK
- Multi-Module Neural Computing Environment - MUME
- FastICA
- Con-x
- NNSYSID
- PDP Software
- nn/xnn
- Roxanne
- Matrix Backpropagation
- NevProp
- Negative feedback neural net - JavaScript
- The ART Gallery
- Stuttgart Neural Network Simulator (SNNS)
- Time Delay Neural Network - TDNN
- DartNet
- NeurDS
- Neural Networks at your Fingertips
- Spike and Neuralog

Shareware

- NeuroForecaster/GA
- NETS - Network Execution and Training Simulator
- WinNN
- Backprop-1.4
- BackBrain

Appendix F

Sample of Data for Analysis of Five Credit Scores

Appendix F

Sample of Data for Analysis of Commercially Available Credit Scores

Tenant Number	Months of lease honored	Date	FICO Risk Score	FICO Advanced Risk Score	FICO Installment Score	FICO Finance Score	Experian Sureview Score
1	Multiple persons	2/8/2001	503	490	495	514	319
2	28	2/8/2001	534	536	556	529	876
3	Multiple persons	5/9/2001	547	504	566	505	496
4	9	5/9/2001	506	468	551	548	447
5	29	8/9/2001	607	601	608	604	465
6	9	9/28/2001	469	479	472	460	207
7	7	10/3/2001	488	431	477	478	124
8	6	10/3/2001	569	495	555	530	264
9	7	10/18/2001	0	0	517	518	464
10	8	12/6/2001	516	527	521	499	199
11	multiple persons	12/10/2001					
12	3	12/10/2001	598	616	574	570	654
13	2	12/10/2001	562	493	547	563	429
14	no scores	12/10/2001					
15	14	12/11/2001	791	850	815	821	901
16	6	12/18/2001	600	553	605	571	783
17	2	12/21/2001	444	468	479	510	71
18	4	1/2/2002	645	687	625	655	922
19	multiple persons	1/2/2002	513	556	550	576	465
20	12	1/3/2002	488	471	513	499	238
21	6	1/31/2002	0	0	543	549	0
22	12	2/5/2002	562	518	528	496	218
23	4	2/5/2002	0	0	0	0	0
24	13	2/11/2002	0	0	534	540	441
25	4	2/19/2002	575	512	537	523	621
26	2	2/27/2002	665	684	672	683	963
27	multiple persons	3/7/2002			535	511	666
28	3	3/8/2002	652	640	654	650	914
29	no scores	3/8/2002					
30	6	3/8/2002	498	464	477	486	197
31	12	3/12/2002	589	513	551	563	159

Appendix G

Important Variables Testing Nine Variables

Appendix G

Important Variables Testing Nine Variables to Identify Most Important Seven Variables

First Run		**Second Run**		**Third Run**		
R Square	Important Inputs	R Square	Important Inputs	R Square	Important Inputs	Combination Number
0.3187	ME	0.4387	MH	0.3472	MH	84
0.4315	HE	0.471	HL	0.3004	HL	83
0.4356	HK	0.2714	HK	0.2845	HK	82
0.4097	HF	0.2128	GE	0.3163	HG	81
0.3827	FK	0.4714	KH	0.4267	HK	80
0.2725	MF	0.3608	FH	0.2923	MF	79
0.5261	ME	0.4087	MH	0.2088	MH	78
0.3122	LH	0.3214	LH	0.304	LG	77
0.468	HK	0.4806	KE	0.6016	KE	76
0.555	KE	0.3861	KE	0.5902	MG	75
0.3188	LF	0.4358	LF	0.3496	FE	74
0.4838	LF	0.4731	ML	0.4374	MF	73
0.4341	MG	0.2585	ME	0.2898	GK	72
0.3448	KG	0.4954	KF	0.3579	LK	71
0.3755	MG	0.3835	FL	0.438	ML	70
0.3676	ME	0.6348	MG	0.3321	MG	69
0.3902	MI	0.3016	ME	0.6257	ME	68
0.5297	LE	0.3438	IE	0.2373	JE	67
0.4954	KG	0.3672	GF	0.4437	GE	66
0.3766	KH	0.6313	KF	0.5396	HF	65
0.5502	LF	0.2834	KH	0.4183	HF	64
0.5272	MH	0.476	LH	0.3624	KH	63
0.2933	KH	0.4277	MK	0.286	KH	62
0.402	LF	0.3181	KH	0.6328	KH	61
0.3326	LK	0.4537	LK	0.4719	ML	60
0.3994	MK	0.4173	KF	0.4031	KF	59
0.5543	MK	0.4719	MK	0.6611	MF	58
0.5432	LK	0.4795	LK	0.3252	LF	57
0.3239	HF	0.589	LH	0.4599	LH	56
0.4375	MH	0.3339	HF	0.4498	LH	55
0.389	ME	0.4807	LF	0.4926	LF	54
0.3233	MH	0.4929	MH	0.3935	MF	53
0.397	LH	0.3237	FE	0.3576	LF	52

0.3672	KH	0.3011	HF	0.3428	KH	51
0.4645	MK	0.4028	KH	0.2818	MH	50
0.6612	KH	0.4069	KH	0.2801	KE	49
0.3498	MH	0.3947	ME	0.3817	MH	48
0.6705	MH	0.3421	MG	0.3834	MH	47
0.2505	KH	0.3859	KG	0.2325	JH	46
0.2885	MK	0.4229	MK	0.2826	MK	45
0.3464	KG	0.2555	ML	0.2655	MK	44
0.5038	KI	0.2972	JI	0.3853	MI	43
0.5308	KE	0.4729	GE	0.346	KG	42
0.3324	ME	2134	MK	0.2846	MK	41
0.2595	MK	0.2348	ML	0.4773	ML	40
0.2399	ME	0.3342	ME	0.2585	ME	39
0.2681	MK	0.4585	MK	0.3643	ME	38
0.3823	KH	0.3015	KE	0.2491	KE	37
0.6276	MH	0.2493	MI	0.3824	MH	36
0.3814	MH	0.4149	MG	0.5296	MH	35
0.3243	LE	0.2907	ML	0.2473	ME	34
0.2545	MH	0.247	ML	0.2496	ME	33
0.5268	ME	0.3717	ME	0.2677	MH	32
0.5281	IH	0.315	JG	0.2809	HI	31
0.3864	KH	0.3199	KH	0.4942	KE	30
0.4365	MH	0.6195	MH	0.3743	FH	29
0.3078	FH	0.2381	LH	0.1889	FH	28
0.4968	KH	0.5977	MH	0.5838	MH	27
0.4354	KF	0.2097	MK	0.3255	HF	26
0.5179	KH	0.4255	KF	0.3594	KH	25
0.3537	GK	0.3159	LI	0.4302	MG	24
0.5468	MF	0.3416	ML	0.3774	GF	23
0.5024	MK	0.5446	KF	0.3778	KF	22
0.2879	LG	0.5157	LF	0.4295	KG	21
0.5335	LH	0.513	KF	0.529	ML	20
0.4027	KF	0.2407	LH	0.3895	LH	19
0.3876	MK	0.3079	LK	0.571	LK	18
0.4842	MF	0.4789	HF	0.5232	KH	17
0.3227	HF	0.3894	KH	0.3948	KH	16
0.264	MH	0.433	MH	0.4981	MG	15
0.4899	MH	0.387	MH	0.4323	MG	14
0.3406	MG	0.5081	MG	0.6019	MG	13
0.3125	MH	0.255	ML	0.2846	LK	12
0.4448	MK	0.3517	MH	0.312	MH	11
0.4238	KG	0.3089	LH	0.3278	KH	10
0.4854	MH	0.2983	LG	0.4481	ML	9
0.4086	MG	0.2458	ML	0.5008	ML	8
0.3022	LG	0.4605	ML	0.3765	LF	7

0.434	MH	0.3994	LF	0.4726	MH	6
0.4059	ML	0.2934	MH	0.4262	MH	5
0.2853	MG	0.2528	MG	0.2958	MH	4
0.2776	HG	0.2349	LF	0.444	HG	3
0.458	KF	0.2228	KG	0.1614	KH	2
0.2538	HE	0.4275	FE	0.3606	GE	1

E = adult, many adults, adult with child
F = total tenant income
G = vehicle *Blue Book* value
H = driving infractions?
I = criminal activity?
J = total loan balance
K = total monthly payment
L = number of credit inquiries
M = percent satisfactory accounts

Appendix H

Important Variables Testing Seven Variables

Appendix H

Important Variables Testing Seven Variables to Identify Most Important Five Variables

First Run		Second Run		Third Run		
R Square	Important Inputs	R Square	Important Inputs	R Square	Important Inputs	Combination Number
0.3864	GH	0.2838	ML	0.4003	LH	21
0.3115	ML	0.388	FH	0.3217	KF	20
0.2466	MF	0.2582	MF	0.3727	MF	19
0.3123	MH	0.4927	HF	0.3318	ML	18
0.5197	MH	0.477	HF	0.4193	KH	17
0.4395	KH	0.3525	LH	0.4438	LH	16
0.1797	ML	0.1668	KM	0.2831	ML	15
0.3681	ML	0.4074	GK	0.3025	KG	14
0.3404	MG	0.2756	ML	0.2977	ME	13
0.2943	KH	0.3625	KM	0.4572	KH	12
0.2293	GL	0.2709	KH	0.1735	HG	11
0.3168	EM	0.3239	ME	0.3516	ME	10
0.4633	LE	0.2837	LF	0.3408	HF	9
0.3677	MF	0.2964	KM	0.4784	KH	8
0.349	KH	0.3375	HK	0.4079	LF	7
0.5768	ME	0.3027	MF	0.3962	ME	6
0.4311	ME	0.4718	FK	0.4298	ME	5
0.447	FK	0.482	FL	0.5908	FE	4
0.5226	MF	0.2138	MH	0.5278	HF	3
0.3138	HL	0.31	EG	0.4412	HF	2
0.2406	KH	0.2076	KH	0.4892	KG	1

E = adult, many adults, adult with child
F = total tenant income
G = vehicle *Blue Book* value
H = driving infractions?
I = criminal activity?
J = total loan balance
K = total monthly payment
L = number of credit inquiries
M = percent satisfactory accounts

Reference List

Abernathy, T. (2004). *Remarks of Treasury Assistant Secretary for Financial Institutions Wayne A. Abernathy to the Exchequer Club of Washington.* Retrieved March 19, 2006, from http://www.ustreas.gov/press/releases/js140.htm

Ainscough, T. L., & Aronson, J. E. (1999). An empirical investigation and comparison of neural networks and regression for scanner data analysis. *Journal of Retailing and Consumer Services*, *6*, 205-217.

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance*, *18*(3), 505-29.

Angel, B. (2000). A score to settle. *U.S. Banker, 110*(8), 34-36

Australia. (1999). *The National Office for the Information Economy*. Retrieved on March 18, 2006, from http://www.noie.gov.au/projects/ecommerce/legal_framework/index.htm

Australia. (2000). *Consumer protection in electronic commerce: Principles and key issues.* Retrieved October 12, 2004, from http://www.treasury.gov.au/publications/ConsumerAffairs/ElectronicCommerceAndConsumerProtection/ConsumerProtectionInElectronicCommerce-PrinciplesAndKeyIssues/protect.asp

Auten, J. H. (2000*). Remarks to the Treasury Borrowing Advisory Committee of the Bond Market Association*. Retrieved March 17, 2006, from http://www.treas.gov/press/releases/ls819.htm

Avery, R. B., Bostic, R.W., Calem, P. S., & Canner, G.B. (2000). Credit scoring: Statistical issues and evidence from credit bureau files. *Real Estate Economics*, *28*(3), 523-547.

Baker, A., & McKenzie, G. (2000). *E-commerce spotlight*. Retrieved March 18, 2006, from http://www.mbc.com/ecommerce.html

Banasik, J., Crook, J., & Thomas, L. (2001). Scoring by usage. *Journal of Operational Research Society*, *52*(9), 997-1006.

Barron, J., & Staten, M. (2003). The value of comprehensive credit reports: Lessons from the U.S. experience. In *Credit Reporting Systems and the International Economy*. Boston: MIT University Press.

Baxt, W.G., & Skora, J. (1996). Prospective validation of artificial neural network trained to identify acute myocardial infarction. *The Lancet*, *347*(8993), 12-15.

Bigus, J.P. (1996). *Data mining with neural networks: Solving business problems from application development to decision support*. New York: McGraw Hill.

Boon, M. E., Kok, L. P., & Beck, S. (1995). Histological validation of neural-network assisted cervical screening: comparison with the conventional approach. *Cell Vision*, *2*, 23-27.

Boritz, J., & Kennedy, D. (1995). Effectiveness of neural network types for prediction of business failure. *Expert Systems Application*, *9*, 504-512.

Boritz, J., Kennedy, D., & Albuquerque, A. (1995). Predicting corporate failure using a neural network approach. *Intelligent Systems in Accounting, Finance and Management*, *4*, 95-111.

Bortolan, G., & Willems, J.L. (1993). Diagnostic ECG classification based on neural networks. *Journal of Electrocardiology*, *26*, 75-79.

Brill, J. (1998). The importance of credit scoring models in improving cash flow and collections. *Business Credit*, *1*, 16-17.

Brockett, P., Shin. S.L.S., & Kellison, B. (2003). *A statistical analysis of the relationship between credit history and insurance losses*. Austin, TX: University of Texas McComb Business School Press.

Bugera, V., Konno, H., & Uryasev, S. (2002). Credit card scoring with quadratic utility functions. *Journal of Multi Criteria Decision Analysis*, *11*, 197-211.

Bylinsky, G. (1993, September 6). Computers that learn by doing. *Fortune*, 96-102.

Cannon, C. (2000). Bill lets consumers see credit scores. *National Mortgage News*, *24*(43), 26-28.

Cate, F. H. (2002, Summer). Principles for protecting privacy. *Cato Journal*, *22*(1), 33-34.

CBC News (2001). *Canadian doctors raise questions about data mining***.** Retrieved November 12, 2004 from http://cbc.ca/cgi-in/templates/view.cgi?/news/2001/08/15/Consumers/medicalprivacy_010815

Chapman, T. (2004). *Impact of FACTA Act*. Retrieved September 28, 2004, from http://www.bankinfo.com/article.html?id=20040714N8NW1NWD&from=CompanyIndex

Chatzky, J. S. (2003). How credit scoring really works and why it matters. *Money Magazine*, *3*, 142-143.

Cheh, J. J., Weinberg, R. S., & Yook, K. C. (1999). An application of an artificial neural network investment system to predict takeover targets. *Journal of Applied Business Research, 15*(4), 33-45.

Cheng, B., & Titterington, D.M. (1994). Neural networks: A view from a statistical perspective. *Statistical Science*, *9*, 2-30.

Churchhill, G. A., Nevin, R. J., & Watson, R. R. (1977). The role of credit scoring in the loan decision. *Credit World*, *3*(3), 6-10.

Cogger, K. O, Koch, P. D., & Lander, D. M. (1997). *A* neural network approach to forecasting volatile international equity markets. *Advances in financial economics*, 3, 117-157.

Cooper, J. C. B. (1999). Artificial neural networks versus multivariate statistics: An application from economics. *Journal of Applied Statistics*, *26* (8), 909-21.

Credit Infocenter (2002). Can your insurer hold your credit report against you? *Credit Infocenter Report*, *2*(27), 1-7.

Desai, V. S., Convay, D. G., Crook, J. N., & Overstree, G. A. (1997). Credit scoring models in the credit union improvement using neural networks and genetic algorithms. *IMAJ Mathematics Applied in Business and Industry, 8*, 323-346.

Duchessi, P, Shawky, H., & Seagle, J. P. (1988). A knowledge engineered system for commercial loan decisions. *Financial Management*, *17*, 57-65.

Durand, D. (1941). *Risk elements in consumer installment financing*. New York: National Bureau of Economic Research.

Dutta, S., & Shekhar, S. (1988). Bond ratings: A non conservative application of neural networks. *IEEE International Conference on Neural Networks*, *2*, 443-450.

Dwinnell, W. (1998). Modeling methodology: Model input selection. *Journal of PC Artificial Intelligence*, *12*(1), 23-26.

Eberhart, R. C., & Dobbins, R. W. (1990). *Neural network PC tools: A practical guide,* San Diego, California: Academic Press, Inc.

Edelman, D. B. (1997). Credit scoring for lending to small businesses. C*redit scoring and credit control conference V*, Edinburgh, Scotland: University of Edinburgh Press, 35-39.

Elgin, P. E. (1994). Pioneers try "neural networks" to pick stocks. *Computer Cashflow Magazine*, *15*(8), 5.

Emel, A.B., Oral, M., Reisamn, A., & Yolalan, R. (2003). A credit scoring approach for the commercial banking sector. *Journal of Socio Economics Planning Sciences*, *37*, 103-123.

EU Report, (2002). *e-Europe: Information society for all action plan*. Retrieved October 12, 2004, from http://europa.eu.int.ISPO/ida/

Experian Inc. (2003). *Best and worst cities for credit scores*. Cleveland, OH: Experian Press:

Fair, Issac & Company (2002). Understanding your credit score. *Fair, Issac & Company consumer brochure.* San Mateo, CA: Fair Issac & Company Documents.

Fensterstock, A. (2001). The application of neural networks to credit scoring. *Business Credit*, *103*(3), 58-60.

Fisher, B. (1999). Mellon creates fraudwatch to predict and manage risk using neural technology. *Journal of Retail Banking Services*, *21*, 1, 34-44.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics 7*, 179-188.

Frame, W. S., Srinivasan, A., & Woosley, L. (2001). The effect of credit scoring on small-business lending. *Journal of Money, Credit and Banking, 33*(3), 23-35.

Freeman, J. (1999). Neural network development and deployment rules of thumb. *Hydrocarbon Processing*, *10*, 101-107.

U.S. FTC (2005). *National and state trends in fraud and identity theft*. Federal Trade Commission report. Retrieved March 18, 2006 from http://www.consumer.gov/sentinel/pubs/Top10Fraud2004.pdf

Furletti, M. (2002). *An overview and history of credit reporting*. Philadelphia Federal Reserve Bank, 6.

Galindo, J., & Tamayo, P. (1997). *Credit risk assessment using statistical and machine learning methods as an ingredient for financial intermediaries risk modeling*. Retrieved March 18, 2006, from http://ideas.repec.org/p/sce/scecf7/31.html

Garcia, R., & Gencay, R. (2000). Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, *94*(1-2), 93-115.

Gayer, J. (2003). *Policing privacy: Law enforcement's response to identity theft*. Los Angeles: CALPIRG Education Fund White paper.

Glorfeld, L.W., & Hardgrave, B.C. (1996). An improved method for developing neural networks: The case of evaluating commercial loan creditworthiness. *Computer Operation Research, 23*(10), 933-944.

Graves, D. (2000). The future of lending: Automated credit decisions in zero time. *Business Credit*, *102*(9), 12-14

Greene, W. (1998). Sample selection in credit scoring models. *Japan and the World Economy Journal*, *10*, 299-316.

Grudnitski, G., & Osburn, L. (1993). Forecasting S&P and gold futures prices: An application of neural networks. *Journal of Futures Markets*, *13*(6), 631-643.

Hamm, L., & Brorsen, B. W. (2000). Trading futures markets based on signals from a neural network. *Applied Economics Letters*, *7*(2), 137-40.

Handzic, M., & Aurum, A. (2001). Knowledge discovery: Some empirical evidence and directions for future research. *5th International Conference on Wirtschafts Informatics*, Augsburg, Germany, 964-974.

Hawley, D.D., Johnson, J.D., & Raina, D. (1990, November). Artificial neural systems: A new tool for financial decision-making. *Financial Analysis Journal*, 63-72.

Haykin, S. S. (1999*). Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall.

Henley, W.E. (1995). *Statistical aspects of credit scoring*. Unpublished doctoral dissertation, The Open University, Milton Keynes, UK.

Henley, W.E. (1996). A k-nearest neighbor classifier for assessing consumer credit risk. *Statistician*, *44*, 77-95.

Hertz, J., Krogh, A., & Palmer, R.G. (1991). *Introduction to the theory of neural computation.* Redwood City, CA: Addison Wesley Publishing.

Hill, T., & Lewicki, P. (2006). *Statistics: Methods and applications.* Tulsa, OK: Statsoft Press Inc.

Hinton, G.E. (1992). How neural networks learn from experience. *Scientific American*, *267*, 144-151.

Hitheesing, N. (1996). The mother of development. *Forbes*, *157*(2), 88.

Hsieh, C. (1993). Some potential applications of artificial neural systems in financial management. *Journal of Systems Management*, *44*(4), 12-16.

Hu, M. Y., & Tsoukalas, C. (1999). Combining conditional volatility forecasts using neural networks: An application to the ems exchange rates. *Journal of International Financial Markets, Institutions & Money*, *9*(4), 407-22.

Internet Policy Institute (2000). *Regulating and legislating the internet and e-commerce*. Retrieved March 18, 2006, from http://internetpolicy.org/briefing/3_00.html

Jensen, H.L. (1992). Using neural networks for credit scoring. *Managerial Finance*, *18*, 15-29.

Jones, W. (2001). Study: Insurance scoring becomes more favored. *Best's Review*, *102*(5), 16

Kaastra, I., & Boyd, M. S. (1995). Forecasting futures trading volume using neural networks. *Journal of Futures Markets*, *15*(8), 953-970.

Kadet, A. (2003). FICO frenzy. *Smartmoney*, *12*(5), 104-109.

Khan, H.A. (2002). *Can banks learn to be rational*? University of Tokyo. Retrieved on March 18, 2006, from http://www.e.u-tokyo.ac.jp/cirje/research/dp/2002/2002cf151.pdf

Kim, C. N., & McLeod, R., Jr. (1999). Expert linear models and nonlinear models of expert decision making in bankruptcy prediction. *Journal of Management Science, 16*(1), 189-206.

Kim, J.W. (1993). Expert systems for bond rating: Comparative analysis of statistical, rule-based, and neural network systems. *Expert Systems 10*, 167–171.

Lacher, R., Coats, P., Sharma, S., & Fants, F.A. (1995). A neural network for classifying the financial health of a firm. *European Journal of Operations Management*, *85*(1), 53-65.

LaFalce, J. (2000, May 4) *LaFalce introduces sweeping legislation to protect consumer privacy.* Democrat news conference on financial privacy. Retrieved March 18, 2006, from http://www.appliedprofit.com/glba.html

Landesberg, M. K., Levin, T. M., Curtin, C. G., & Lev, O. (1998). *Privacy online: A report to Congress*. Federal Trade Commission Whitepaper. Washington: Government Printing Office**.**

Lee, K.W., & Lam, H. N. (1995). Optimal sizing of feed forward neural networks: Case studies. *IEEE 2nd Conference on Neural Networks and Expert Systems*, *2*, 79-82.

Leonard, K. (1996). Information systems and benchmarking in the credit scoring industry. *Benchmarking for Quality Management & Technology*, *3*(1), 38-44.

Lew, J. (1999) *Privacy policies on federal web sites. Memorandum m-99-18 for the heads of executive departments and agencies*. Retrieved March 18, 2006, from http://www.whitehouse.gov/omb/memoranda/m99-18.html

Lewis, E. M. (1992). *An introduction to credit scoring*, San Rafael, CA: Aetna Press.

Marakas, G.M. (1999). *Decision support systems in the twenty-first century*. Upper Saddle River, NJ: Prentice Hall.

Marteli, J.P., Panichelli, R., Strauch, R., & Taylor-Schoof, S. (1997). The effectiveness of credit scoring on low to moderate income and high minority area populations. *Fair, Issac and Company Research Report*, Rafael, CA: Fair, Issac and Company Press.

Mason, A. B., & Lalor, J. J. (1877). *The primer of political economy in sixteen definitions and forty propositions*. Chicago: Jansen, McClurg & Company.

Mazer, E.E. (2001). Insurers defend use of credit scoring. *National Underwriter Property & Casualty*, *105*(52), 19.

McDonald, C. (2003). NCOIL model addresses credit score issues. *Journal of National Underwriter*, *107*(16), 17-22.

Mechling, J., & Applegate, L. (2001). *Eight imperatives for leaders in a networked world*. Boston: Harvard College Press.

Mester, L.J. (1997, September). What the point of credit scoring. *Federal Reserve Bank of Philadelphia Business Review*, *10*, 3-16.

Monahgan, J. (2000). The impact of personal credit history on loss performance in personal lines. *Conference on ratemaking of the Casualty Actuarial Society*, *3*, 79- 107.

Moshiri, S., Cameron, N. E., & Scuse, D. (1999). Static, dynamic, and hybrid neural networks in forecasting inflation. *Computational Economics*, *14*(3), 219-35.

Mozer, M. & Smolensky, P. (1997). Skeletonization: A technique for trimming the fat from a network via relevance assessment. In *Advances in Neural Information Processing Systems*, *1*, 107–115

MPF Research, (2004, September). Statistics and U.S. overview. *Southeast Apartment Markets Conference,* Atlanta, GA, 25.

Nechyba, M. C., & Yangsheng, X. (1995). Human skill transfer: Neural networks as learners and teachers. *IEEE Conference on Robots and Systems*, *4*, 314-319.

Nelson, M. M., & Illingworth, W. T. (1991). *A practical guide to neural nets.* Reading, MA: Addison-Wesley Publishing Company.

Olmstead v. U.S., 277 U.S. 438(1928). Justice Brandeis, dissenting. Retrieved April 2, 2006 from http://caselaw.lp.findlaw.com/scripts/getcase.pl?court=us&vol=277&invol=438

Ormerod, P., Taylor, J.C., & Walker, T. (1991). *Neural networks in economics*. Cambridge, MA: Blackwell Publishing.

Penrose, P. (1993, February 26). Star dealer who works in the dark. *The London Times*, 28.

Pfleeger, S. L. (2001). *Software engineering: Theory and practice*. Upper Saddle River, NJ: Prentice Hall.

Piramuthu, S. (1998). Feature selection for financial credit risk evaluation decisions. *Operations and Information Management*, *98*(11), 23-33.

Platts, G., & Howe, I. (1997). A single European scorecard. C*redit scoring and credit control conference V.* Edinburgh, Scotland: University of Edinburgh, 17-19.

Principe, J. C., Euliano, N.R., & Lefebrve, W.C. (2000). *Neural and adaptive systems: Fundamentals through simulation*. New York: Wiley Publishing.

Quinlan, J. R. (1993). Comparing connectionist and symbolic learning methods. In *Computational learning theory and natural learning systems: Constraints and prospects.* Cambridge, MA: MIT Press.

Reichert, A. K. (1983). An examination of the conceptual issues involved in developing credit scoring models. *Journal of Business and Economic Statistics*, *1*, 101-114.

Rental Property Reporter, (2005). *Owners and managers survey results*. Retrieved on January 22, 2006 from http://www.rentalprop.com/survey-results.htm

Rosenberg, E., & Gleit, A. (1994). Quantitative methods in credit management. *Operations Research Journal*, *42*(4), 589-613.

Rumelhart, D. E., Hinton, D. E., & Williams, R. J. (1986). *Learning internal representations by error propagations*. Cambridge, MA: MIT Press.

Saloma, J. S. (1984). American effective law enforcement. *Ominous Politics*, *29*, 16-17.

Sarle, W. S. (1994). Neural networks and statistical models. *Nineteenth Annual SAS Users Group International Conference*, Cary:NC, 1538-1550.

Schwartz, T. (1995). Applications on parade. *Electronic Design*, *43*(16), 68.

Scranton, D. F. (2001). Lock up your credit files; FCRA is back in town. *The Banking Law Journal*, *118*(6), 580-591

Setiono, R., Leow, W. K., & Thong, J. Y. L. (2000). Opening the neural network black box: An algorithm for extracting rules from function approximating artificial neural networks. *IEEE twenty first international conference on information systems, 12*, 176-186.

Shaw, M., & Gentry, J. (1998). Using an expert system with inductive learning to evaluate business loans. *Financial Management*, *17*(3), 45-56.

Shtub, A., & Versano, R. (1999). Estimating the cost of steel pipe bending: A comparison between neural networks and regression analysis. *International Journal of Production Economics, 62*(3), 201-207.

Shurkan, A., & Singleton, J. (1991). Neural networks for bond rating improved by hidden layers. *IEEE International Conference on Neural Networks, 2*, 157-162.

Singapore. (2000). *Misuse of the Internet*. Retrieved September 23, 2004, from http://www.sba.gov.sg/work/sba/internet.nsf/pages/misuse

Singapore. (2001). *Singapore launches e-commerce master plan*. Retrieved March 18, 2006, from http://www.ida.gov.sg/idaweb/media/infopage.jsp?infopagecategory=ncbarchivemediareleases.mr:media&versionid=5&infopageid=I1008

Singleton, J.C., & Surkan, A. J. (1995). Bond rating with neural networks. In A. P. Refenes (Ed.), *Neural networks in the capital markets*. Chichester, England: Wiley Publishing, 301–307.

Singleton, J. C., & Surkan, A. J. (1990). Neural networks for bond rating improved by multiple hidden layers. *IEEE international conference on neural networks*, 163–168.

Skapura, D. M. (1996). *Building neural networks*. Reading, MA: Addison-Wesley.

Soman, D., & Cheema, A. (2002). The effect of credit on spending decisions: The role of the credit limit and credibility. *Marketing Science*, *21*(1), 32-53.

South Africa. (1999). *Green paper for electronic commerce in South Africa*. Retrieved September 23, 2004, from http://www.polity.org.za/govdocs/green_papers/greenpaper/introduction.PDF

Star Systems (2001). *Financial privacy: Beyond the Gramm-Leach Bliley Act*. New York: Internal Whitepaper.

Swales, G., & Yoon, Y. (1992). Applying neural networks to investment analysis. *Financial Analysts Journal*, *48*, 78-80.

Synovate Research (2003). *Federal Trade Commission identity theft survey report*. Retrieved March 18, 2006, from http://www.ftc.gov/os/2003/09/synovatereport.pdf

Takita, H. (1995). Pennies from heaven: Selling accurate weather predictions. *Today (Japan), 63*(7), 14.

Tam, K.Y., & Kiang, M. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, *38*(7), 926-947.

Taylor, I. (2000). Credit scoring gets an "A". *Black Enterprise*, *31*(4), 57

Terna, P. (1997). Neural network for economic and financial modeling: Summing up ideas emerging from agent based simulation and introducing an artificial laboratory, *Cognitive Economics*, *1*, 271-309.

Thomas, L. C., Hand, D. J., & Jacka, S. D. (1998). *Statistics in finance*, London: Arnold Press.

Trippi, R. R., & DeSieno, D. (1992). Trading equity index futures with a neural network. *Journal of Portfolio Management*, *19*(1), 27-33.

Trippi, R. R., & Turban, E. (1996). *Neural networks in finance and investing 2nd Ed.*, New York: McGraw Hill Irwin Publishing Company.

Turban, E., & Aronson, J. E. (2001). *Decision support systems and intelligent systems*. Upper Saddle River, N.J.: Prentice Hall

U.S. Census Bureau. (1996). *Property owners and managers survey*. Retrieved January 30, 2006, from http://www.census.gov/hhes/www/housing/poms/pomtsum3.html

U.S. Chamber of Commerce. (2003). *The Fair Credit Reporting Act: Access, efficiency & opportunity*. Washington: Information Policy Institute Press.

U.S. Federal Trade Commission. (2000) *Privacy online: Fair information practices in the electronic marketplace—A report to Congress*. Gaithersburg, MD: Federal Trade Commission.

U.S. Federal Trade Commission. (2003). *FTC overview report and timeline of the IDT program.* Retrieved March 18, 2006, from http://www.ftc.gov/os/2003/09/timelinereport.pdf

U. S. General Accounting Office. (2000). *Federal agencies' fair information practices*. GAO/AIMD-00-296R. Gaithersburg, MD: General Accounting Office.

U. S. West, Inc. v. Federal Communications Commission (1999). *Certificate denied on Petition for review of an order of the Federal Communications Commission.* 10th Circuit. Retrieved on March 18, 2006, from http://www.kscourts.org/ca10/cases/1999/08/98-9518.htm

Van Der Spek, R., & Spijkervet, A., (1997). *Knowledge management: Dealing intelligently with knowledge management*. Edinburgh, Scotland: Knowledge Management Network

Walczak, S. (1999). Gaining competitive advantage for trading in emerging capital markets with neural networks. *Journal of Management Information Systems*, *16*(2) 177-192.

Wallison, P.J. (2001, July 24). When privacy is a credit risk. *The New York Times*, A23.

Weiss, S. M., & Kulikowski, C.A. (1991). *Computer systems that learn,* San Mateo, CA: Morgan Kaufmann Publishing.

West, D. (2000). Neural network credit scoring models. *Journal of Computer and Operations Research*, *27*, 1131-1152.

Westin, A. (1967) *Privacy and freedom*. New York: Atheneum Publishing.

Weston, L. P. (2004). *Free credit reports are just the start*. Retrieved March 18, 2006, from http://moneycentral.msn.com/content/Banking/Yourcreditrating/P100244.asp

White, H. (1988). Economic prediction using neural networks: The case of IBM daily stock returns. *IEEE international conference of neural networks*, San Diego, CA, 451-458.

White, H. (1992). *Artificial neural networks: Approximation and learning theory*. Oxford, England: Blackwell Publishing.

Widrow, B., Rumelhart, D. E., & Lehr, M. A. (1994). Neural Networks: Applications in Industry*, Business and Science*, Journal A, *35*(2)17-27.

Witkowska, D. (1995). Neural networks as a forecasting instrument for the Polish stock exchange. *International Advances in Economic Research*, *1*(3), 232-241.

Witten, I., & Eibe, F. (2005). *Data mining: Practical machine learning tools and techniques*, San Francisco, CA: Morgan Kaufman Publishing.

Wray, B., Palmer, A., & Bejou, D. (1994). Using neural network analysis to evaluate buyer-seller relationship. *European Journal of Marketing*, 28(10), 36.

Yen, G., & Lu., H. (2002). Hierarchical genetic algorithm for near optimal feed forward neural network design. *International Journal of Neural Systems*, *12*(1), 31-43.

Yin, W., & DeVaney, S.A. (1999). A comparison of two credit scoring models: Commercial versus consumer education. *American Council on Consumer Interests*, 45th Conference, *1*, 136-145.

Yobas, M.B., Crook, J.N., & Ross, P. (1997). *Credit scoring using neural and evolutionary techniques*. University of Edinburgh. Retrieved March 18, 2006, from http://www.dai.ed.ac.uk/papers/documents/mt9687.html

Yoon, Y., & Swales, G. (1997). Predicting stock price performance. *24th Hawaii International Conference on System Sciences*, *4*, 156-162.

Zahedi, F. (1993). *Intelligent systems for business: Expert systems with neural networks.* Belmont, WA: Wadsworth Publishing Company, 10-11.