MODELING SAFETY PERFORMANCE OF SPECIALTY CONTRACTORS

By

BRENT ROGERS

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN BUILDING CONSTRUCTION

UNIVERSITY OF FLORIDA

2009

To my family

ACKNOWLEDGMENTS

I would like to thank Dr. Jimmie Hinze for his direction and assistance in my research. I am also very grateful for Dr. Ian Flood's and Dr. R. Raymond Issa's contributions in the area of empirical modeling to my research.

I would also like to thank my family for their unwavering support of my decision to pursue an advanced degree.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS 4

LIST OF TABLES 7

LIST OF FIGURES 8

LIST OF ABBREVIATIONS 10

ABSTRACT 11

CHAPTER

1 INTRODUCTION 13

Background 13
Scope 15
Objective 15
Organization 15

2 LITERATURE REVIEW 17

Introduction 17
Costs of Injuries 17
Recordable Injury Rate 19
Experience Modification Rate 21
Drug Testing 23
Nature of the Firm 26
Worker Orientation 27
Safety Professionals 28
Empirical Modeling 29

3 METHODOLOGY 33

Introduction 33
Data Collection 33
Selection of Variables 34
Model Development 35
Safety Scorecard 36
Multivariate Linear Regression 37
Artificial Neural Network 37
Training the Models 38
Selecting the Best Model 39

4 RESULTS 40

Introduction 40
Input Set One 40
Input Set Two 41
Input Set Three 42
Summary 45

5 CONCLUSIONS 58

Introduction 58
Predictive Model 59

6 RECOMMENDATIONS 61

Introduction 61
Recommendations for the Industry 61
Recommendations for Further Study 62

LIST OF REFERENCES 64

BIOGRAPHICAL SKETCH 66

LIST OF TABLES

Table page

4-1 Standard errors and correlation coefficients for input set 1 41

4-2 Standard errors and correlation coefficients for input set 2 42

4-3 Standard errors and correlation coefficients for input set 3 43

4-4 Summary table showing standard errors and correlation coefficients across all models and input sets 46

LIST OF FIGURES

Figure page

2-1 Experience modification rate calculation 30

2-2 Risk scores of various building trades 31

2-3 Structure of a feed forward artificial neural network..... 32

4-1 Actual RIR and scorecard 1 predicted RIR for each example 46

4-2 Actual RIR vs. scorecard 1 predicted RIR 47

4-3 Actual RIR and multivariate linear regression 1 predicted RIR for each example 47

4-4 Actual RIR vs. multivariate linear regression 1 predicted RIR 48

4-5 Actual RIR and ANN 1 predicted RIR for each example..... 48

4-6 Actual RIR vs. ANN 1 predicted RIR 49

4-7 Actual RIR and scorecard 2 predicted RIR for each example 49

4-8 Actual RIR vs. scorecard 2 predicted RIR 50

4-9 Actual RIR and multivariate linear regression 2 predicted RIR for each example 50

4-10 Actual RIR vs. multivariate linear regression 2 predicted RIR 51

4-11 Actual RIR and ANN 2 predicted RIR for each example..... 51

4-12 Actual RIR vs. ANN 2 predicted RIR 52

4-13 Actual RIR and scorecard 3 predicted RIR for each example 52

4-14 Actual RIR vs. scorecard 3 predicted RIR 53

4-15 Actual RIR and multivariate linear regression 3 predicted RIR for each example 53

4-16 Actual RIR vs. multivariate linear regression 3 predicted RIR 54

4-17 Actual RIR and ANN 3 predicted RIR for each example..... 54

4-18 Actual RIR vs. ANN 3 predicted RIR 55

4-19 Absolute error as a function of annual volume (million $) 55

4-20 Absolute error as a function of full-time field workers employed..... 56

4-21 Absolute error as a function of percent of work subcontracted..... 56

4-22 Absolute error as a function of holding toolbox talks 57

4-23 Absolute error as a function of percent of time a full-time safety director spends in the field..... 57

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
EMR	Experience Modification Rate
RIR	Recordable Incident Rate

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science in Building Construction

MODELING SAFETY PERFORMANCE OF SPECIALTY CONTRACTORS

By

Brent Rogers

December 2009

Chair: Jimmie Hinze
Cochair: Ian Flood
Major: Building Construction

Small specialty contractors perform much of the work in the building construction industry. As such, employees of these firms are the ones being exposed to jobsite hazards. However, little research into construction safety has been focused on small and medium-sized specialty contractors. Smaller-sized contractors are generally thought of as being either financially unable or unwilling to invest in safety improving measures. A predictive model which can demonstrate numerically the positive impact of certain safety policies and practices may encourage such investment, and as a result, promote the health of the construction workforce.

The objective of this research was to develop a model that can predict the recordable incident rate (RIR) of small and medium-sized specialty contractors. The tools used to create the model were a scorecard developed by the researcher, multivariate linear regression, and an artificial neural network. Data was collected from a previous study performed at the University of Florida. Inputs for the models were leading indicators that impact safety performance. RIR was the output of the models.

The study found that an artificial neural network was the most accurate predictor of RIR. The artificial neural network's accuracy improved as variables that appeared to be

confusing the model were dropped from the input set. However, given the small sample size, it is impossible to conclusively state that the neural network is the best tool for modeling safety performance. Further exploration with a much larger sample size would be needed to confirm the results of this study.

CHAPTER 1
INTRODUCTION

Background

The construction industry has long suffered from high injury rates. According to the United States Bureau of Labor Statistics (BLS), the construction industry had an Occupational Safety and Health Administration (OSHA) recordable incident rate (RIR) of 5.4 in 2007. In comparison, the private sector as a whole had an RIR of 4.2 during that same time (BLS 2008). In earlier times, it was thought that injuries and even fatalities incurred during the construction process were an unpleasant but unavoidable part of doing business. This perception has gradually given way to the idea that severe injuries and fatalities on construction are unacceptable, rather than simply unpleasant. While incidents still occur, construction companies are increasingly making efforts, both on their own and as required by legislation, to minimize their frequency and severity.

Large construction firms can more easily absorb the cost of implementing safety practices than can small and medium-sized firms. However, In the United States, the vast majority of companies are small (less than 20 employees). These companies often operate as specialty contractors, performing one or several particular construction specialty tasks. Such companies have the most difficulty implementing measures that are intended to improve safety performance. This can be true as a result of financial issues or simply a lack of understanding of the value of reducing injuries. Smaller companies may assume that the investment needed to reduce incidents will not provide a good return or that such expenditures would make them uncompetitive when bidding projects. This may explain why specialty trade contractors suffer from higher injury and illness rates than the construction sector as a whole (BLS 2008).

Computer modeling may be a way to demonstrate to small and medium-sized construction firms the impacts of safety initiatives on improving safety performance. Computer modeling is used in a wide variety of research, from predicting crop yields in agriculture to battle simulations for the military. It is also extensively used in construction with applications to a variety of problems, including predicting activity durations and financial outcomes of construction projects.

Of particular use are artificial neural networks. Artificial neural networks make predictions in the same way humans do, by learning from experience. Neural networks are trained on data to find the relationships between variables and outcomes or outputs. The neural network can find patterns and relationships that might otherwise be missed with simpler modeling techniques. The models developed by the neural network are tested with inputs which have a known output. This allows the researcher to determine which model is the most accurate by comparing the model's predicted output with the known actual output. After models are tested the best model is selected and validated with additional data. The best model may then be applied to the situation it was trained to predict by collecting and inputting the information and then running the model to generate an output prediction.

While prior construction safety research has utilized conventional statistical analysis techniques to determine the effectiveness or ineffectiveness of certain policies and practices on the RIR in the construction industry, little of this work has been focused on small and medium-sized companies. Furthermore, there is a scarcity of prior research on predictive models for construction safety. Construction companies are for-profit businesses, and as such, make investment decisions based on the expected

rate of return. A predictive model that shows the impact of safety policy changes on the RIR could prove to be a valuable resource for decision-makers in the construction industry.

Scope

The study is focused on examining how attributes and safety policies of small and medium sized specialty contractors impact safety performance. The models developed will use leading indicators as inputs. The inputs will be identified through a literature review. Analyses will be conducted on the available data to determine if the impact of the inputs is sufficiently substantial to warrant including them in the models. An artificial neural network will be used as the primary modeling tool, along with a scorecard developed by the researcher and multivariate linear regression.

Objective

The objective of the research is to develop a model that can predict the RIR for small and medium construction companies. The most accurate model will be offered to the industry to show numerically the impact of their safety policy decisions on the RIR.

Organization

The research is organized into six chapters. Chapter 1 is the introduction. Chapter 2 is a review of literature related to construction safety and modeling. Literature reviewed will provide a brief history of safety in the construction industry, the financial impacts of incidents, lagging indicators of safety performance, leading indicators of safety performance, definitions of modeling and artificial neural networks, and the applications of artificial neural networks in the construction industry. Chapter 3 describes the methodology this researcher used in the study. Results and analysis of

the study are located in Chapter 4. Conclusions based on the results are contained in Chapter 5. Finally, Chapter 6 provides this researcher's recommendations for further study.

CHAPTER 2
LITERATURE REVIEW

Introduction

This literature review will explore past research conducted in the area of construction safety. The costs of workplace injuries and illness in construction will be examined first to underscore their importance for financial stability. Lagging indicators of safety performance and their suitability for use as an output in an empirical model will then be studied. A review of leading indicators of safety performance follows this to provide an understanding of potential inputs into the model. Finally, empirical modeling and specifically, artificial neural networks will be briefly examined to understand the basic principles behind developing, testing, and selecting models. Whenever possible the literature review will focus on data specific to small and medium-sized or specialty contractors.

Costs of Injuries

Ethically speaking, it is undeniable that firms should make efforts to prevent injuries and fatalities from occurring on their projects. While ethical conduct is important, it is also critical to understand the financial benefits of creating a safe work environment. Promoting worker safety can result in the avoidance of substantial costs, resulting from both the incident itself and lost productivity due to the event.

Hinze and Appelgate sought to identify and quantify the costs associated with construction injuries in 1991. They state that the costs of an injury can be divided into two basic categories, direct and indirect. Direct costs are readily identifiable and can have a dollar value applied to them without much difficulty. These are the costs that are likely covered by the workers' compensation insurance carrier and are a directly

connected to the injured person. The indirect costs of an injury encompass all other negative financial impacts as a result of the incident.

Indirect costs of injuries are broad in scope and difficult to quantify with accuracy. The most obvious indirect costs are associated with lost productivity as a result of an injury. Injured workers will continue to be paid after they are injured even though they are no longer productively working. That is, injured workers and any fellow workers assigned to drive them to receive medical treatment will remain stay “on the clock” on the day of the injury. Additionally, after injured workers have recovered sufficiently to return to the project, they may still be required to be on light-duty by a doctor or they may not be able to perform their regular job duties as well as prior to being injured. In such cases, little meaningful work may be accomplished during that time, but the worker is still being paid a regular wage.

The crew of the injured worker is also negatively affected by the injury accident. Psychologically, after the injured worker has been taken to receive medical treatment, the remaining workers may be shaken by the event. Concern for their co-worker and fear that something similar could occur to them are feelings that can distract from the work at hand. Additionally, at the time of the injury, workers on the crew will likely come to help the injured worker, ceasing their productive activities for some time. Other workers in the area, even those in other crews, will likely stop their normal work activities when an injury accident occurs.

Large amounts of paperwork often result from injuries. Witnesses may need to fill out written statements. In addition, a supervisor will need to conduct an accident

investigation and prepare a report on the event. All of this time spent on the incident is time that would otherwise have been spent on productive work.

The actual dollar amount of an injury varies considerably with the type of injury. Hinze and Appelgate found the costs of an injury which did not require days away from work or restricted activity to be approximately $519 and $442 in direct and indirect costs, respectively. For injuries that did result in restricted activity or lost days, the costs soared to approximately $6,910 and $1613 for direct and indirect, respectively. Neither of these figures factored in the cost of liability claims. When claims costs are factored in, the costs rise much higher.

In 2007, Waehrer et al. analyzed the costs, including claims, of lost time injuries in construction from 2002 data. The entire construction industry had a cost of $42,093 per incident. Painters, electricians, masons, and miscellaneous specialty contractors all had higher costs per incident than the industry as a whole. Mechanical contractors had less expensive injuries on a per incident basis, but more than made up for it with the volume of injuries, making their trade the most expensive of the entire industry in terms of total costs of serious injuries. These facts indicate not only the importance of minimizing injuries as a part of being financially prudent, but also the heightened need for such efforts among specialty trade contractors, as their incidents can be both more costly and more numerous.

Recordable Injury Rate

The OSHA recordable injury rate (RIR) is considered one of the simplest and clearest measurements of determining safety performance. It is a lagging indicator as it reflects performance after-the-fact. The computation of RIR is easy and can be quickly done if data are readily available. RIR equals the number of recordable injuries and

illnesses multiplied by 200,000 and divided by man-hours worked. The basic calculation is that it is the number of recordable incidents per 100 employees per year. However, since many construction employees may not be working for an entire year for the same firm, worker-hours are used. The 200,000 figure comes from a total of 100 employees that are working 40 hours per week for 50 weeks per year (BLS 2009).

The determination of what constitutes a "recordable injury" is more complex. First, the incident must meet the criteria for consideration as work-related. OSHA states that

> An injury or illness is considered work-related if an event or exposure in the work environment caused or contributed to the condition or significantly aggravated a preexisting condition. Work-relatedness is presumed for injuries and illnesses resulting from events of exposures occurring in the workplace, unless an exception specifically applies.

Once it is determined that an injury or illness meets OSHA's criteria, the next step is to determine if it is recordable. Recordable injuries and illnesses range from death to anything that requires more than first aid. OSHA defines as recordable those work-related injuries and illnesses that result in the following:

- Death
- Loss of consciousness
- Days away from work
- Restricted work activity or job transfer, or medical treatment beyond first aid

Additionally, any work-related injury or illness that meets any of the following criteria must be recorded:

- Anything that has been diagnosed by a licensed health care professional
- Cancer
- Chronic Irreversible Disease
- A fractured or cracked bone

- A punctured eardrum
- Any needle-stick injury or cut from a sharp object that is contaminated with another person's blood or other potentially infections material
- Any case requiring an employee to be medically removed under the requirements of an OSHA health standard
- Tuberculosis infection as evidenced by a positive skin test or diagnosis by a physician or other licensed health care professional after exposure to a known case of active tuberculosis

It is important to note that visits to the doctor solely for observation or diagnostic procedures (when negative) are not considered as recordable (OSHA 2004).

The OSHA recordable injury rate makes an excellent criterion for judging safety performance as it reflects recent changes in safety performances. It is not biased against or in favor of any particular contracting firm. The input variables are related solely to the total work hours and the total number of recordable injuries. The injury rate is also a standardized measurement that is used across all industries in the United States, so easy comparisons can be made from one type of work to another to determine how well an industry is faring compared to the country as a whole.

Experience Modification Rate

The experience modification rate (EMR) is a lagging indicator of safety which is directly related to the financial impact of injuries. The EMR, commonly just referred to as the "mod rate," of a contractor is the multiplying factor use to modify the manual rate that is to be paid in worker's compensation insurance premiums. Everett and Thompson reviewed EMR computation extensively in 1995. Their research provides a solid foundation for understanding the impact of EMR on insurance premiums and what factors affect the EMR.

The basic formula insurers use in determining the amount contractors pay in premiums is calculated by multiplying the Manual Rate by Payroll Units and then multiplying by EMR. The manual rate is based on the type of work being performed within a particular state. Every year a rating bureau will set the value of the manual rate for each and every trade based on the history of injuries and payouts for employees performing that trade. Simply put, the manual rate reflects the insurance risks of the various specific trades. A payroll unit is an employer's non-overtime labor cost divided by 100. Everett and Thompson make note that it is non-overtime labor costs instead of including overtime costs because the risk is just related to the amount of time working. However, this may be misleading as some studies have shown that overtime is associated with an increased risk of injury (Dembe et al. 2005). Finally, the EMR comes into play in the formula. The EMR is a reflection of the contractor's accident and injury history. Contractors with poor safety performances will have EMR values greater than one. Contractors with good safety records will generally have EMR values considerably less than one, e.g. 0.6 or 0.7. One can see that whenever the value of EMR increases, the premiums will increase by that same proportion. Thus contractors will either pay more or less than the manual rate based on their EMR.

> EMR is computed based on the safety performance of three of the previous four years (the most recent year is excluded). The exact formula is shown in figure 2-1. For the purposes of this study, it is just necessary to understand that safety performance has a direct impact on the EMR of a company, which affects the financial bottom line of construction firms. Not only does it impact their payments to insurance companies for workers' compensation coverage, but it can also affect their ability to get work, as owners may use the EMR as a method for evaluating and/or qualifying contractors for selection on their projects (Hinze et al. 1995).

The usefulness of the EMR as an output value in a computer model to evaluate safety is questionable as EMR does not take into account the most recent year's safety

performance, i.e. the EMR does not reflect current or recent performance. Thus, recent changes in safety policy will have a lessened or no impact on a company's EMR.

Drug Testing

Construction workers are disproportionately high drug abusers and alcohol users. Workers under the influence of drugs or alcohol are a danger to themselves and to other workers in their vicinity. Their impaired senses can easily cause an accident which results in injury to innocent individuals. This can be particularly true if the impaired worker is operating heavy or dangerous equipment. The Office of Applied Studies (OAS) analyzed results from a National Household Survey on Drug Abuse (NHSDA) from 2000 and 2001. Their analysis showed that on average 13.2% of the construction workers 18 and older had used at least one kind of illicit drug during the past month. Marijuana was by far the most used drug with 10.9% reporting such use in the past month. Illicit drugs other than marijuana were used by 4.2% of the workers in the past month. Alcohol use is an even more widespread problem among construction workers, with 43.2% reporting binge drinking during the past month. For the purposes of the survey, binge drinking was classified as consuming five or more drinks within a few hours at least one time during the past month. Heavy alcohol use was reported by 14.7% of the respondents. Heavy alcohol use was defined as binge drinking five or more times during the past month. The total percent of workers who had used alcohol during the past month was 63.3%. Younger workers aged 18 to 24 were more likely to abuse drugs and alcohol than their older counterparts, with the 18 to 24 age group leading them in every single category of substance abuse except for sedatives, where they trailed by .1% (SAMHSA 2009).

Twenty-six years ago drug testing was performed on less than 1% of all workers in the United States. In 1999, the Substance Abuse and Mental Health Services Administration (SAMHSA) released a report showing that roughly half of all workers were employed by companies that had implemented some form of drug testing. This large rise in drug testing was related to the increasing perception that drug use had a negative impact on both safety and productivity in the workplace (SAMHSA 2008).

There are a variety of drug testing programs companies can implement.

- Pre-employment – employees are required to take a drug test before beginning work.
- Periodic – all company employees undergo drug testing at regular intervals.
- Random – a portion or percentage of the employees are selected randomly to be tested (tested at varying intervals).
- For cause – employees who are witnessed behaving in a manner that suggests that they are under the influence of illicit drugs or alcohol can be directed to take a test based on this suspicion.
- Post-incident – employees that are involved in an incident are required to undergo a drug test, often regardless of fault.
- Follow-up – employees that have previously failed a drug test are tested before returning to work, often after completion of a substance abuse program or suspension.

Combinations of differing types of testing can be used for a more comprehensive program.

As previously mentioned, the effect drugs have on the senses of construction workers, particularly those wielding dangerous tools or equipment, is not one to be ignored. Studies have been conducted showing a decline in incident rates when drug testing had been implemented. A survey was conducted by Gerber and Yacoubian with the goal of evaluating the effectiveness of drug testing in construction. Companies that

implemented drug testing had an average incident rate of 8.79 two years prior to the implementation of drug testing, and this dropped to 4.61 one year after drug testing began. This is nearly a reduction of one half. The data from 1998 showed companies with drug testing programs had an incident rate of 4.15 compared to incident rates of 6.32 in companies which did not conduct drug testing (2001).

Of note is that within the same survey, employers were asked about why they implemented dug testing. The reason that was rated the highest was to "promote the safety of their workers," indicating perhaps that financial issues were either secondary to the health of their workforce or that no financial benefit was perceived. Financial impacts were also studied as contractor EMR data were collected by the researchers. Again, companies which drug tested fared better than those who did not. On average, EMR values fell from .973 in 1995 to .862 in 2000 after companies had implemented drug testing. Companies which did not have drug testing programs in place during the same time period saw their EMR values rise from .935 to .95 (Gerber and Yacoubian 2001).

Large companies were quicker to adopt drug screening policies, but smaller companies have been gradually catching up as the importance and benefits of drug testing have become more widely known. Indeed, studies show that the benefits of drug testing are not reserved only for large firms, but apply to smaller and specialty contractors as well. A study conducted by Hinze and Gambatese revealed that Florida roofing contractors that used for-cause drug testing had a median injury rate of 10.37. Those who did not test on a for-cause basis had a median injury rate of 39.29, nearly four times higher. Only four out of the 24 roofing contractors surveyed did not use for-

cause testing, demonstrating its prominence. The same study also examined mechanical contractors' regional divisions and also Nevada-based specialty contractors. Results from those groups reflected the same trend found in Florida. Mechanical contractors' divisions conducting post-incident testing had a median incident rate of 3.06. Those divisions that did not utilize post-incident testing reported a substantially higher rate at 13.33. Of the Nevada group, firms that drug tested had a median injury rate of 9.94, considerably lower than those that did not, which had a median injury rate of 14.71. Thus, within all three groups that were surveyed, the median injury rate of organizations that drug tested was lower than organizations that did not (2003).

Nature of the Firm

The size of construction firms has long been thought to affect safety performance. Chi et al. showed that there was a significant negative correlation between the number of workers employed by a firm and that firm's standardized mortality rate (2005). A study in 1997 showed a similar trend among Ontario construction companies. The study was conducted over a six year period, which covered both a robust economy and a recession, showed an increasing firm size to be associated with decreasing frequency of injuries. This trend was consistent throughout the span of the study (McVittie et al. 1997).

The reason for this difference is thought to be due to the larger firms having more financial resources and thus can afford improvements in employee safety. Additionally, smaller specialty contractors are usually performing more dangerous work such as steel erection, plumbing, roofing, and electrical distribution systems installation. The nature of their work is often inherently more dangerous than that of large general contractors.

Many trades have risks specific to their particular line of work. This is evident as the injury and fatality rates vary widely across different types of work. Baradan and Usemen analyzed injury data to determine a risk score for 16 different building trades. The risk score they developed took into account the potential for injuries as well as fatalities. The summary chart of the results of their risk analysis is shown in Figure 2-2. Trades that are more likely to be working in elevated positions generally have the worst risk ratings as their injury and fatality rates are high.

Worker Orientation

Since construction is a dangerous industry with many hazards unique to each project, new workers are particularly vulnerable to being injured. Workers who have been on a jobsite for weeks or months are much more likely to know their way around, know where dangerous areas, and know the best paths to avoid them. A new hire that is on a project site on the first day of employment will have none of this valuable knowledge. One way to mitigate this problem is to provide orientation for new workers. The orientation training can vary greatly in scope and length, which often will depend on who is providing it. General contractors may provide an orientation for all workers new to their project. In addition to this, the specialty trade contractors may provide an orientation more specific to their particular line of work. The orientation training should include information related to general hazards and hazards specific to the jobsite, minimizing the chance of injury due to those hazards, proper usage of personal protective equipment (PPE), proper usage of tools, and safety rules for that particular project.

One study explored the impacts of safety orientation training on plumbers and pipefitters in Ohio. A total of six employers were examined, representing an aggregate

of nearly four million worker-hours worked from 1996-1998. The analysis of the findings revealed that workers that were provided a safety orientation had an injury rate of 3.4%. Those who did not receive a safety orientation had an injury rate of 11.1% (Kinn et al. 2000).

Safety Professionals

Safety professionals also play an important role in construction safety. While all workers should have appropriate safety training, and as such, should conduct their work activities in a safe manner, infractions are common if this is not stressed on a daily basis. If workers understood and followed all safety guidelines, injury and fatality rates would be miniscule. The few incidents that would occur in such a scenario would likely have be the result of an unforeseeable and uncontrollable outside events. The reality is that the tradesworkers in construction often violate safety rules, whether they were established by OSHA or by stricter company or jobsite policies. As such, safety professionals are needed to monitor the workforce and identify and mitigate existing hazards. Such individuals can devote the full focus of their efforts towards ensuring a safe work environment. Firms with full-time and part-time safety directors have been reported to have superior safety performance records when compared with firms which do not employ safety directors (Hinze and Raboud 1988).

Financial constraints due to firm size or project size may make utilizing a safety director economically infeasible. In such cases, the foreman of a crew can assist in ensuring the safety of workers in the crew. Among mechanical contractors, those firms which had foremen conducting safety inspections benefited with substantial reductions in injury rates compared with organizations that did not require their foreman to conduct safety inspections (Hinze and Gambatese 2003). This situation is not ideal as the

foremen may sense a conflict of interest between maximizing production and ensuring that the jobsite is safe. However, it is a reality that some smaller firms, especially specialty trade contractors may not have the capability of hiring a safety professional.

Empirical Modeling

An accurate predictive model of safety performance could provide decision-makers within construction firms the data needed to justify the implementation of new safety policies. Such a model could also demonstrate where a firm's resources could best be applied to maximize the improvement in safety performance.

Empirical models are those that are developed on the basis of observations. In construction, empirical models are used for a variety of purposes including, but not limited to predicting costs (Wilmot and Mei 2005), productivity (Song and AbouRizk 2008; Zayed and Halpin 2005), and outcomes of claims (Chau 2007). Despite this, the use of empirical modeling techniques for predicting the recordable injury rates of construction firms is not well researched.

Flood and Issa reviewed empirical modeling with a focus on its use in the construction industry in 2009. Their work provides a basis for understanding the structure, development, evaluation, and validation of empirical models. Ideally, models should have three distinct sets of data. These include the following:

1. Fitting data – observations which are used for training the model on the patterns and relationships between the variables so that a prediction can be made.

2. Testing data – observations which are used to make judgments on the performance of models for the purposes of identifying the most accurate predictor.

3. Validation data – observations which are tested on the model that was selected during the testing phase to further confirm its accuracy and ensure that the selected model did not benefit from any bias in the testing data.

It is vital that the data cover the full scope of the problem that is being investigated. This is important as the model may only be valid within the range of the data that it has been utilized for training.

When testing and selecting a model, the performance should not merely be examined in terms of its overall performance. Each model's performance needs to be carefully evaluated across the entire scope of the problem. A model that performs the best overall, may fail miserably at predicting in a restricted area of the problem. If this is the case, the model cannot be selected as it may not be suitable for all of its intended applications.

Artificial neural networks are a type of empirical modeling. In simple forward-feed neural networks the information moves in one direction from input nodes to hidden nodes to output nodes. Alternatively, they can be described as using "nonlinear mapping between input vector and output vector via a system of simple interconnected neurons. It is fully connected to every node in the next and previous layer" (Chau 2007). A diagram of such a network is shown in Figure 2-3.

$$\frac{\text{Actual Primary Losses} + \text{Ballast Value} + \text{Weighting Value} \times \text{Actual Excess Losses} + (1 - \text{Weighting Value}) \times \text{Expected Excess Losses}}{\text{Expected Primary Losses} + \text{Ballast Value} + \text{Weighting Value} \times \text{Expected Excess Losses} + (1 - \text{Weighting Value}) \times \text{Expected Excess Losses}} = \frac{\text{Total A}}{\text{Total B}}$$

Figure 2-1. Experience modification rate calculation (Adopted from Everett and Thompson. Experience Modification Rating for Workers' Compensation Insurance, 1995)

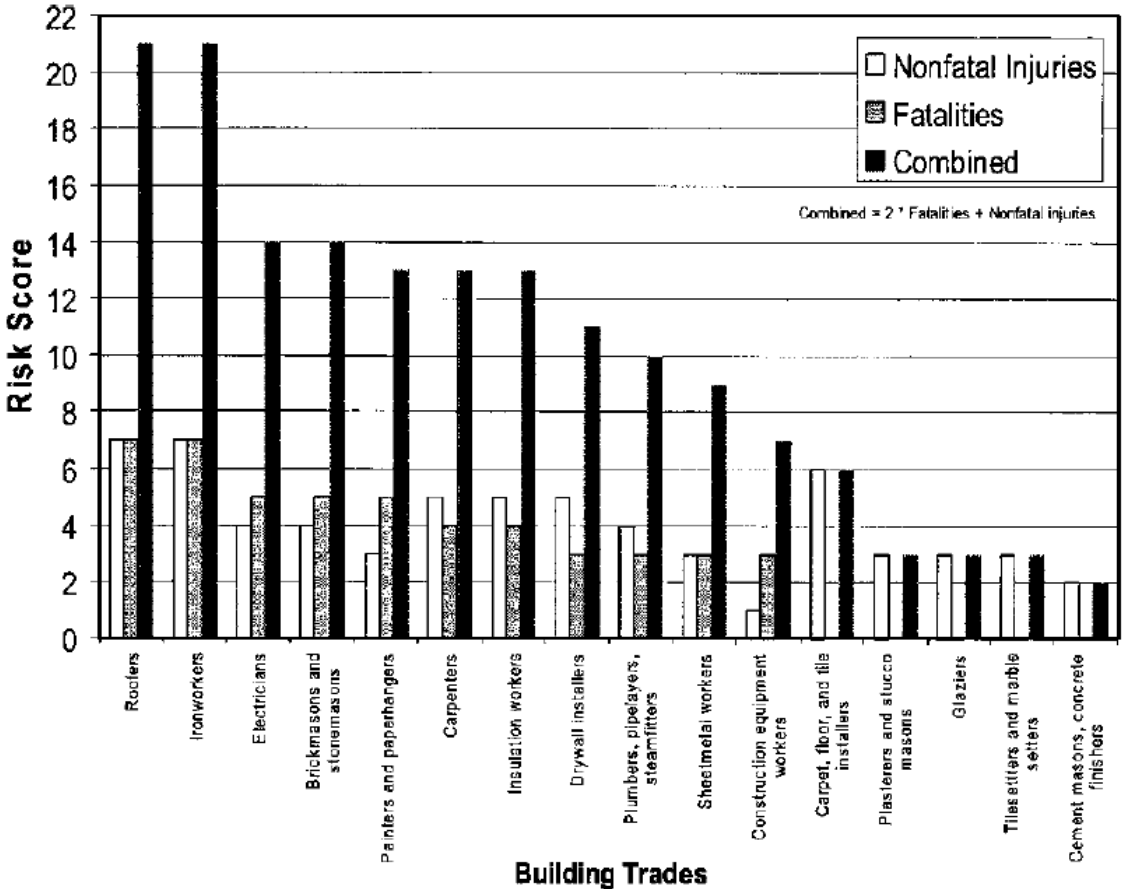

Figure 2-2. Risk scores of various building trades (Adopted from Baradan and Usmen. *Comparative Injury and Fatality Risk Analysis of Building Trades*, 2006)

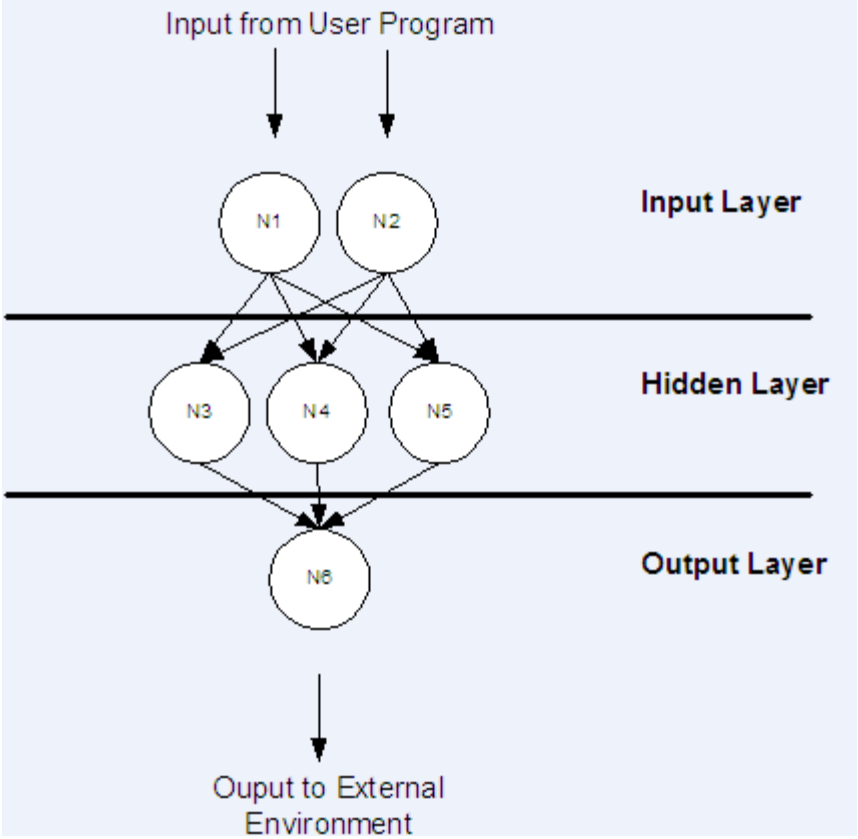

Figure 2-3. Structure of a feed forward artificial neural network (Adopted from Heaton. *A Feed Forward Neural Network,* 2007)

CHAPTER 3
METHODOLOGY

Introduction

The objective of this study is to develop a model which will predict the recordable injury rate of specialty contractors. The study also aims to explore the use of artificial neural networks as a tool for developing this model. Two other methods of modeling will also be explored to determine if the effort expended in using the artificial neural network is worthwhile. Multivariate linear regression will be used along with linear regression of a "scorecard" that this researcher developed based on his interpretation of the data.

Data Collection

Data for this study was collected from a previous master's thesis completed at the University of Florida's Rinker School of Building Construction. Thomas Feronti's thesis, "Construction safety practices of specialty contractors" published in 2006 under the advisement of Dr. Jimmie Hinze, is the source of the data. Feronti's research involved the distribution of a survey to small and medium-sized (< $100 million of annual volume) specialty contractors to examine their safety practices and to evaluate their association with recordable injury rates.

There were 131 responses to the survey. Not all respondents provided all information necessary for this research, so some missing data did exist in the responses. Unfortunately, some contractors did not know, or refused to provide their RIR values. Others did not respond to questions which were used as inputs for the model. Additionally, some respondents reported RIR values of zero. These were treated as outliers and were excluded from the analysis. Those companies which

reported having an RIR of zero were included if they also reported their RIR for five years ago. This was done to prevent outliers from confusing the models. Some company responses were deemed nonresponsive for the purposes of this research and were eliminated from further analysis. The final resultant size of the sample was reduced to 38.

Selection of Variables

Variable selection was performed after paring down the data. The researcher determined that RIR was a superior measure of safety performance compared to the EMR. This is because the EMR will not reflect recent safety policy changes. The RIR is calculated on an annual basis and reflects only the previous year, as opposed to EMR which takes into account several years, excluding the immediately preceding year. As such, the RIR was used as the output value in all models.

Inputs were selected based on a sensitivity analysis of the data. Scatter plots and correlation coefficients were examined to determine the impact of each potential input on the RIR. The sensitivity of the data was deemed the most important, with secondary consideration being given to how well the various areas of each problem were represented. For example, an input with a binary “yes or no” response may have a strong correlation with the output. However, if there are 37 responses of “yes” and just one response of “no” this input is not very useful as the impact will not be very well understood by the model.

After this analysis was conducted, seven factors were identified for use in the model:

- Annual volume – continuous variable
- Number of full-time field workers – continuous variable

- Percent of work which is subcontracted out – continuous variable
- Does the firm drug test – binary variable
- Does the firm provide orientation for all workers – binary variable
- Does the firm hold weekly toolbox talks – binary variable
- Percent of time a full-time safety director is in the field – continuous variable

Model Development

Three separate sets of inputs were identified, with each model being tested with all sets. Three different sets were used as some inputs with weaker or had lesser understood effects on the RIR and could be left out in subsequent models. Of particular concern are the binary inputs. It was quite easy for the binary inputs to skew the results of the poorer performing contractors in the model. By having multiple input sets, the impact of each input on the model can be better understood, and changes in accuracy can be observed. The three sets of inputs used were as follows:

1. Set one
    a. Annual volume,
    b. Number of full-time field workers
    c. Percent of work subcontracted
    d. Does the company drug test
    e. Does the company hold weekly toolbox talks
    f. Does the company provide orientation for all workers
    g. Percent of time a full-time safety director spends in the field
2. Set two
    a. Annual volume
    b. Number of full-time field workers
    c. Percent of work subcontracted
    d. Does the company hold weekly toolbox talks
    e. Does the company provide orientation for all employees
    f. Percent of time a full-time safety director spends in the field
3. Set three
    a. Annual volume
    b. Number of full-time field workers
    c. Percent of work subcontracted
    d. Does the company hold weekly toolbox talks
    e. Percent of time a full-time safety director spends in the field

Safety Scorecard

The “scorecard” of safety performance was the first model developed. This was a subjective method of modeling safety performance. This researcher carefully examined the scatter plots of each input variable and the corresponding RIR. From this, judgments on the strength of the input’s impact were made. Additionally, by looking at the scatter plots showing the entire range of data, the researcher saw outliers and made attempts in the development of the scorecard to mitigate them.

The scorecard simply awarded points, ranging from one to three, based on meeting a certain criteria for each input. The following awards were developed based on the researcher’s analysis of the data:

- $50 million or more in annual volume – three points
- 25 or more full-time field workers – two points
- 50% or more of the work subcontracted – one point
- The company conducts drug tests in some form – one point
- The company holds weekly toolbox talks – two points
- The company provides orientation for all workers – two points
- The company has a full-time safety director that spends 50% or more of their time in the field – two points

The total score of each company was calculated and paired with its RIR to develop a simple one variable regression model. The intercept and coefficient of the score variable were found using the regression function in the data analysis pack in Microsoft Excel. The score and RIR were used as x and y-values, respectively. The predicted value was determined by using the following formula:

$$\text{Predicted value} = \text{intercept} + (\text{total score} * \text{coefficient})$$

The total score encompassed all variables, and as such allowed for use of the one variable regression technique.

Multivariate Linear Regression

A model utilizing multivariate linear regression was subsequently developed. This model was also developed using the regression function in the data analysis pack of Microsoft Excel. All variables in the input set were fed into the regression function as x-values, with the RIR as the y-value. The predicted value was calculated using the following formula:

$$\text{Predicted value} = \text{intercept} + (x_1 * \text{x coefficient}_1) + (x_n * \text{x coefficient}_n)\ldots$$

The intercept and coefficients were derived from Excel's analysis.

Artificial Neural Network

The artificial neural network was constructed using NeuroSolutions software, which was developed by NeuroDimension, Incorporated. NeuroSolutions was chosen for its ease of use. Gaining insight into the inner workings and processes of artificial neural networks is outside the scope of this study. NeuroSolutions provides an add-in for Microsoft Excel allowing for the quick selection of the various data sets. It also provides a network builder which creates a network based on the intended use.

In this case, the artificial neural network structure was built based on the manufacturer's recommendation for a function approximation network. The basic structure was the set of inputs, three perceptrons, and then the single output. Back-propagation was used to aid in developing the weights for proper learning. The number of epochs varied according to the report of mean squared error of the training and cross-validation set. Over-learning was a major issue that required mitigation. Carefully controlling the number of epochs helped to prevent this. When a large number of

epochs were used, especially on this small sample of data, the neural network moved from learning from the data to simply memorizing the points. The cross-validation set was used during the training phase to determine the appropriate number of epochs. Mean squared error of the cross-validation and training sets determined when to cut-off training at a given epoch. The range of epochs used in training ranged from 13 to 250. The length of time required to train the model was negligible for all models.

Training the Models

In order to compensate for the small sample size, four examples (~10%) were withheld from each set of data as testing data. This was repeated three times for each model, for each set of inputs, with a different four examples being withheld each time. This allowed the researcher to triple the testing data, while keeping the amount used in training the same.

An example of this is that for input set one, the multi-variable regression model was tested using cases 5, 12, 21, and 33. The rest of the data were reserved as training data. Those four reserved testing cases previously mentioned would then be tested by comparing the actual RIR versus what the model predicted their RIR would be based on the inputs. After this was completed, the model was retrained. This time, however, cases 10, 20, 30 and 36 were withheld as testing data, with the rest being used as training. The same testing procedure was repeated. Finally, a third iteration was conducted with another set of data, namely cases 7, 13, 18, and 35.

The testing cases were not selected at random. They were selected to ensure that the full spectrum of the problem was covered. The case numbers were assigned by the researcher in ascending order, thus the lower the case number, the lower the RIR. Thus, for the sample size of 38, each testing set has a good spread of RIR values.

It is vital that the model be able to perform across the entire spectrum of the problem. If cases are not selected with a good spread of RIR, the most accurate model after testing may be one that is simply good at predicting within the small range of the cases tested.

For the artificial neural network (ANN), a slightly different approach was used. The testing data that were used were the same as for the other models; however, three examples from the training set were withheld each time as cross-validation data. This allowed the researcher to quickly determine if the model was over-learning during the training process. Thus, while the testing set was the same for each model, the training set was slightly different in the artificial neural network.

Selecting the Best Model

Selection of the best model will be based upon minimization of the standard error and the highest correlation coefficient. The models will be evaluated within each input set and overall. Each input set will also be evaluated in comparison to the other two to determine which one resulted in the most accurate models.

CHAPTER 4
RESULTS

Introduction

Results of the modeling are provided in this chapter. Graphical representations of the aggregated predicted RIR versus the actual RIR for each model, for each input set are shown. Their accuracy will be presented and the best model will be chosen. Some models predicted RIR values of less than zero, an obvious impossibility. This underestimation was not adjusted to zero, but was left alone for the purposes of evaluating accuracy.

Input Set One

The scorecard performed relatively well with input set one. The standard error was 4.01, the lowest among all models in the input set. Given that the scorecard's weights and requirements for points were specifically tailored to each input, it is not surprising that it would perform relatively well with the complete set of inputs. Figures 4-1 and 4-2 show the actual and predicted RIR line graphs and the scatter plot of predicted versus actual RIR, respectively.

The multivariate linear regression model followed a pattern similar to the scorecard. It overestimated the RIR values in the middle range and underestimated it at the higher values. Overall, it had the strongest coefficient of correlation.

The neural network model also showed a trend of overestimating the RIR of the middle-range values while underestimating the higher values. It performed well at predicting the lowest two values, but severely deviated from the actual RIR on the next four values.

Table 4-1 summarizes the results of the testing for each model. The multivariate linear regression had the strongest correlation, and the scorecard had the lowest standard error. The ANN had a much lower correlation coefficient and higher standard error. Given the data found in Table 4-1, it is determined that the scorecard regression produced the most accurate results.

Table 4-1. Standard errors and correlation coefficients for input set 1

	Scorecard Regression	MV Regression	ANN
Standard Error	**4.012138**	4.337107	4.705738
Correlation Coeff.	0.761556	**0.776082**	0.613545

Input Set Two

Input set two was virtually the same as set one, except without the drug testing input. Overall, the performance improved compared to the previous input set. The scorecard's performance degraded. The same pattern that was observed previously with the scorecard regression is seen again with the overestimation in the middle of the range and underestimation at the higher RIR values.

The multivariate linear regression model's performance improved as a result of removing the drug testing input. This model had the lowest standard error and highest correlation coefficient out of the group for input set two. It did especially well at mimicking the trend, except for the last three RIR values.

The ANN model also improved over its performance on the previous input set. The predictions of the ANN were the most accurate, especially at the extremes. It steadily over-estimated the middle RIR values.

The standard errors and correlation coefficients for all three models are found below in Table 4-2. The ANN showed a great improvement, but the multivariate linear

regression had the best result with respect to its correlation coefficient and standard error. As such it is selected as the best predictor for input set two.

Table 4-2. Standard errors and correlation coefficients for input set 2

	Scorecard Regression	MV Regression	ANN
Standard Error	4.66865	**4.219175**	4.231964
Correlation Coeff.	0.666976	**0.772946**	0.763195

Input Set Three

The final set of inputs contains all continuous variables and one binary variable. Given the reduction in inputs, the scorecard model performed poorly at predicting the outer values. This occurred because as the number of inputs declined, the number of contractors with a total score of zero rose. It is visible on the line graphs in Figures 4-13 and 4-14 that after just a few examples the prediction essentially “flat-lines”. All of those values along the horizontal line had a total score of zero in the model. As such, they all were assigned the intercept value of the regression analysis from their particular training set. Despite this, the correlation coefficient actually increased over the previous set of inputs and the standard error fell.

Performance of the multivariate linear model degraded as well, but not by a very large amount. Again, the trend continued for overestimating the middle range of RIR values and underestimating the larger RIR values.

The ANN’s accuracy increased greatly with two less binary variables compared to the first set of inputs. Based on this, it seems that the ANN was confused by the binary variables. Figures 4-17 and 4-18 show a relatively close set of predictions with only a few exceptions.

The ANN performed the best with the third and final set of inputs. While the correlation coefficient was not the best of the group, its standard error was substantially lower than that of the scorecard and multivariate linear regression models'.

The researcher also explored using this model on only those firms with less than 16 RIR. The reason for this being that there appeared to be disproportionate errors at the higher range. The result was the model that produced a standard error of 2.96, lower than the original model, but as a proportion of range of RIR it was higher. This exercise was merely exploratory, and the reduced sample model was not compared with other models in selection of the best predictor.

Table 4-3. Standard errors and correlation coefficients for input set 3

	Scorecard Regression	MV Regression	ANN
Standard Error	4.564484	4.361675	**3.98072**
Correlation Coeff.	0.724305	**0.762012**	0.695456

The best performing individual model was the ANN with input set three. The dominance of continuous variables helped to reduce some of the inaccuracies that it experienced with the other two input sets. It had the lowest standard error out of all models with any input set.

As this ANN model and input set had the lowest standard error, the relationship between the inputs and the error were explored. The researcher hoped to gain insight into what was causing the errors so that these factors could be mitigated. Given the small sample size, and thus the small testing set, it is difficult to identify problem areas conclusively. However, firm size and errors are reasonable at the lower and upper third of the range, but are substantially higher in the middle of the range. This trend is evident when firm size is measured as annual volume or the number of full-time field

workers employed. These errors may arise from the fact that there is great variability of RIR among companies in the middle range of firm size. In general, large companies have good safety performance and smaller companies have poor safety performance. The companies in the middle vary between these two extremes. As such, it may be difficult for the ANN to understand the relationship in that section of the range.

The other three inputs, percent of work subcontracted, weekly toolbox talks, and percent of time safety director spends in the field showed a common trend. Most firms self-performed 100% of the work, did not hold toolbox talks, and did not have a safety director, or at least did not have one who spent time in the field. There was lots of variability of the RIR for these same input values (0% subcontracted, no weekly toolbox talks, and safety director spends 0% of their time in the field). As such, the ANN had larger errors at these inputs values compared to the rest of the inputs. It is important to note that while percent subcontracted and percent of time safety director spends in the field are continuous variables, the inputs were generally confined to only a few different values across the range. For the testing set, percent subcontracted was essentially a binary variable, given that all companies either had subcontracted out 0% of 100%. Percent of time safety director spends in the field did not have the same magnitude of the problem, but the issue was similar. Responses were usually estimates such as 25%, 50%, 75%, etc. Thus, while technically, the input for percent of time the safety director spends in the field could fall anywhere from 0% to 100%, in practice the responses reflected a rounding of the actual value to the nearest 5%, 10%, or even 25%. This may partially explain the error associated with this input.

Summary

When examining the performance of the different types of models across all input sets, the multivariate linear regression model was the best performer on average, with the lowest standard error and the highest coefficient of correlation. It consistently held high correlation throughout testing, with each set of inputs yielding no coefficient of correlation less than .762. On average over the three input sets, its standard error was narrowly better than that of the ANN and substantially better than that of the scorecard.

The input sets performed similarly on average, with each having a standard error of greater than 4.3 and less than 4.4. Likewise coefficients of correlation had a relatively tight range from .717 to .734. Each set of inputs catered differently to the models, and as such the degradation of the accuracy of one model in another set of inputs was offset by the increase in accuracy of another model, resulting in a reasonably tight pattern of mean accuracy between the three sets of inputs.

Overall, the ANN with input set three performed the best out of all models, with the lowest standard error. The ANN's performance steadily increased as binary variables were dropped from the input set. While some patterns can be seen in the errors, it is not possible to infer much from this as the sample size is so small. A much larger sample size showing the same trends in errors would be much more conclusive.

Table 4-4. Summary table showing standard errors and correlation coefficients across all models and input sets

Input Set	Measurement	Scorecard Regression	MV Regression	ANN	Mean
1	Standard Error	4.012138	4.337107	4.705738	4.351661
	Correlation Coeff.	0.761556	0.776082	0.613545	0.717061
2	Standard Error	4.66865	4.219175	4.231964	4.373263
	Correlation Coeff.	0.666976	0.772946	0.763195	0.734372
3	Standard Error	4.564484	4.361675	3.98072	4.302293
	Correlation Coeff.	0.724305	0.762012	0.695456	0.727258
Mean	Standard Error	4.415090667	4.305985667	4.30614067	
	Correlation Coeff.	0.717612333	0.770346667	0.690732	

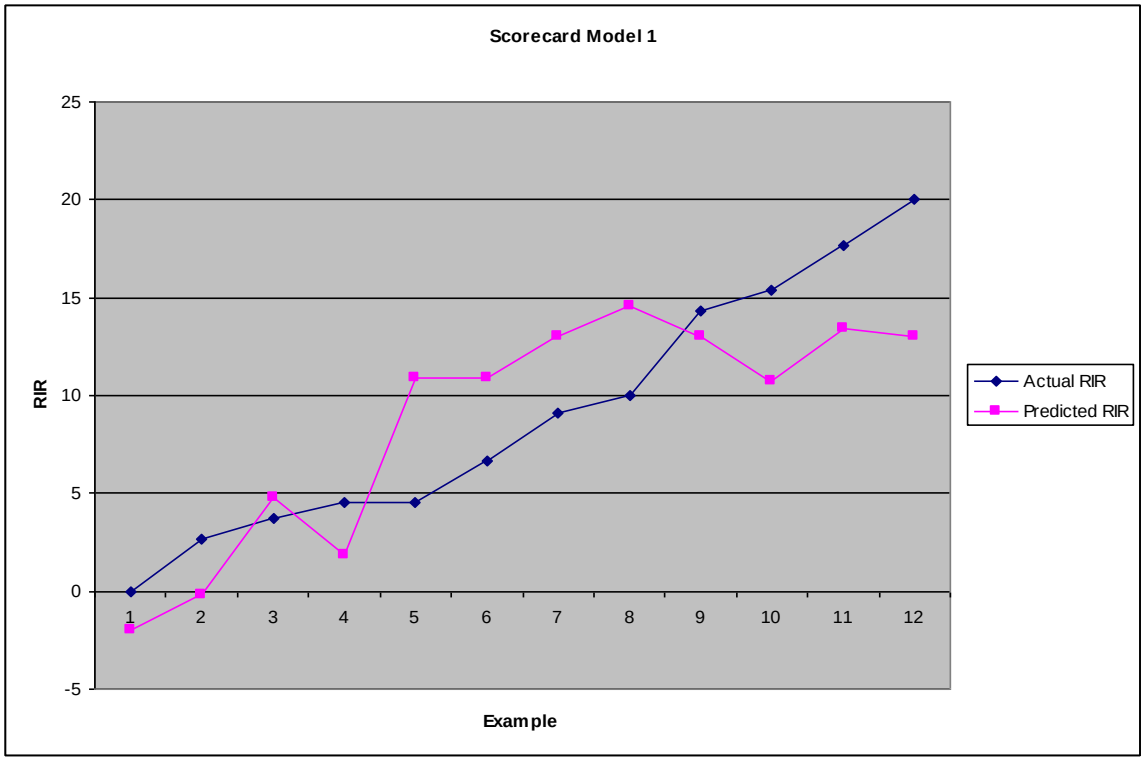

Figure 4-1. Actual RIR and scorecard 1 predicted RIR for each example

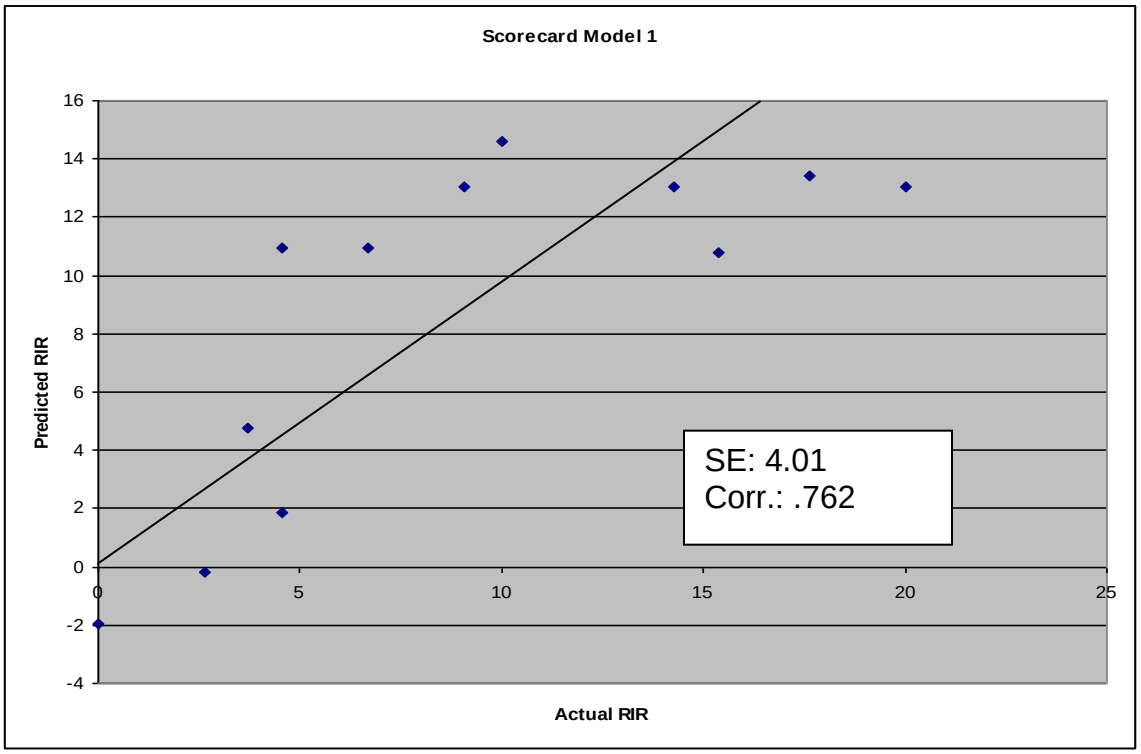

Figure 4-2. Actual RIR vs. scorecard 1 predicted RIR

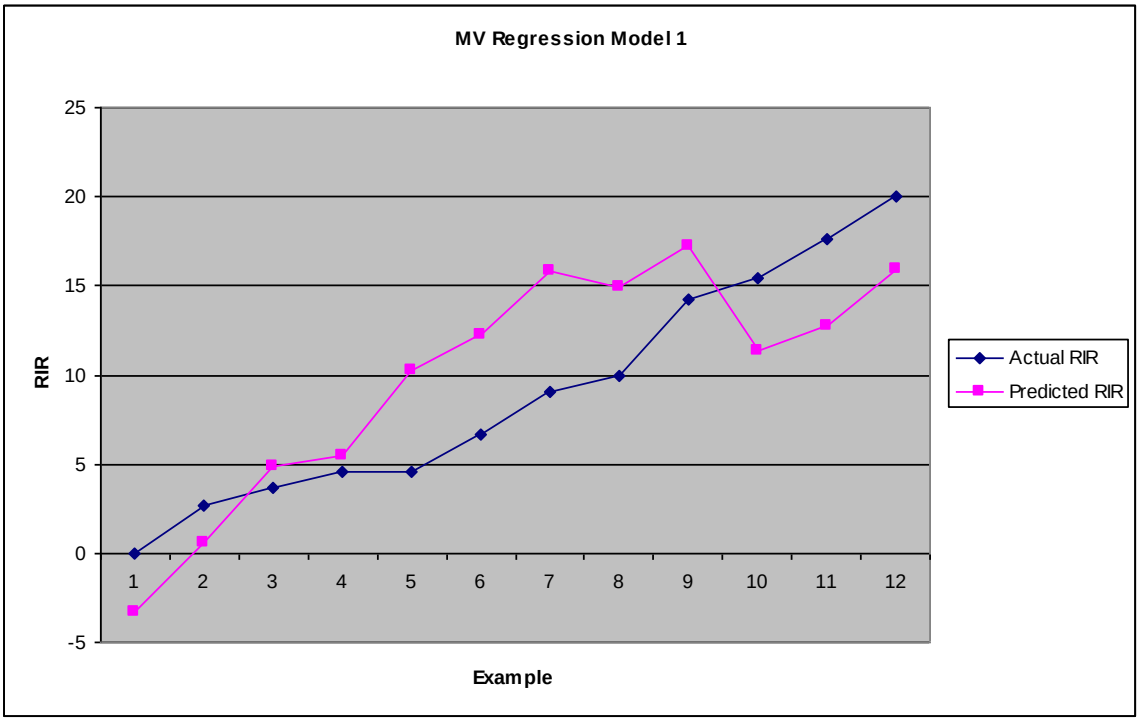

Figure 4-3. Actual RIR and multivariate linear regression 1 predicted RIR for each example

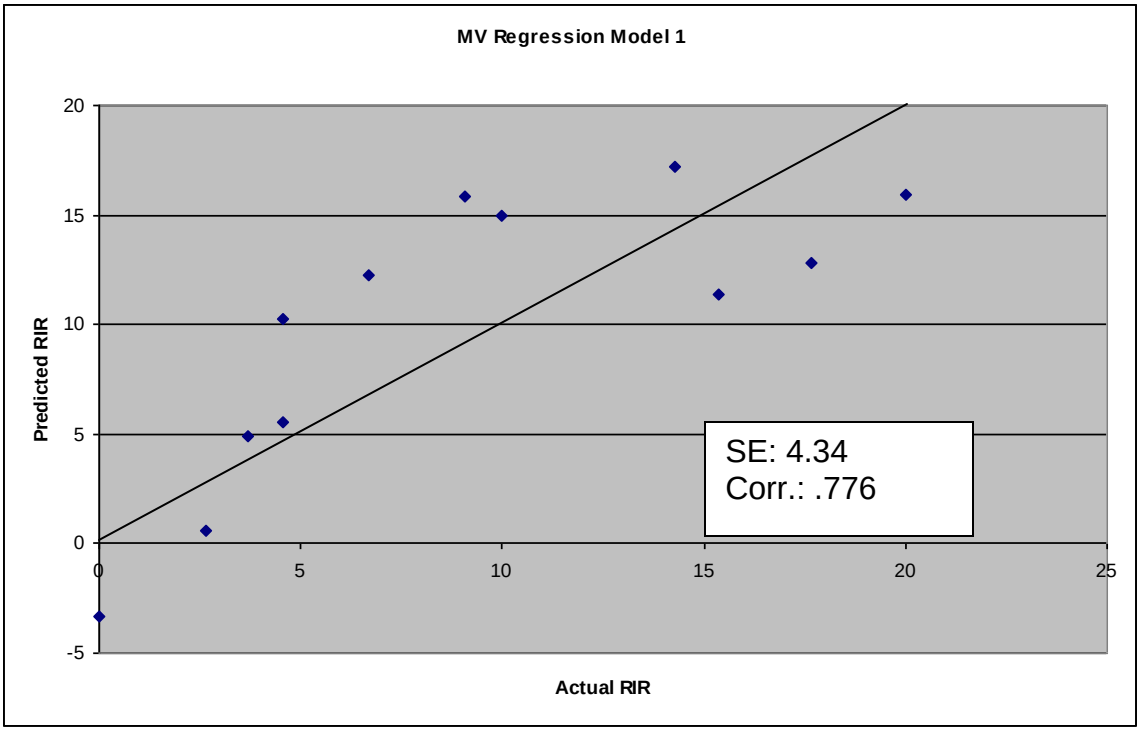

Figure 4-4. Actual RIR vs. multivariate linear regression 1 predicted RIR

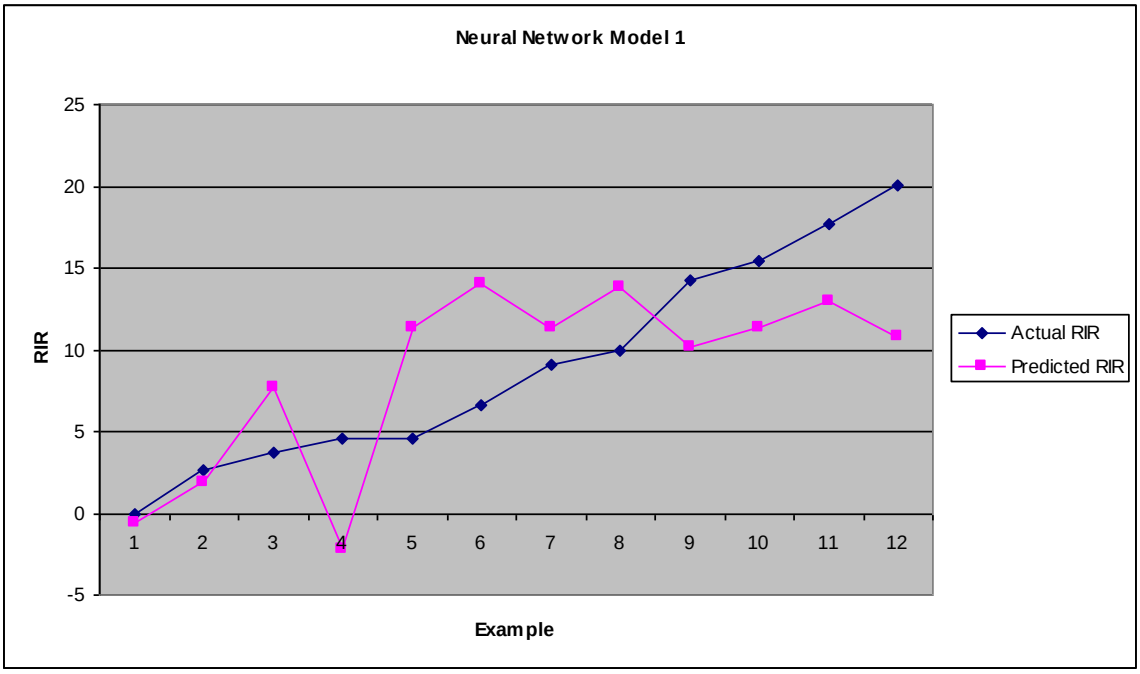

Figure 4-5. Actual RIR and ANN 1 predicted RIR for each example

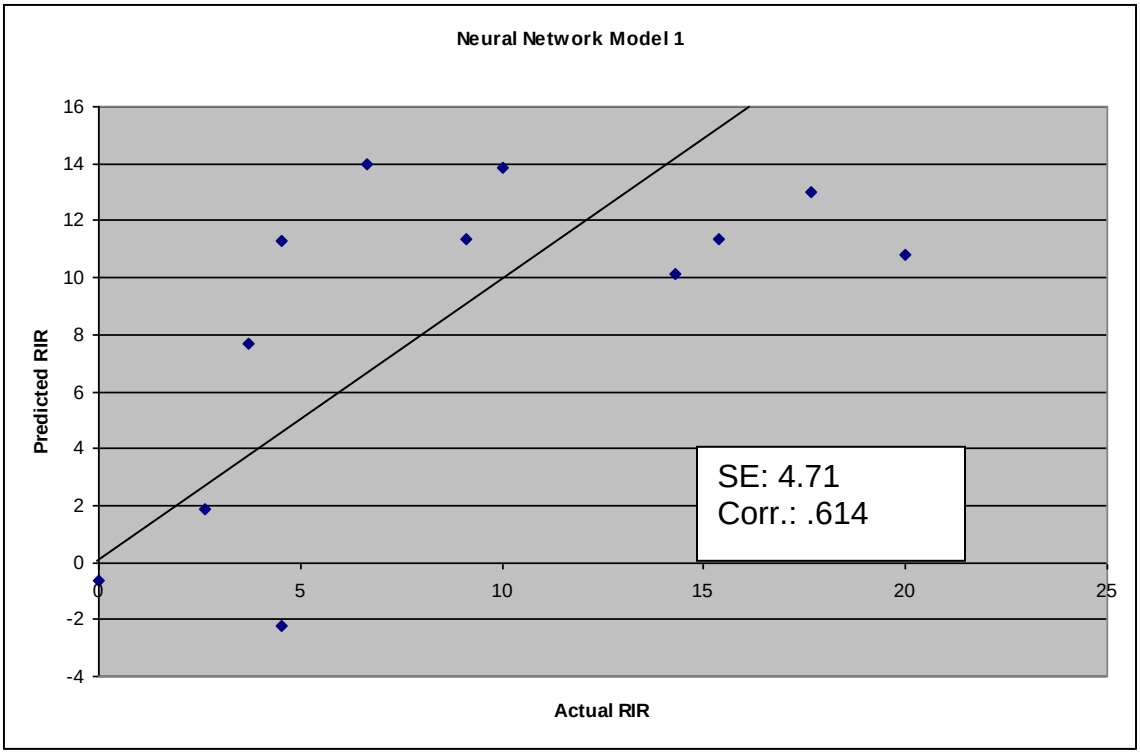

Figure 4-6. Actual RIR vs. ANN 1 predicted RIR

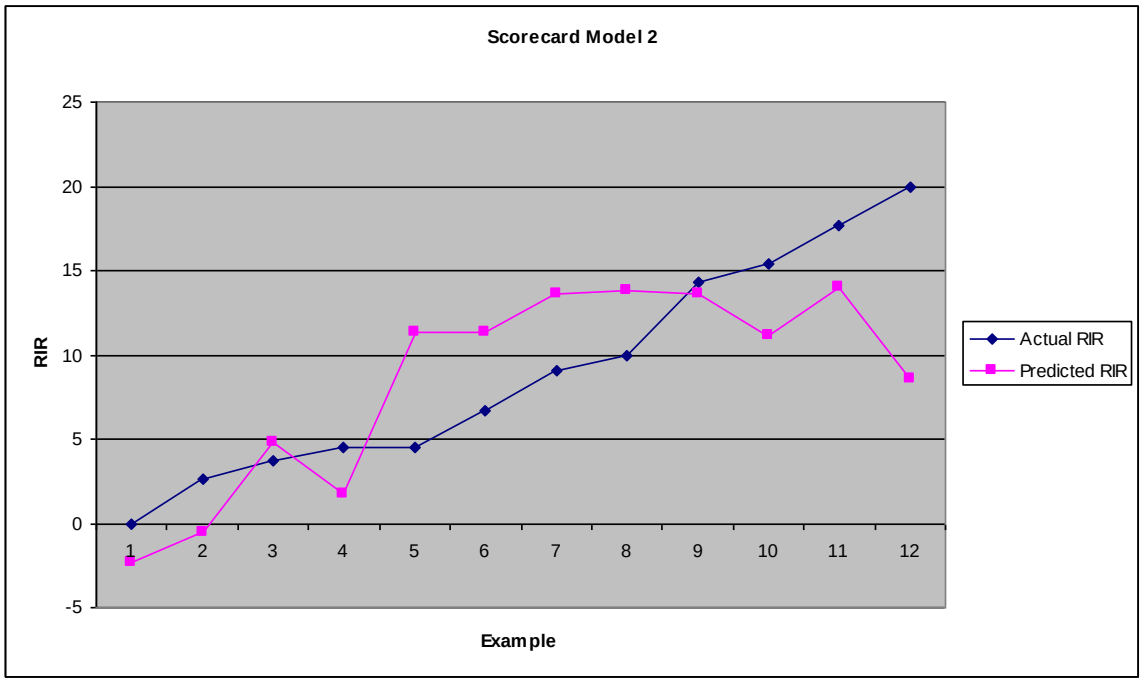

Figure 4-7. Actual RIR and scorecard 2 predicted RIR for each example

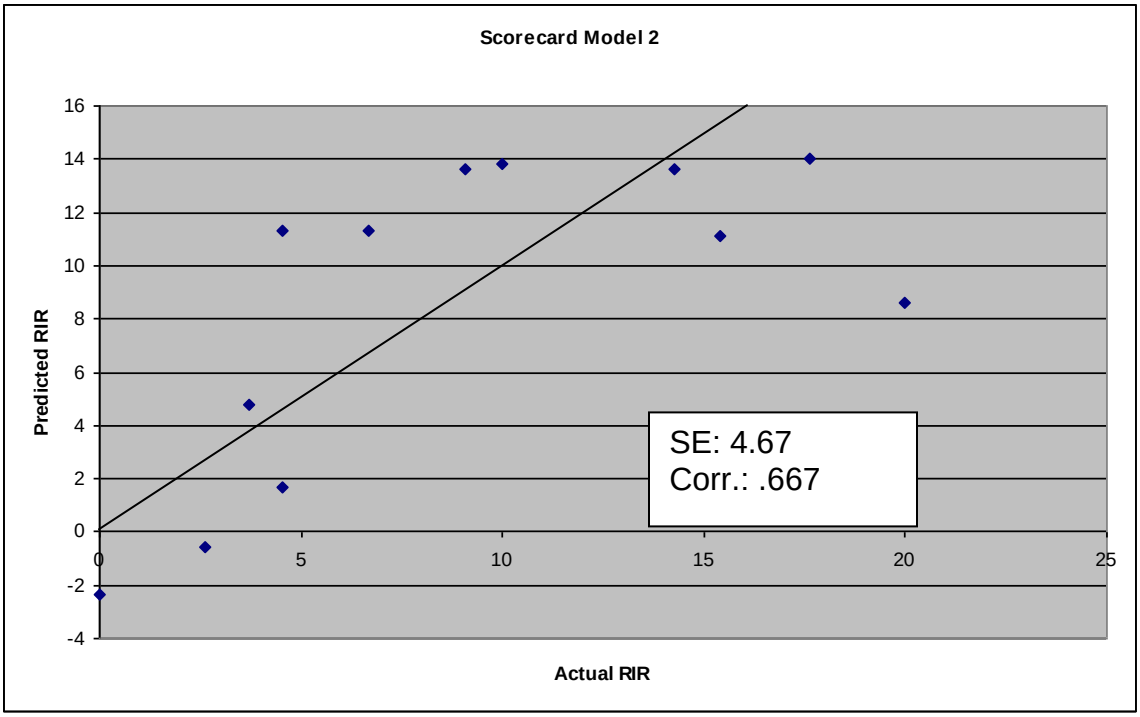

Figure 4-8. Actual RIR vs. scorecard 2 predicted RIR

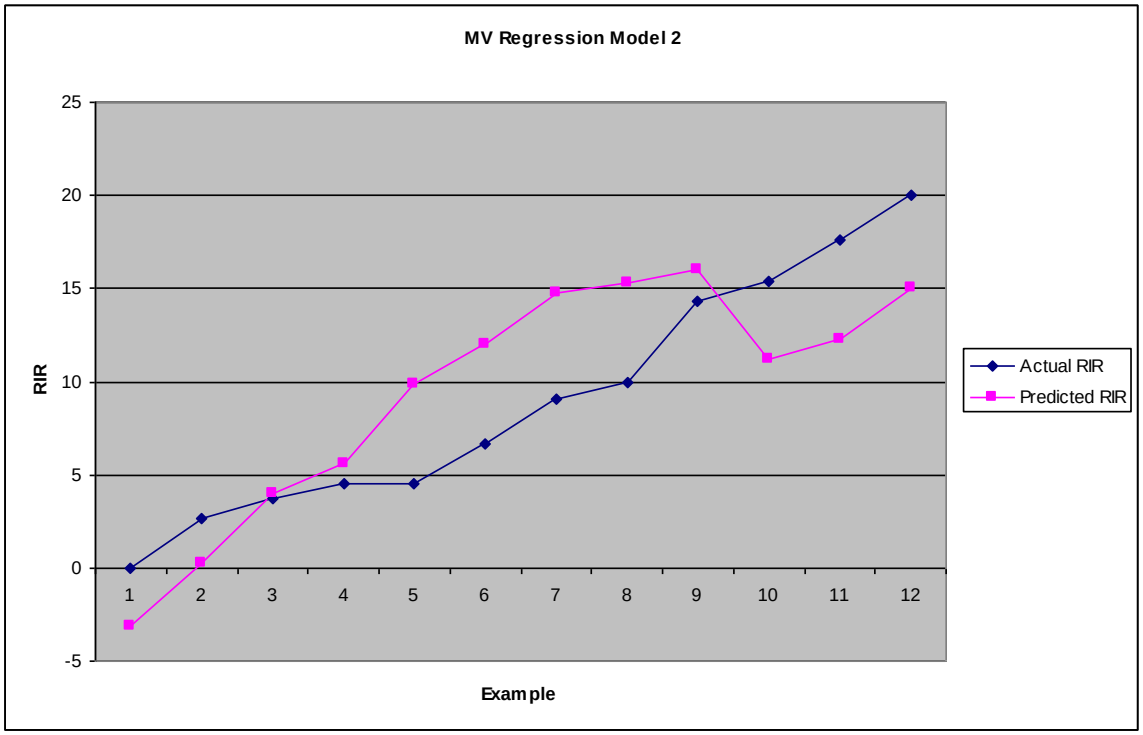

Figure 4-9. Actual RIR and multivariate linear regression 2 predicted RIR for each example

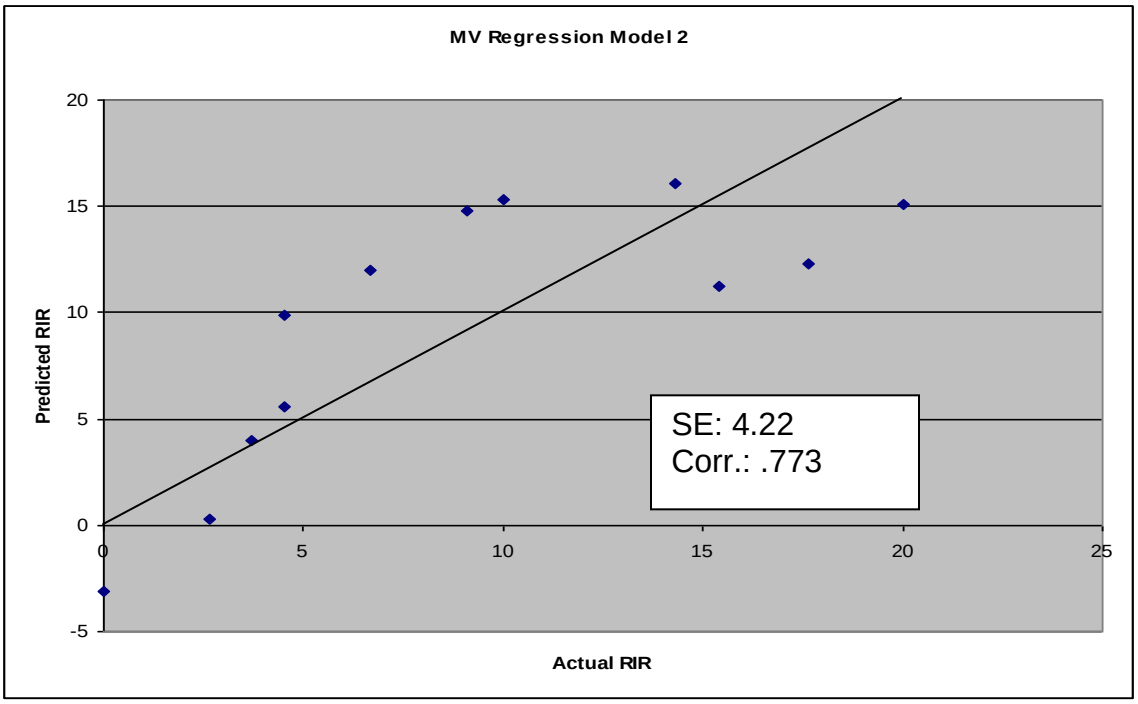

Figure 4-10. Actual RIR vs. multivariate linear regression 2 predicted RIR

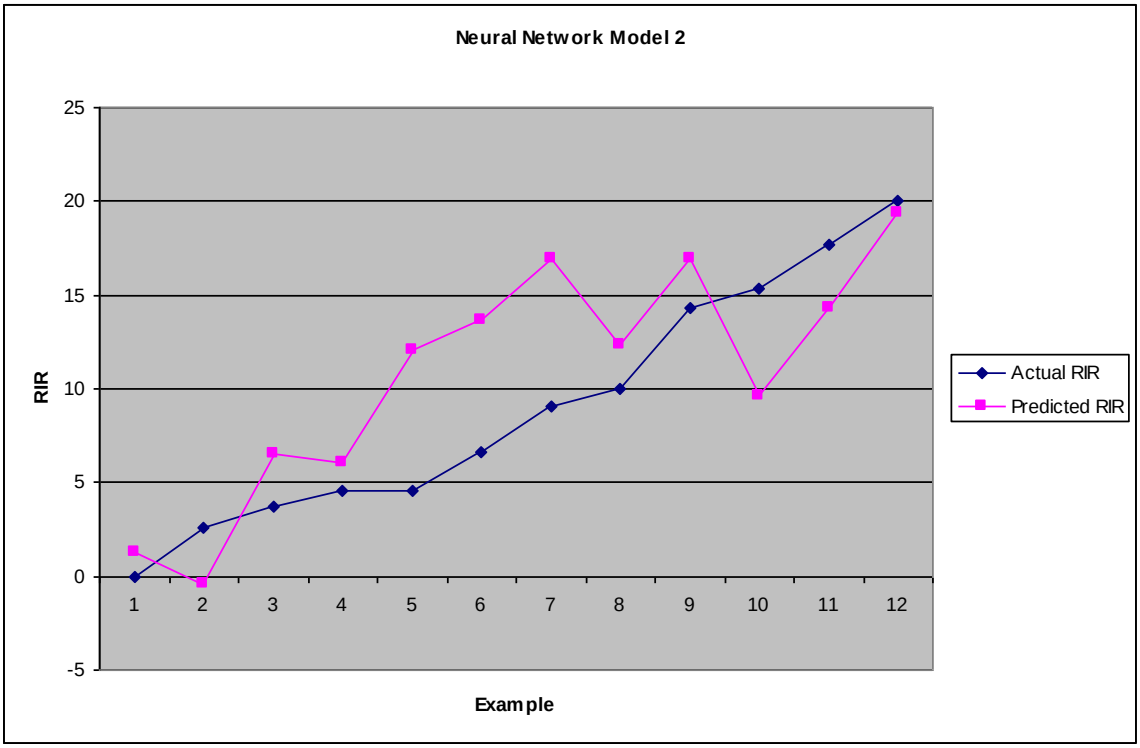

Figure 4-11. Actual RIR and ANN 2 predicted RIR for each example

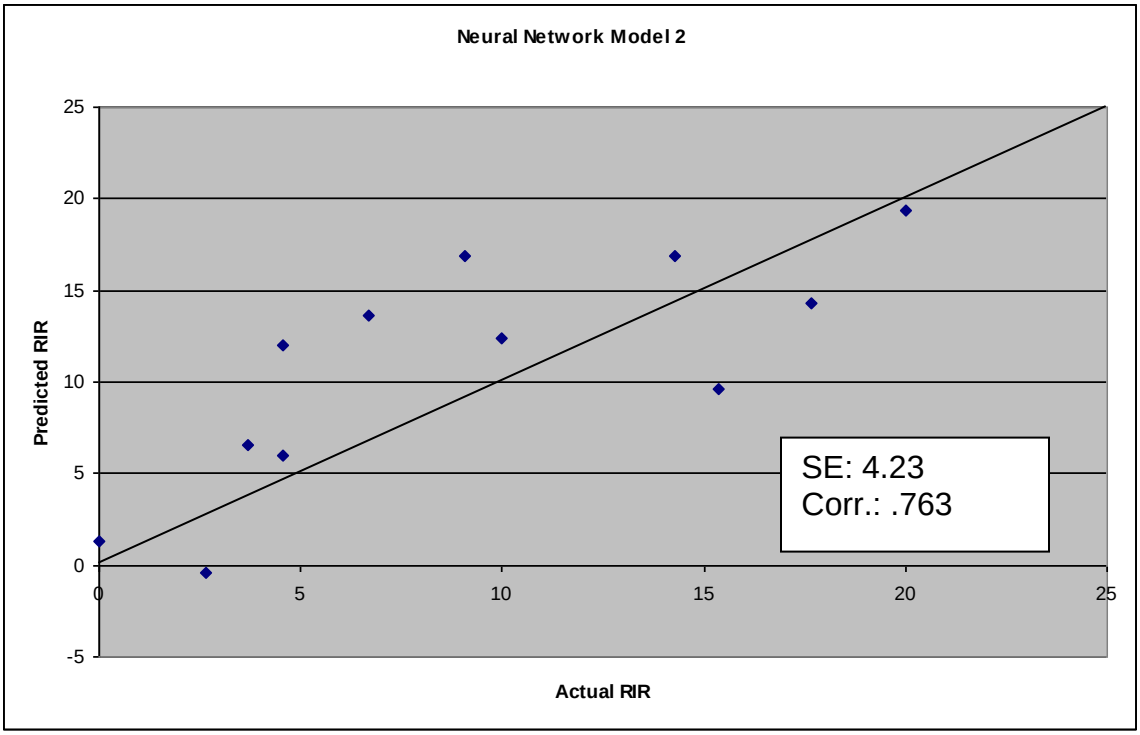

Figure 4-12. Actual RIR vs. ANN 2 predicted RIR

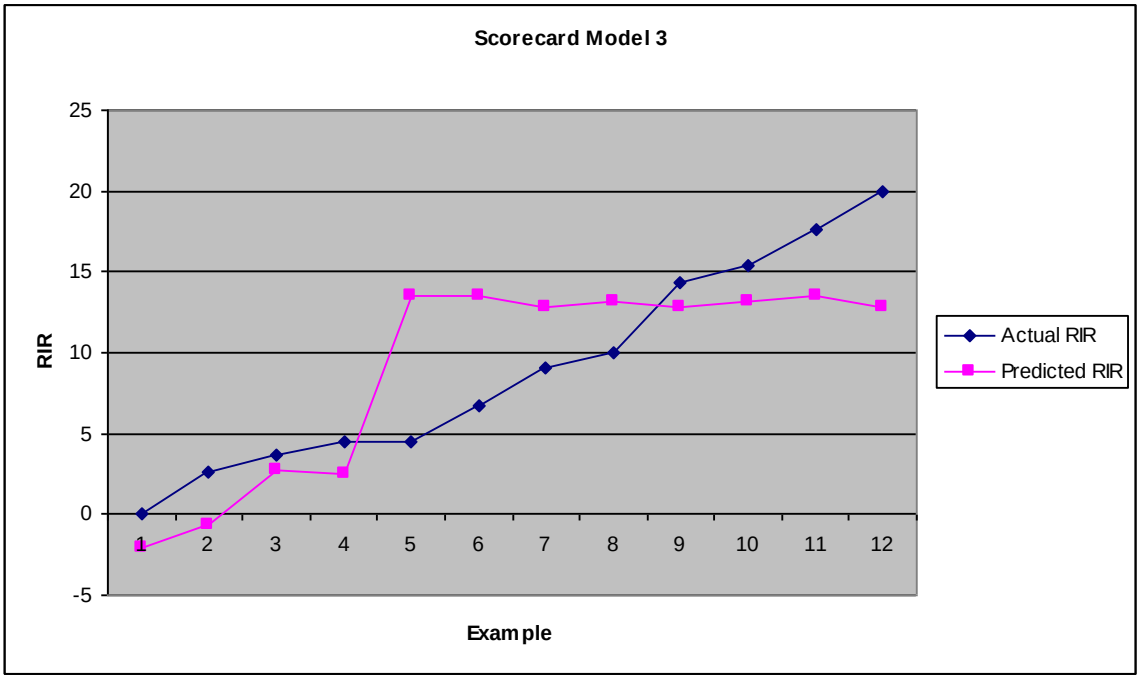

Figure 4-13. Actual RIR and scorecard 3 predicted RIR for each example

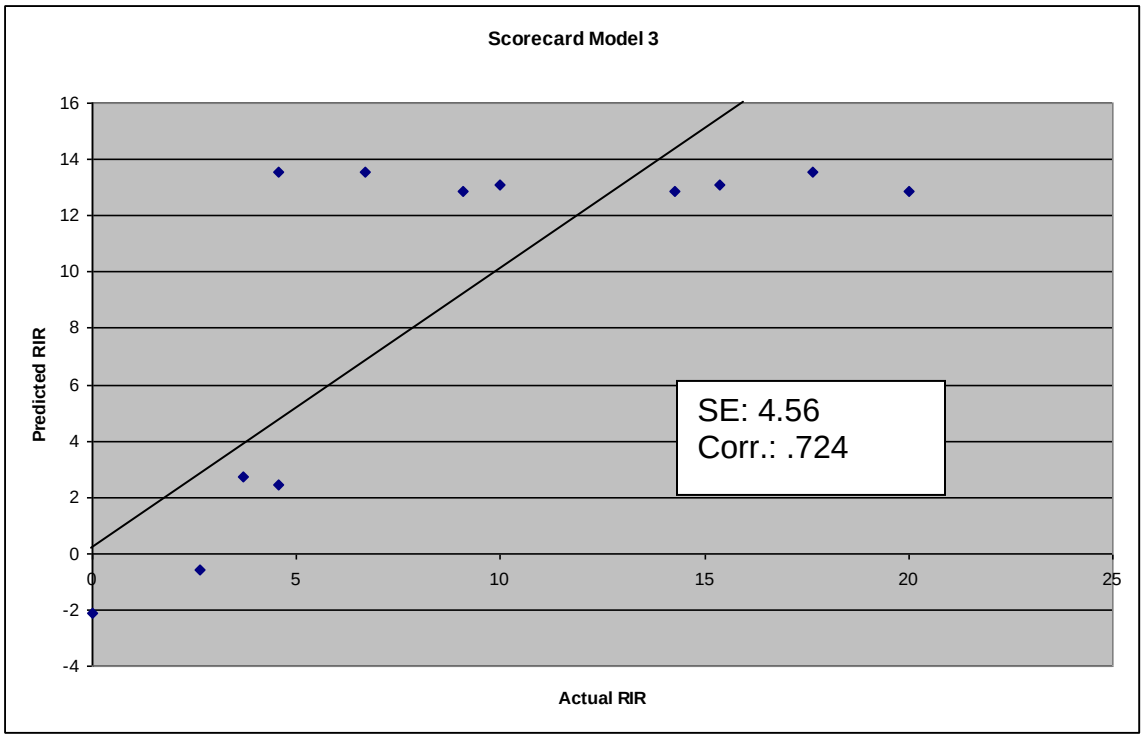

Figure 4-14. Actual RIR vs. scorecard 3 predicted RIR

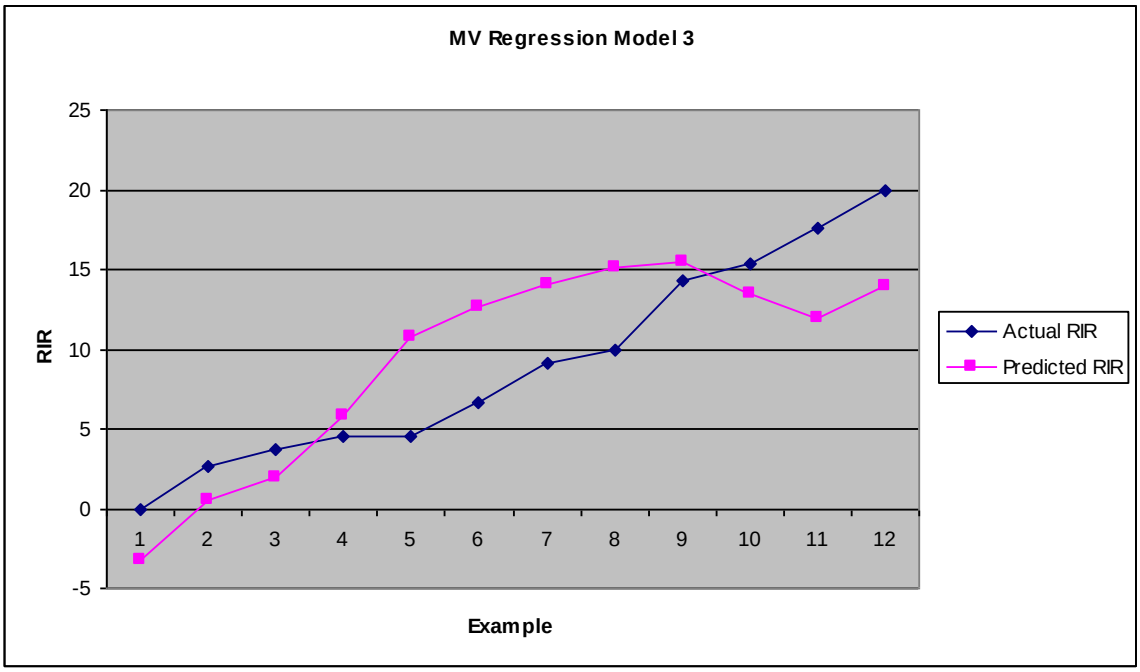

Figure 4-15. Actual RIR and multivariate linear regression 3 predicted RIR for each example

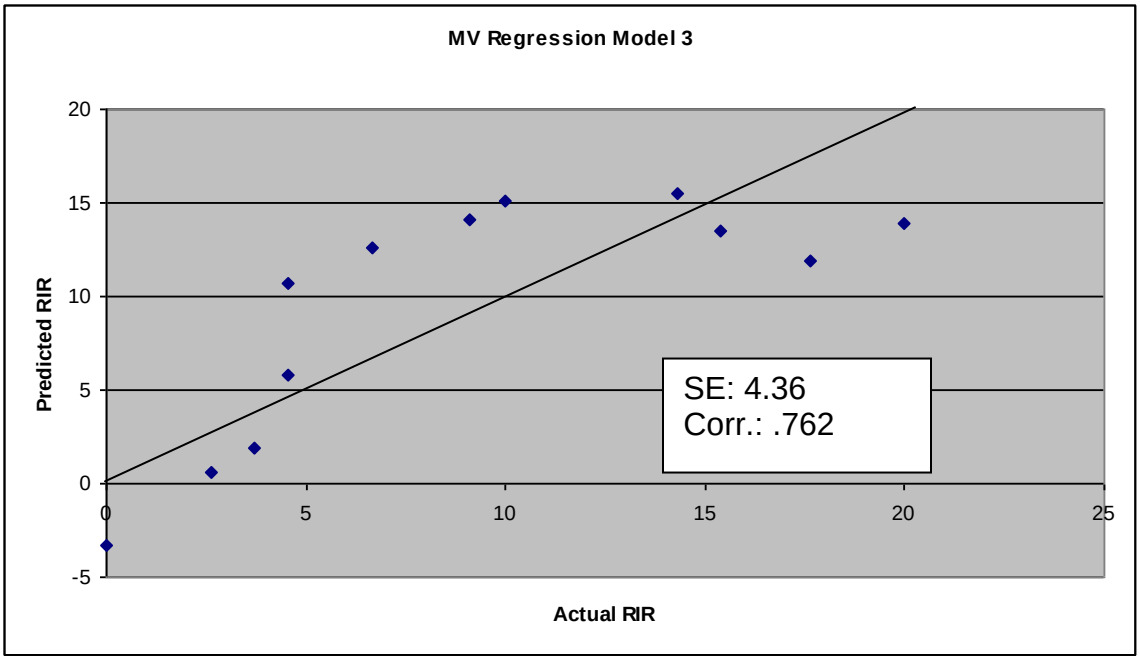

Figure 4-16. Actual RIR vs. multivariate linear regression 3 predicted RIR

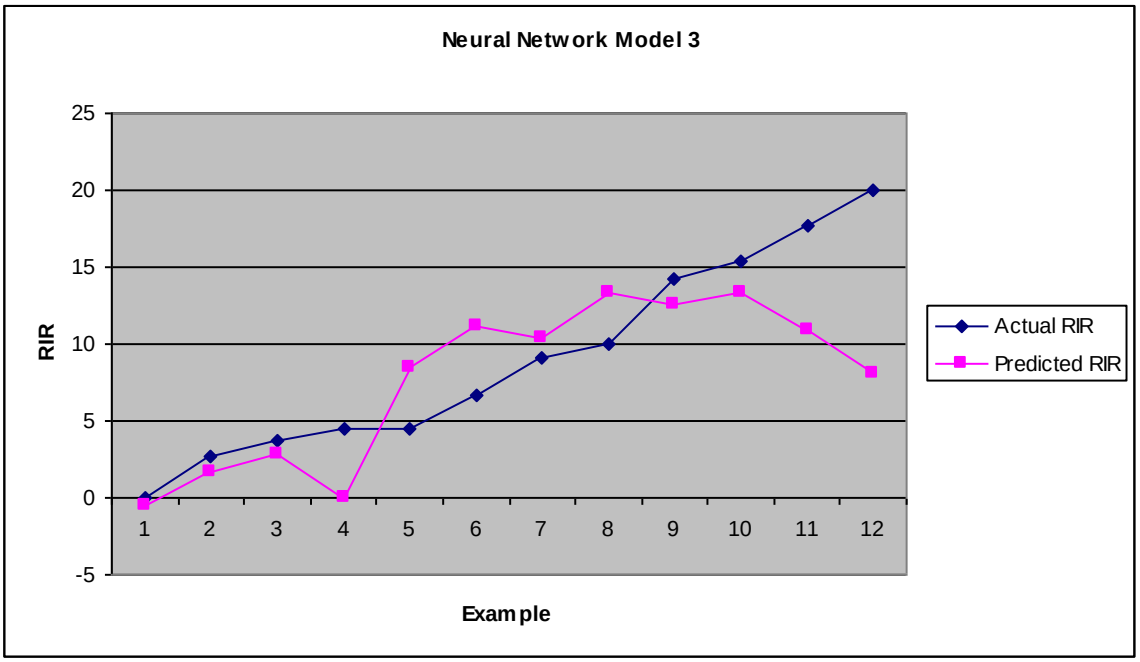

Figure 4-17. Actual RIR and ANN 3 predicted RIR for each example

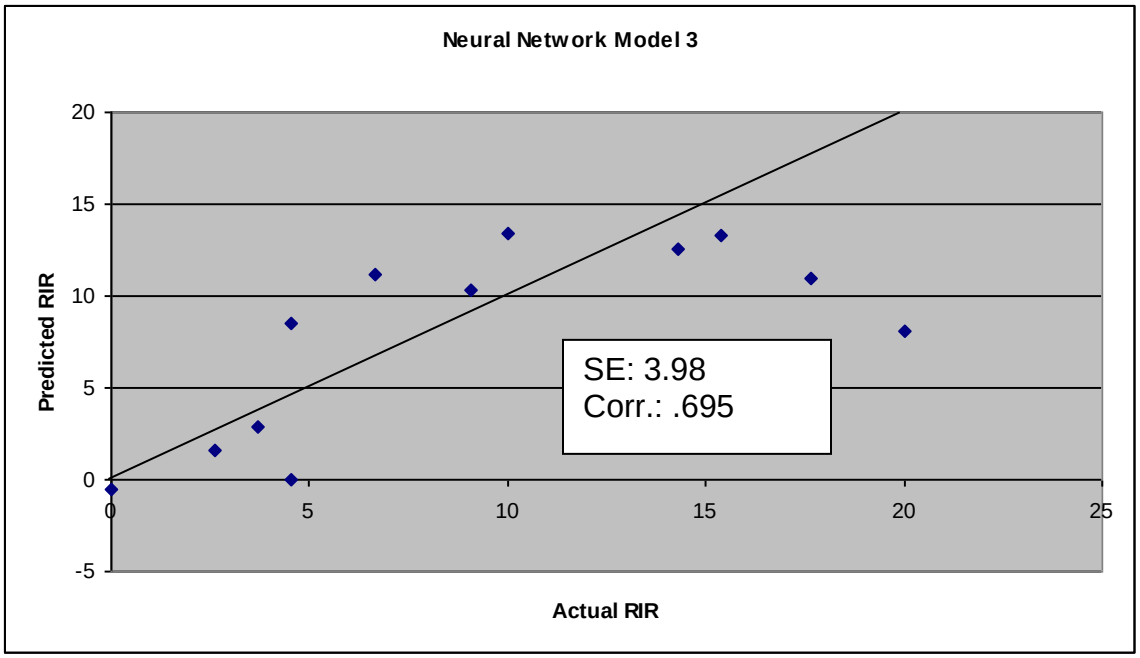

Figure 4-18. Actual RIR vs. ANN 3 predicted RIR

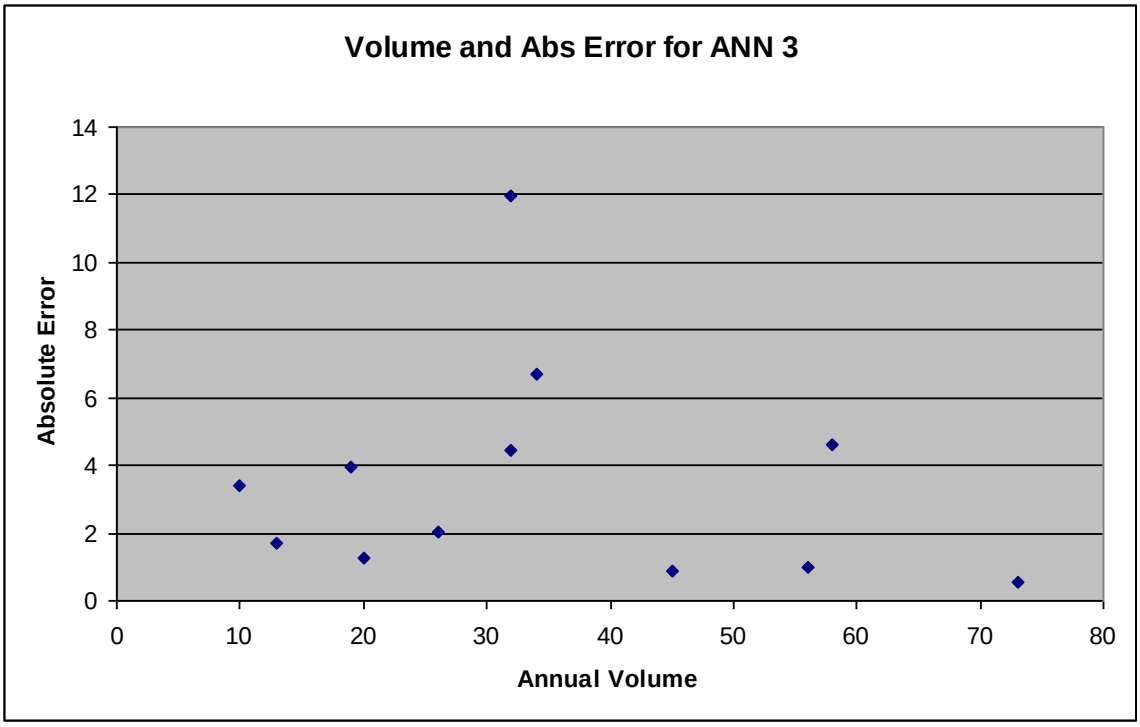

Figure 4-19. Absolute error as a function of annual volume (million $)

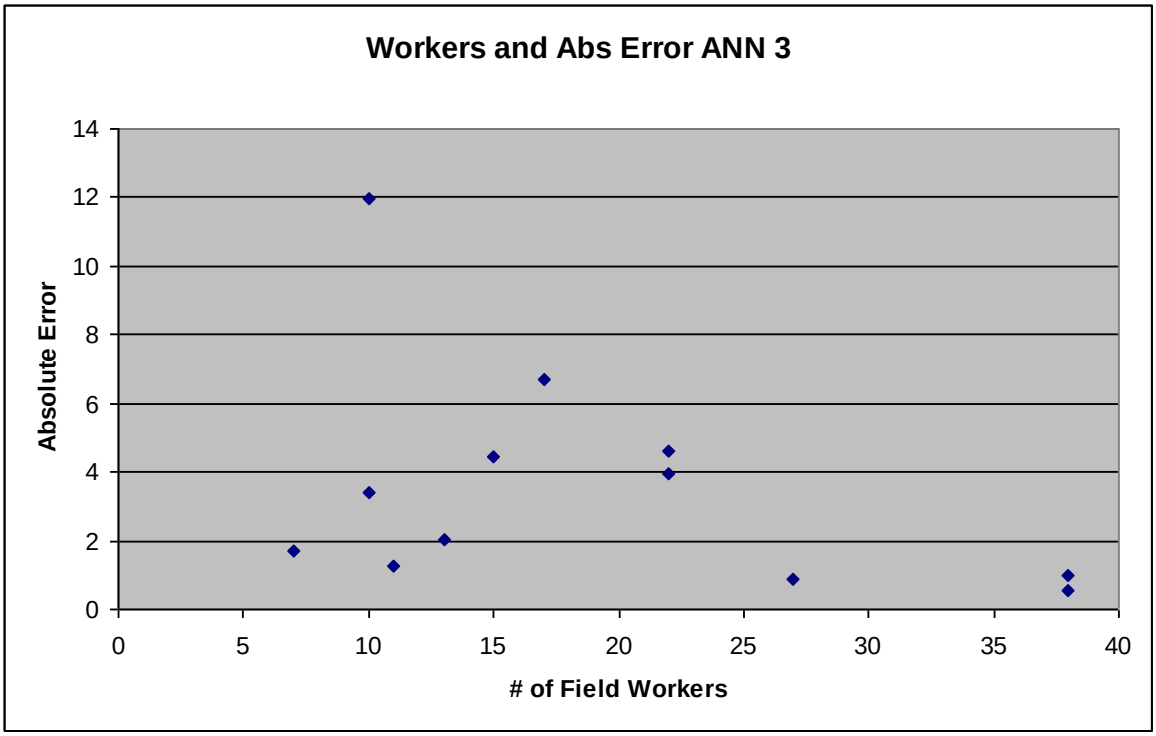

Figure 4-20. Absolute error as a function of full-time field workers employed

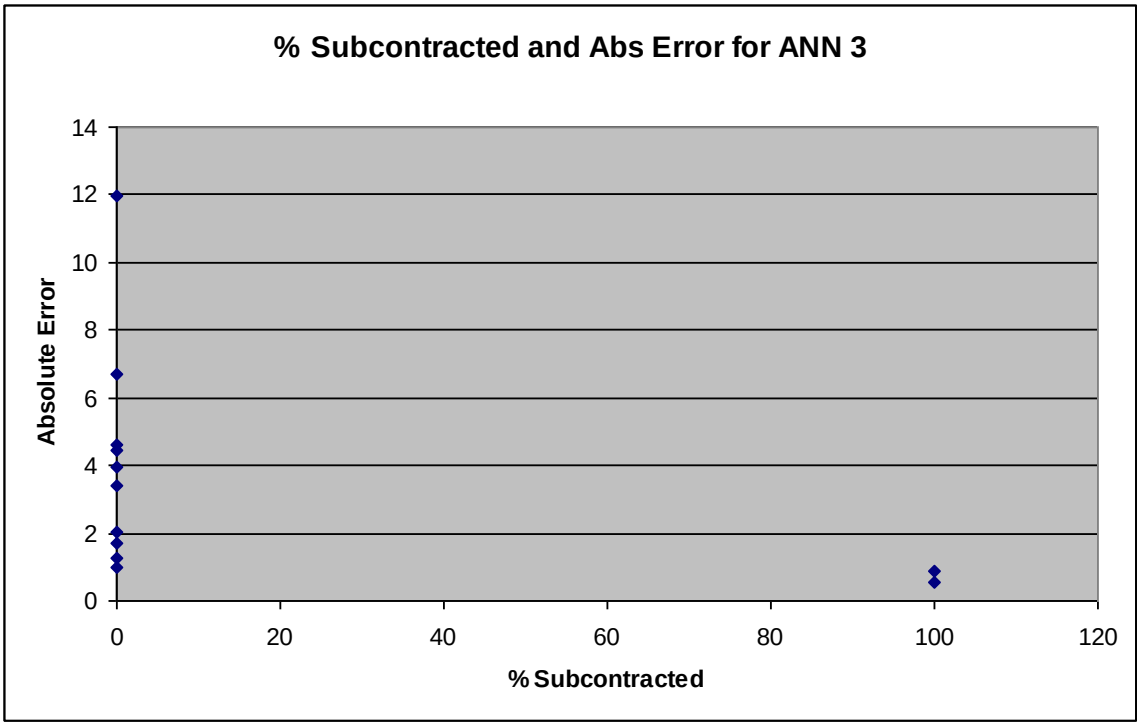

Figure 4-21. Absolute error as a function of percent of work subcontracted

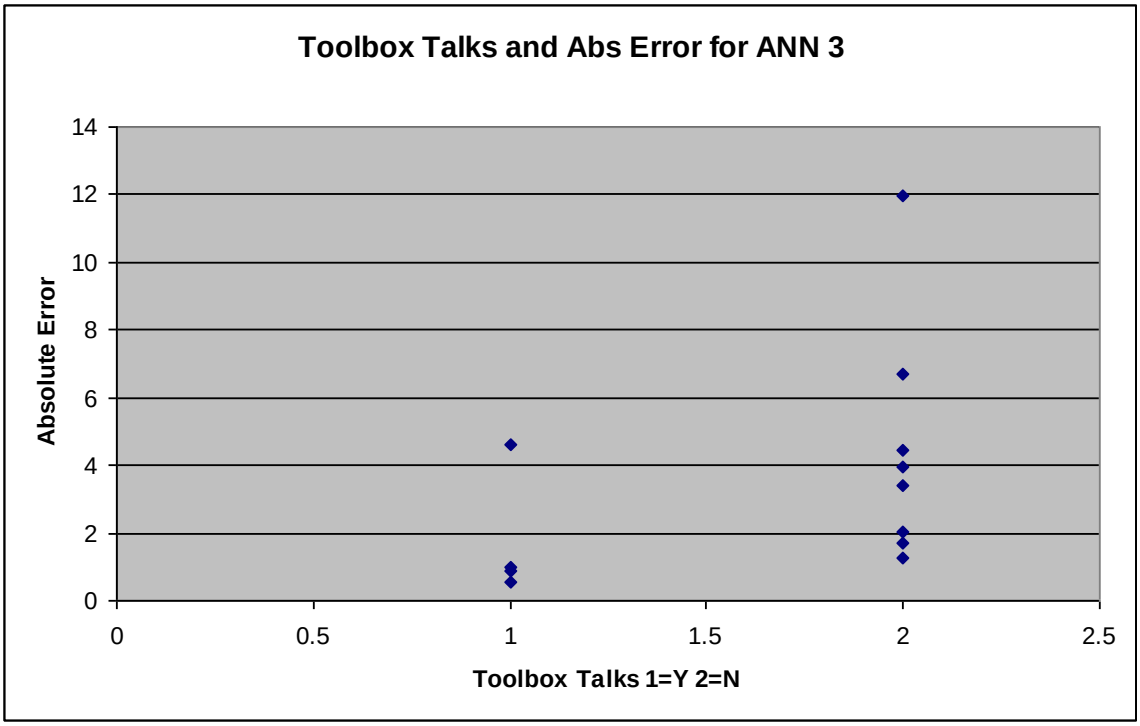

Figure 4-22. Absolute error as a function of holding toolbox talks

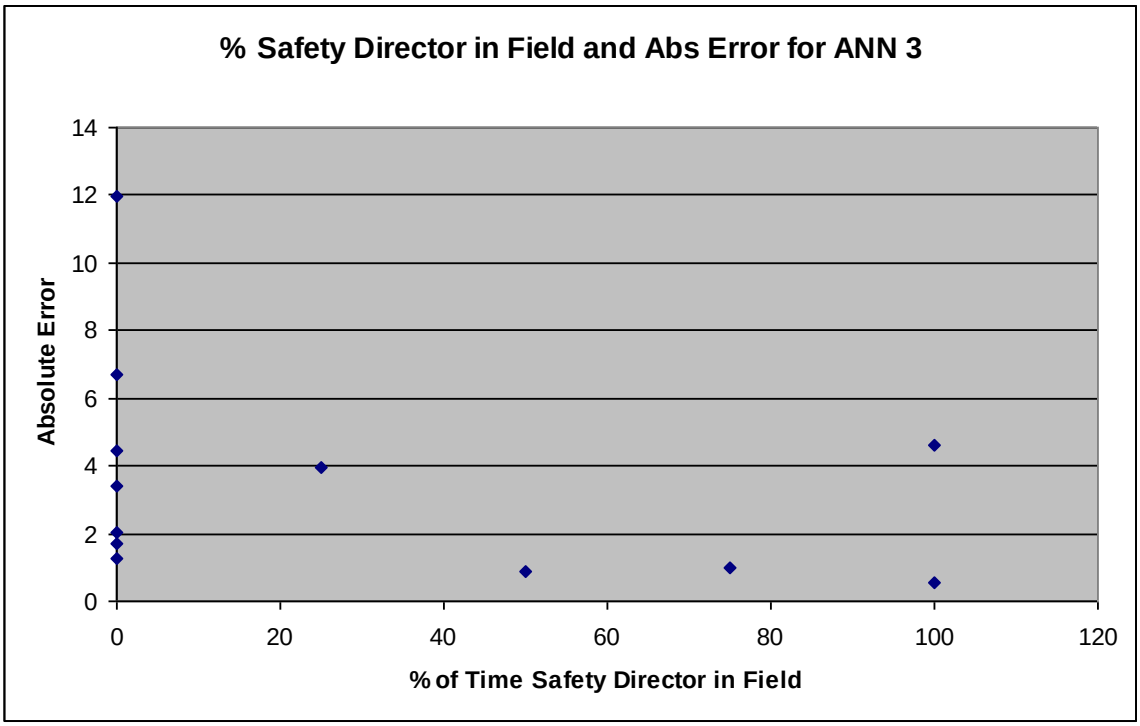

Figure 4-23. Absolute error as a function of percent of time a full-time safety director spends in the field

CHAPTER 5
CONCLUSIONS

Introduction

While considerable research has been conducted in the area of construction safety and the various factors that impact safety performance, few studies have focused on the specialty trade contractors which perform the majority of the actual work in constructing most buildings. Even less work has been done in the area of predictive safety modeling. Specialty contractors suffer from higher injury rates compared to the construction industry on average. When this is combined with the fact that they are often responsible for performing most of the work on construction projects, it would appear prudent for researchers to seek to reduce injuries and fatalities in construction by focusing on this often-neglected sector of the construction industry.

As noted in the literature review, some speculate that the benefit of vigorously pursuing a safer workforce is not well understood by some specialty contractors. As such, this researcher outlined the costs of jobsite injuries in construction. The costs of injuries are not just those covered by workers' compensation insurance. Indirect costs resulting from lost productivity of the injured worker, the crew, and the supervisors occur as well. Liability claims resulting from injuries can dramatically raise the costs of injuries even further. Additionally, more injuries will result in higher premiums paid to insure all employees because of rising EMRs. Thus, the costs of injuries to workers ultimately impact the entire company.

This research also examined various factors that impact safety performance. This was done to understand the impact of the potential inputs of a predictive RIR model. Drug testing, the nature of the firm, worker orientation training, and the use of

safety professionals were explored. Research in those areas showed that they impact the RIR of companies. These were selected as some of the inputs for the model. Other inputs were identified through a sensitivity analysis of the data from the survey.

Predictive Model

Three sets of inputs and three models were created for developing an effective predictive model of RIR. The three sets of inputs were chosen to try to mitigate any bias of the inputs for a particular model. Three models were used so that performances could be compared between them. The three models were a subjective "scorecard" developed by the researcher, a multivariate linear regression analysis, and an artificial neural network.

The artificial neural network ended up yielding the lowest single standard error. Initially its performance was weak. This may have been due to the amount of binary variables used as inputs. Inputs such as drug testing and providing orientation training could skew the model's predictions. While such inputs were associated with improved safety performance, some outlying poorly-performing companies had had such programs in place. As such, when the model made a prediction on these companies it underestimated their RIR. The ANN appeared much more suited to the continuous variables and understanding the relationships between them and the output values.

The multivariate linear regression model performed the best across all data sets on average. While it could not match the ANN's best performance on any single input set, it consistently performed relatively well regardless of the various inputs used. This technique is easier to use than the ANN for making predictions as it can be set up in minutes using Microsoft Excel by any novice user. The accessibility of this model to the

industry and the public should not be underestimated. Ease of use is an important factor for broad adoption as a safety policy evaluation tool in the industry.

The scorecard was developed to compare a subjective model to the objective ANN and multivariate linear regression models. It had the highest standard error, which was to be expected given its development. The mean correlation was good at .7176, but this does not fully describe the statistics. As inputs were reduced, the scorecard would increasingly rate companies as having a total score of zero. The model could not differentiate between companies with the same score. Twenty-two companies scored zero with the third input set. As such, this is not a good predictor. The more variables on which to base the scores, the better the performance of the scorecard was. This was the opposite trend of the artificial neural network which was perhaps confused by binary inputs.

Overall, the models were able to pick up the trend well, showing good correlation coefficients in all cases. Within this research, the ANN showed the most promise for use as a safety modeling tool, as it held the lowest error. However, little can be stated conclusively regarding the results of the analysis as the sample size was small.

CHAPTER 6
RECOMMENDATIONS

Introduction

As previously mentioned, specialty contractors have, for the most part, been under-represented in safety research with preference given to large general contractors. The significance of this study was limited by the small amount of data available that had responses for all desired inputs and the output. However, this research can be a starting point for future efforts into modeling safety performance.

Recommendations for the Industry

While the predictive model developed was perhaps not accurate enough to be a decision-making tool for specialty contractors, trends in safety performance can still be seen through this research. Indeed, all models had fairly good correlation coefficients, indicating the trend was well understood by the models. As such, the industry can view this research to see the impacts of various safety policies and characteristics of firms on RIR. The models developed illustrated that as size of the firm as measured by annual volume or number of full-time field workers increased, RIR decreased. Providing orientation for workers, holding weekly toolbox talks, and hiring a full-time safety director that spends most of his time in the field were other factors that the models showed had a positive impact on safety performance. The impact of the percent of work subcontracted and drug testing programs was not well understood by the models. Generally, those firms that subcontracted out a substantial portion of the work performed exceptionally well in terms of RIR. However, the number of such companies was small relative to the sample size, and as such the relationship cannot be well established. Drug testing programs were not well understood by the model; however,

the literature review showed conclusively that companies which conduct drug testing have lower RIR and EMR than those that do not. Firms should implement those policies that the model associated with decreasing RIR. While the precise decrease in RIR may not be predicted, a decrease should occur, nonetheless.

Recommendations for Further Study

This researcher recommends that a similar modeling study be conducted with the use of a much larger sample size, upwards of 300. Hopefully with a large sample size the outliers would have less of a negative impact on the performance of the model. In general, the models predicted the trends fairly well given their small training data set. A much larger sample size would allow for greater training and thus a more accurate prediction of RIR from the model. It would also allow for the ideal three training, testing, and validation sets of data. The small sample size of this study meant that only training and testing sets could be used.

Artificial neural networks should be further studied for their usefulness in predicting safety performances. A future study utilizing ANN as a predictor should focus on collecting as much data as possible of continuous variables as opposed to binary variables. The binary variables appeared to confuse the ANN as a few poorly performing contractors might have a positive value for one or more of the binary answers and thus it skewed their prediction, heavily underestimating their true RIR.

The multivariate linear regression model should be used in combination with the ANN for comparative purposes to further understand if any accuracy is gained by using the more complex tool. If the relations are relatively simple, utilization of the ANN may be unnecessary.

Another suggestion that may enhance the prediction of safety performance would be to use a classification model. For example, an ANN could be set to classify company safety performance as “very good,” “good,” “average,” “bad,” or “very bad.” Instead of using a continuous output value, which at the higher extremes becomes difficult for a model to accurately predict, the model could simply classify them into groups. The goal of accurately predicting safety performance as measured by RIR seems rather ambitious, particularly with small contractors where there is a high variability of RIR values. Larger contractors have more stable RIR values as their workforces are larger. One injury in a very large workforce will have much less of an impact that one on a workforce of seven workers. Thus, the future of safety modeling of smaller and medium-sized specialty contractors may be in the classification of performance into categories.

If several years were available, an additional study could be conducted to track safety policy/practice changes and how they impact safety performance. Information gathered over the course of years could allow a better understanding of the impacts of policy changes. Research could determine exactly how various changes in safety policies/practices affect RIRs and EMRs. Impacts of safety policy/practice changes on the costs related to injuries could also be examined. With this information, a study might be able to provide estimated dollar values of how much a company could anticipate in cost savings as a result of reduced injuries resulting from implementing various safety enhancing measures.

LIST OF REFERENCES

Baradan, .S, and Usmen, M. A. (2006). "Comparative Injury and Fatality Risk Analysis of Building Trades." *Journal of Construction Engineering & Management*, 132(5), 533-539.

Bureau of Labor Statistics. (2009). “How to Compute your Firm’s Incidence Rate for Safety Management.” <http://www.bls.gov/iif/osheval.htm> (September 2009).

Bureau of Labor Statistics. (2008). “Table 1. Incidence rates of nonfatal occupational injuries and illnesses by industry and case types, 2007.” <http://www.bls.gov/iif/oshwc/osh/os/ostb1917.txt> (August 2009).

Chau, K.W. (2007). “Application of a PSO-based neural network in analysis of outcomes of construction claims.” *Automation in Construction*, 16(5), 642-646.

Chi, C. F., Chang, T.C., and Ting, Hsin-I. (2005). "Accident patterns and prevention measures for fatal occupational falls in the construction industry." *Applied Ergonomics*, 36(4), 391-400.

Dembe, A., Erickson, J., Delbos, R., and Banks, S. (2005). “The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States.” *Occupational and Environmental Medicine*, 62(9): 588–597.

Hinze, J. and Raboud, P. J. (1988). “Safety on Large Building Construction Projects.” *Journal of Construction Engineering & Management*, 114(2), 286-293.

Everett, John G., and Willard S. Thompson. (1995). "Experience Modification Rating for Workers' Compensation Insurance." *Journal of Construction Engineering & Management*, 121(1), 66-78

Feronti, T. K. (2006). “Construction safety practices of specialty contractors.” Master’s thesis, University of Florida, Gainesville, FL.

Flood, I. and Issa, R. R. (2009). “Empirical Modeling Methodologies for Construction”, accepted for publication in *Journal of Construction Engineering and Management*, 30.

Gerber, J. K. and Yacoubian, G. S. Jr. (2001). "Evaluation of Drug Testing in the Workplace: Study of the Construction Industry.” *Journal of Construction Engineering & Management*, 127(6), 438-444.

Heaton, J. (2007) “A Feed Forward Neural Network.” <http://www.heatonresearch.com/articles/5/page2.html> (September 2009).

Hinze, J. and Appelgate, L. L. (1991). “Costs of Construction Injuries.” *Journal of Construction Engineering & Management*, 117(3), 537-550.

Hinze, J., Bren, D. C., and Piepho, N. (1995). "Experience Modification Rating As Measure of Safety Performance." *Journal of Construction Engineering & Management,* 121(4), 455-458.

Hinze, J. and Gambatese, J. (2003). "Factors That Influence Safety Performance of Specialty Contractors." *Journal of Construction Engineering & Management*, 129(2), 159-164.

Kinn S., Khuder, S. A., and Bisesi, M. S., et al. (2000). “Evaluation of safety orientation and training programs for reducing injuries in the plumbing and pipefitting industry.” *Journal of Occupational and Environmental Medicine*, 42(12), 1142–1147.

McVittie, D., Banikin, H., and Brocklebank, W. (1998). “The effects of firm size on injury frequency in construction.” *Safety Science*, 27(1), 19-23.

Mohamed, Sherif. (2002). "Safety Climate in Construction Site Environments." *Journal of Construction Engineering and Management*, 128(5), 375-384.

Occupational Health and Safety Administration. (2004). “Forms for Recording Work-Related Injuries and Illnesses.” <http://www.osha.gov/recordkeeping/new-osha300form1-1-04.pdf> (August 2009).

Song, L., and AbouRizk, S. M. (2008). "Measuring and Modeling Labor Productivity Using Historical Data." *Journal of Construction Engineering & Management*, 134(10), 786-794.

Substance Abuse and Mental Health Services Administration. (2008). “Worker drug use and workplace policies and programs: Results from the 1994 and 1997 National Household Survey on Drug Abuse.” <http://www.oas.samhsa.gov/nhsda/A-11/toc.htm> (September 2009)

Substance Abuse and Mental Health Services Administration. (2009). “Illicit Drug Use Among Construction Workers.” <http://www.oas.samhsa.gov/construction.htm> (September 2009).

Waehrer, G. M., Dong, X. S., and Miller, T., et al. (2007). "Costs of occupational injuries in construction in the United States." *Accident Analysis & Prevention*, 39(6), 1258-1266.

Wilmot, C. G., and Mei, B. (2005). "Neural Network Modeling of Highway Construction Costs." *Journal of Construction Engineering & Management*, 131(7), 765-771.

Zayed, T. M. and Halpin, D. W. (2005). "Pile Construction Productivity Assessment." *Journal of Construction Engineering & Management*, 131(6), 705-714.

BIOGRAPHICAL SKETCH

The author is from Neptune Beach, Florida. After completing a Bachelor of Science degree in Building Construction, the author decided to continue his education in construction at the graduate level. In December of 2009 the author was awarded a Master of Science degree in Building Construction.