ARTIFICIAL NEURAL NETWORKING AS A DECISION TOOL FOR
NATURAL GAS INVESTMENT

A Thesis
presented to
the Faculty of California Polytechnic State University,
San Luis Obispo

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Industrial Engineering

by
Micah Denecour
January 2011
COMMITTEE MEMBERSHIP

TITLE: Artificial Neural Networking as a Decision Tool for Natural Gas Investment

AUTHOR: Micah Denecour

DATE SUBMITTED: January 2011

COMMITTEE CHAIR: Dr. Lizabeth Schlemer, Ph.D.

COMMITTEE MEMBER: Dr. Roya Javadpour, Ph. D.

COMMITTEE MEMBER: Dr. Kurt Colvin, Ph. D
ABSTRACT

Artificial Neural Networking as a Decision Tool for Natural Gas Investment

by

Micah Denecour

With the growing interest in the Marcellus Shale and its natural gas deposits, there are opportunities to purchase and hold land for investment purposes. A robust decision tool is needed to help guide investors towards the most profitable properties. Artificial neural networks have many unique benefits that make them an ideal candidate for this purpose.

The artificial neural networks created in this study had nine independent variables. Combinations of these nine variables were created to describe 300 theoretical properties available for purchase. Each of these properties were then evaluated by an expert in the field and given a score from one to five to rate its investment potential, which was the dependent variable.

Sixteen different network architectures were used to create over 200 neural networks. However, none of these networks met the criteria established to determine success. This is likely due to the unreliability in the data used to train the network, evidenced by the expert’s inability to reproduce previously assigned scores.

Keywords: Artificial neural networking, natural gas, investment analysis, Marcellus shale, decision model.
TABLE OF CONTENTS

LIST OF TABLES .............................................................................................................. viii

LIST OF FIGURES ............................................................................................................ ix

Chapter 1: Introduction ........................................................................................................ 1
  Objective ....................................................................................................................... 1
  Introduction .................................................................................................................. 1
  Research Question ...................................................................................................... 6
  Hypothesis ................................................................................................................... 6

Chapter 2: Literature Review ............................................................................................... 8
  Introduction to Artificial Neural Networks .................................................................. 8
  Benefits of Artificial Neural Networks ....................................................................... 9
    Artificial Neural Networks vs. Conventional Approaches ......................................... 10
    Data Assumptions ...................................................................................................... 12
    Development Effort ................................................................................................... 13
    Capturing Extreme Values ....................................................................................... 14
  Limitations of Neural Networks .................................................................................. 15
  Data Collection ........................................................................................................... 17
    Randomized Data Selection ..................................................................................... 17
    Contingent Rating .................................................................................................... 18
  Building an Artificial Neural Network ....................................................................... 19
    Number of Hidden Units .......................................................................................... 20
    Data Set Size ............................................................................................................ 21
    Training ..................................................................................................................... 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous versus Integer Values</td>
<td>53</td>
</tr>
<tr>
<td>Network Architecture</td>
<td>54</td>
</tr>
<tr>
<td>Random Weight Initialization</td>
<td>57</td>
</tr>
<tr>
<td>Chapter 6: Conclusion</td>
<td>58</td>
</tr>
<tr>
<td>Chapter 7: Recommendations</td>
<td>60</td>
</tr>
<tr>
<td>Use of Historical Lease Data</td>
<td>60</td>
</tr>
<tr>
<td>Reducing Subjectivity</td>
<td>61</td>
</tr>
<tr>
<td>Comparison to Other Shale Plays</td>
<td>62</td>
</tr>
<tr>
<td>Use of a Broader Rating Scale</td>
<td>63</td>
</tr>
<tr>
<td>Multiple Experts</td>
<td>63</td>
</tr>
<tr>
<td>Bibliography</td>
<td>65</td>
</tr>
<tr>
<td>Appendix</td>
<td>70</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Input variable summary ..........................................................30
Table 2: NYMEX Natural Gas Forecast (ajmpc.com)..............................34
Table 3: Expert scoring rating scale .........................................................70
LIST OF FIGURES

Figure 1: Marcellus Isopach (thickness of shale) (Andrews 2009) ..........2
Figure 2: Horizontal well with hydraulic fracturing (Geology.com) .............4
Figure 3: Example cross plot generated by Matlab ................................43
Figure 4: Example training performance plot ......................................44
Figure 5: Summary of Network Results .............................................48
Figure 6: Comparison of Neural Network, Regression, and Expert Re-
score ..........................................................................................50
Figure 7: Statistical descriptors for Cost .............................................70
Figure 8: Statistical descriptors for Other Annual Revenue ......................71
Figure 9: Statistical descriptors for Distance to Pipeline/Wells ...............71
Figure 10: Statistical descriptors for Gas Price ....................................72
Figure 11: Statistical descriptors for Geologist Score ............................72
Figure 12: Statistical descriptors for Landsmen Score ...........................73
Figure 13: Statistical descriptors for Existing Lease Expires ....................73
Figure 14: Statistical descriptors for Royalty .......................................74
Figure 15: Statistical descriptors for Lease Terms Score ........................74
Figure 17: Equation used to calculate Royalty % .................................75
Figure 18: Scatterplot of Royalty vs. Existing Lease Expires .................75
Figure 16: Equation used to calculate Landsmen Score ........................75
Figure 19: Scatterplot of Landsmen Score vs. Existing Lease Expires ....76
Figure 20: Scatterplot of Landsmen Score vs. Gas Price

Figure 21: Scatterplot of Landsmen Score vs. Distance to Pipeline and Wells

Figure 22: Initial Matlab program code

Figure 23: Probability plot of percent error for Network 5

Figure 24: Histogram of percent error for Network 5

Figure 25: Probability plot of percent error for Network 16

Figure 26: Histogram of percent error for Network 16

Figure 27: Regression analysis
Chapter 1: Introduction

Objective

This research effort investigated the efficacy of artificial neural networks as a decision model to mimic expert evaluations of natural gas investment properties. An expert was given theoretical properties to rate, with these ratings being used to train a neural network. The predictions made by the neural network were then compared to those of the expert to determine whether or not the neural network is a suitable decision aid to investors.

Introduction

Until recent technological breakthroughs, many oil and gas companies believed natural gas trapped in deep shale to be unreachable in an economic manner. Although advanced techniques such as hydraulic fracturing and directional drilling had long been used in conventional oil and gas fields, they had not been applied to shale formations. As more companies enter this relatively new and undeveloped shale gas play, they seek to optimize their spending to maximize return on investment.
Consequently, it is necessary to develop suites of tools to track, plan, and implement strategic land purchases for natural gas production.

The Marcellus shale in particular has generated considerable interest in the last few years due to its large reserves of natural gas and emerging state. Located primarily in Pennsylvania, Virginia, and New York, the Marcellus shale has the added benefit of being close to the large markets in the northeastern United States as shown in Figure 1.

![Figure 1: Marcellus Isopach (thickness of shale) (Andrews 2009)](image)

It is estimated that there is about 363 trillion cubic feet (tcf) of recoverable gas in the Marcellus shale, which would be enough to supply the entire United States for the next 15 years at the current rate of consumption (Soeder and Kappel 2009). Also, compared to other shale
deposits across the United State, the Marcellus is relatively undeveloped with many speculators hoping to “get in early.” For example, in Bradford County, PA, there are currently 100 permits on file for horizontal natural gas wells. However, 10,000 wells are estimated to be in production in the county in the next five years. These reasons among others make the Marcellus shale land an acquisition target for oil and gas companies and individual investors alike.

Natural gas shales are a fine-grained organic rock that holds gas within small pore spaces. The shale formations, often up to 300 feet thick, contain natural gas in these small pore spaces distributed across a large area. Whereas more traditional gas sources flow easily, shales are relatively impermeable to gas flow unless fractures exist in the shale. To capitalize on this, artificial fractures are introduced using a process known as hydraulic fracturing. In hydraulic fracturing or “fracing,” explosive charges are set of in an underground well to create fissures or fractures within the shale. Then, a combination of water, sand, and other additives are pumped into the well to keep the fractures open, known as “propping.” This allows the natural gas to flow into the well and up to the surface for collection.

Because the natural gas is distributed across a large area, it is necessary to increase the collection area of a well. To do this, a process known as directional drilling is used to drill both vertically and horizontally. First, a vertical well is drilled into the shale deposit as shown in Figure 2,
often 6,000-8,000 feet down. Then, up to six horizontal wells are drilled off of the vertical well and go laterally up to one half mile. After encasing the walls of the well in cement or steel, the well is ready for fracking as described previously. By drilling several horizontal wells in such a way, it is possible for a single vertical well to collect gas from a large area of approximately 640 acres or one square mile.

![Figure 2: Horizontal well with hydraulic fracturing (Geology.com)](image)

As it is uncommon for an oil and gas company to own a perfectly square 640 acre parcel of land, a process called unitization controls the way that gas wells are drilled and the profits are divided. If a gas company does not own the land but is interested in drilling, it must lease the land from the landowner. The landowner typically receives an upfront cash
bonus for each acre leased in additional to a royalty payment based on a percentage of the gas extracted. The gas company will pool a collection of adjacent leases it holds into a production unit, with a typical production unit being between 600 and 1000 acres in size. The landowners receive payments based on the amount of land that they own in the production unit and the royalty percentage dictated in the leasing agreement.

Currently, a landowner in the core area of the Pennsylvania Marcellus shale can expect a lease bonus of $4,000 to $6,000 for leasing the land with a 12% to 22% royalty payment. Consequently, it can be very profitable to own property that is located within a production unit. One such investment fund wishes to capitalize on the developing Marcellus shale play by purchasing land or mineral rights and then leasing that land to drilling companies. While prime pieces of land sell for $6,000 to $8,000 per acre, if the land is then leased to a drilling company substantial profit can be gained in the form of a lease bonus and royalty payments. The risk associated with this strategy is that land purchased by the investment fund may not be sought after by a drilling company.

Consequently, it is important for the investment fund managers to ensure that all land purchases meet certain strategic and tactical requirements. In addition to landmen and geologists on the ground, an objective analysis tool based on scientific data would aid decision makers. More traditional heuristic methods are not sufficient for this application, as the relationship between variables is unknown and likely to be non-linear.
Artificial neural networking is a unique non-parametric statistical modeling tool that could meet this needs and will be investigated in this study.

Research Question

Can artificial neural networks be used as a decision model to mimic expert evaluations of natural gas investment properties?

Hypothesis

When compared to an expert’s evaluation of a set of properties for natural gas investment, a properly trained neural network will predict a score for that same set of properties with an acceptable level of error. For the purpose of this study, the mean absolute percent error and the absolute percent error will be used to determine the success of the neural network. The mean absolute percent error (MAPE) measures the total error for an entire data set and should be less than five percent (Nguyen and Cripps 2001). The absolute percent error (APE) measures how the error deviates between properties and should be no greater than ten percent for any property (Mann and Ayala 2009). The MAPE and APE are calculated as shown below:
\[ MAPE = \left( \frac{\sum_{i=1}^{n} \left( \frac{P_i - A_i}{A_i} \right) \times 100}{n} \right) / n \]

\[ APE = \left| \frac{(P_i - A_i)}{A_i} \times 100 \right| \]

Where \( P_i \) and \( A_i \) are the network’s prediction and the actual score given by the expert for property \( i \) in the data set of size \( n \). Additionally, the percent error should be normally distributed about zero and exhibit minimal variation when plotted on a histogram (Mann and Ayala 2009). If the neural network created in this study achieves these metrics, it will be said to sufficiently approximate expert evaluations and will be a suitable decision tool.
Chapter 2: Literature Review

Introduction to Artificial Neural Networks

Artificial neural networks are information processing models that attempt to mimic the learning functionality of the human brain. When the body senses an input, the nervous system sends electrical signals to the brain describing the input. The brain receives that input and interprets it through a series of inter-connected and parallel neurons that transform and process the information. After processing the information, the brain reacts by creating what it deems to be an appropriate response. By repeatedly being exposed to the same stimuli, the brain receives feedback and learns the optimal processing and response.

Much like the human brain, artificial neural networks (ANN) consist of inter-connected neurons passing signals between each other. Most networks consist of an input layer, a hidden layer, and an output layer. In a feed-forward network, the input layer consists of data for the known, or independent variables. This data is passed through the hidden layer where it is processed and sent to the output layer, which represents the dependent variable. By iteratively mapping inputs to outputs and comparing the results to historical data, an ANN learns to process signals to achieve the desired result. Neurons are the basic computing units that
perform local data processing inside the network (Samarasignhe 2007). The function of the network is determined by its structure; how the neurons are connected to one another, the connection strength, and the processing performed by the neurons. The most common neural network is a feed-forward type in which the data flows from input layer to output layer (Fadlalla and Lin 2001). There are other possible configurations such as recurrent networks and radial-based networks, but they are not as common or practical.

After a network has been constructed and trained to correctly map inputs to outputs, it can be used to perform a variety of tasks including function approximation, clustering, forecasting, and prediction. The prediction capabilities of neural networks will be explored in this study. The power of neural networks is that they can acquire, store, and use experiential knowledge. With advances in computing power, ANNs have become adaptive, distributive, and massively parallel systems that have proven potential for solving problems (Mohaghegh et al 1996). They have been used across a variety of industries to solve complex problems with a high degree of accuracy.

Benefits of Artificial Neural Networks

Throughout the literature there are many examples showing the benefits of neural networks over other methods. By mimicking the power
of the human brain, they can solve very complex and non-linear problems that are either impossible or unfeasible to solve using conventional methods (Mohaghegh et al 2001). Neural networks have been used frequently to create expert systems that are capable of learning complex relationships. Such a system can help make informed decisions and reduce subjectivity when it is trained correctly (Worzala et al 1995). Neural networks have many benefits that make them an attractive option when compared to more traditional analysis techniques.

**Artificial Neural Networks vs. Conventional Approaches**

Prior to the advent of neural networks, complex analysis problems involving multiple variables were usually solved using regression or other hedonic models. In the real-estate industry, the standard approach to constructing pricing models was based on linear regression (Din et al 2001). Additionally, many managers utilized discriminant analysis or other quantitative techniques to make more accurate decisions (Aiken and Bsat 1999). However, many of the techniques are not reliable, easy to use, or easy to develop. For this reason, many studies have been done to compare neural networks directly to these conventional approaches.

In one such study, Odom and Sharda (1990) used neural networks to predict the risk of bankruptcy. They compared their results to those obtained through discriminant analysis, the traditional method for the field,
and found that neural networks predicted bankruptcy more accurately. In their survey of neural networks in the financial industry, Aiken and Bsat (1999) found that neural networks were at least as accurate as competing techniques and easier to develop for a variety of problems.

In the real-estate valuation industry, similar results were found. Din et al (2001) constructed a model to predict the value of individual residential properties using both linear regression and neural networking. They found the ANN to have more potential for realistic pricing of individual properties, even with a small sample size. To further study the effects of different network structures on such a comparison, Nguyen and Cripps (2001) performed 108 direct comparisons using different functional model specifications, sample data, and evaluation criteria. The conclusion reached was that ANN performs better than regression when a moderate to large sample size is used (over 500 samples). Two separate studies performed by Do and Grudnitski (1992) and Tay and Ho (1992) found neural networking to be almost twice as accurate as regression for predicting real-estate values.

However, there is some literature that suggests neural networks are not more effective than traditional methods. Worzala et al (1995) found that neural networks only slightly outperformed regression, if at all, for real-estate valuations. One possible reason for this outcome is that they used a relatively small sample size, which later studies showed to be an important factor in the success of neural networks. Additionally, when the
study was performed, the number of neurons in the hidden layers was restricted due to computational power limitations. Recent literature, as discussed previously, suggests that neural networks can perform better than conventional methods if the right parameters are used.

**Data Assumptions**

Another strength of ANNs is that they do not require assumptions to be made about the form or function of the data to be forecasted. For conventional methods based on mathematical calculations, the experimenter must assume a linear or modeled non-linear relationship (Mohaghegh et al 1996). This is problematic, as the form of the function is often unknown or is non-parametric. For example, Nguyen and Cripps (2001) noted that in real-estate applications, age and square footage is not linear with respect to the housing value. In this case, neural networks may perform better because they do not need predetermined functional form based on determinants. While some authors argue that non-linear parameters can be modeled using semi-log, log-log, cubic, or quartic functions, it is still necessary to assume that one of these functions approximates the parameter correctly. As Din et al (2001) found, neural networks are non-parametric and non-linear statistical modeling tools that are directly applicable to many fields for this reason.
In addition, tools such as multilinear regression require that the data have a certain level of homoscedasticity or variance, multicollinearity and independence of variables. Other forecasting techniques require other assumptions for the data (Aiken and Bsat 1999). There can also be issues with outliers significantly skewing results, which is not the case with neural networks. When using neural networks, no tests need to be conducted to check for any of these assumptions that are required for other statistical analysis techniques.

Due to their adaptive nature, neural networks have the added benefit of being able to work with partial or incomplete data sets. In other modeling techniques, missing data is a serious problem. In their study of real-estate valuation, Worzala et al (1995), highlighted the challenges of appraising property with unreliable an unverifiable data and suggested neural networks as a solution to this problem.

**Development Effort**

When compared to other techniques used for statistical modeling, neural networks have the added benefit of being relatively fast to build and execute. As discussed previously, they do not require as much manipulation or testing of the data set, which is a significant time savings. Mohaghegh et al (1999) used neural networks to create an expert system capable of designing natural gas well fracturing jobs. This expert system
served as a replacement for the previous method of using an engineer and computer simulation software. They found that the neural network replicated the functionality of the previous system and reduced the amount of time to complete the design from one day to almost nothing.

This conclusion was reiterated by Mann and Ayala (2009) who pointed out that traditional analysis of the complex parameters involved in natural gas storage facility design required the use of rigorous numerical simulation. This process took a considerable amount of computational time. Neural networks provided “fast, reliable, and robust” (Mann and Ayala 2009) predictions of optimal operating conditions. They were able to create a system that negated the need for simulation and reduced computational time. With the increasing power of computers, neural networks become very feasible and quick solutions to engineering design and decision problems.

**Capturing Extreme Values**

Using traditional analysis techniques, outliers are often accounted for but result in a skewed model. According to Din et al (2001), this inability to capture more extreme conditions is due to the global nature of linear models. They believe that neural networks are more promising in this respect because they can spatially disentangle the nuances of a complex input parameter space. This strength also results in the
consideration of parameter configurations that might not have even been considered using typical analysis techniques (Mann and Ayala 2009). The ability to accurately account for a wider range of values increases the prediction capacity of neural networks when compared to other methods.

**Limitations of Neural Networks**

While there are many benefits to neural networks that make them an effective statistical modeling tool, there are also some key limitations to consider. Essentially, neural networks can be very sensitive to the model setting and parameters that determine its architecture. Allen and Zumwalt (1994) concluded that their model’s success for stock predictions was dependent on the data set, hidden neurons, and other settings. They recommend caution during the development and use of neural networks for financial models.

The first such limitation is that if the training set is too small, the network will memorize the sample as opposed to finding the underlying patterns. As a result, extreme data points will have a disproportionately large influence on the model. Goutte (1997) suggests using a technique known as the k-fold cross validation training method to correct this.

Conversely, if too many training samples and hidden neurons are used, the model is less likely to generalize, or predict new data. This is due to model memorizing the training set to the point that it is over
training. If too few hidden units are used, the training error will be high and the network will not generalize as accurately as possible because of underfitting and statistical bias (Nguyen and Cripps 2001). Throughout the literature, no conclusive formula was found to determine the correct sample size and number of hidden units. Most authors suggest a trial and error procedure to generate the best possible network architecture.

Another concern found regarding neural networks was the estimation error inherent in any statistical model. Lenk et al (1997) expressed their concern in a paper titled “High-tech valuation: should artificial neural networks bypass the human valuer?” They argue that mass valuation techniques like neural networks save information processing resources, but that savings might not outweigh the cost of estimation error. They acknowledged that this error is present equally in neural networks and regression.

Mohaghegh et al (1999) indirectly addressed this concern by noting that their neural network application is entirely data driven, but that the addition of expert knowledge for real-time decision-making may enhance the process. It is important to note that a neural network alone should not be responsible for decision-making, but should be an objective tool to aid those making the decisions.
Data Collection

To train a neural network that is capable of accurate predictions, a satisfactory data set must be collected and used. And, as Mohaghegh et al (2001) found, the more data that is available the better a neural network will perform. But, evaluating all possible combinations of variables is not feasible, especially when using a ranking evaluation model. So, it is important to consider the methods by which the data is gathered and evaluated before it is exposed to a neural network.

Randomized Data Selection

As the number of variables in a problem increases, so does the number of possible combinations that must be evaluated to determine the optimal solution. At a certain point, it becomes infeasible to perform an exhaustive search and a subset of the available data must be used. It is important that the data set has minimal bias and is substantial enough to train the neural network. In the literature, one method of doing this is to randomly select a certain number of data sets from the possible permutations of input variables (Mann and Ayala 2009). This method was used successfully by Nguyen and Cripps (2001) for the evaluation of real estate. Having data selected randomly from amongst the possible
combinations is important if the data set is to then be rated or scored by experts (Harrell 1993).

Contingent Rating

Since there is no publicly available historical data on which to base this study, it is necessary to have an expert artificially assess the value of theoretical natural gas properties. The idea of providing hypothetical choices to experts to rate or score is known as a stated preference contingent rating model. Essentially, an expert is given a list of options and asked to give each one a score on a continuous scale (i.e. from 1 to 10). Such methods are generally considered easier to design and analyze than other ways of rating and ranking.

Washington et al (2006) used this type of model for traffic experts to rate the anticipated effectiveness of accident prevention techniques. They relied on the law of large numbers by using many experts and many data points to diminish the problems associated with interpreting expert’s statements. Demange (2010) identified a pitfall of stated preference methods, known as intensity invariance. One expert may inadvertently inflate their statements by consistently over or under rating. Dividing every expert’s rankings by their average rank can diminish this effect and scale the responses appropriately.
Harrell (1993) made several suggestions on how to best design a stated preference exercise. One suggestion from his work was to provide the subject with a list of all possible alternatives at once. This allows the subject to consider the size of the difference between the options. Also, it was noted that subjects should only be asked about what is familiar to them. Harrell found that eight attributes or variables were generally the maximum subjects should be allowed to consider simultaneously. He noted that as the number of attributes and levels increases, the number of replications needed increases more rapidly.

**Building an Artificial Neural Network**

While there is no right way to build an ANN, many suggestions were found in the literature that served the respective authors well. According to Aiken and Bsat (1999), building a neural network is a three-stage process. First, decisions must be made about what input variables will be studied. In many cases, this is done by consulting with experts to see what is significant based on their experience. Statistical methods can also be used to weed out less influential variables. The architecture or model parameters such as the number of layers, type of transfer functions, and training method must also be selected at this point.

Next, the network must be trained on a data set until there is minimal error between model’s predictions and the actual values. Once
this is complete, the model must be exposed to a validation data set that was not used during training. This is to test the model’s generalization abilities. If the model does not perform correctly, the variables or architecture are changed and the process starts over. It is common throughout the literature that several iterations are necessary before a sufficient network is created.

**Number of Hidden Units**

In terms of network architecture, the most basic parameters are the number of hidden layers to be used and the number of neurons that will be in each layer. Geman et al. (1992) pointed out that a small network with only one hidden neuron in one hidden layer is likely to be biased. Conversely, if the network is too large, the bias will be reduced but there is a risk of significant variance contribution to the error. Wilamowski (2003) clarified that there is no set method for determining the number of hidden units to use, but that as the size grows, the ability of the network to generalize decreases. A network with a larger number of neurons has the possibility of mapping noise supplied to the network from the inputs. Nguyen and Cripps (2001) recommend trial and error; if too few hidden units are used, the training and generalization error will be high, if too many hidden units are used, the training error will be low but the generalization error will be high.
In a summary of forty finance uses of neural networks, Fadlalla and Lin (2001) found that backpropogation training with one hidden layer was widely used. Twenty-nine of the studies used one hidden layer, seven used two hidden layers, two used no hidden layers, and five did not report the number of layers used. Mann and Ayala (2009) used two hidden layers with 50 and 25 neurons in each layer. To characterize gas reservoirs, Mohaghegh et al (1996) used one hidden layer with 28 hidden neurons. While the number of hidden units varies widely, it is generally accepted that trial and error experimentation will help researchers determine the correct configuration.

Data Set Size

There is no consensus on the required number of data points to correctly train a neural network, but many researchers share what has worked well for them. Mohaghegh et al (1999 and 2000) used close to 600 different data points to train their models, with 40% of that set being reserved for verification and testing of the network. Similarly, Din et al (2001) divided their 285 sample data set with 60% used to training and 40% for testing and validation. Mann and Ayala (2009) found that 500 training sets, with 100 of those reserved for testing, was enough to train their network. Wu et al (2009) had difficulty finding data and used only 100 training samples, with 10 being used for testing. Their results showed the
network was able to successfully aid decision makers in the development of heavy oil reservoirs.

Training

The process of training a neural network involves iteratively exposing the network to a set of inputs for which the output is known. The network’s predicted output is then compared to the actual output and the weights connecting the neurons are altered to minimize the error. Many different training algorithms are available for use with neural networks and vary in the way that they measure and adjust for the error found in predictions. The error surface of a neural network is a multi-dimensional function that represents the amount of prediction error present in the model for a given set of weights that connect the neurons. Different training algorithms traverse the error surface differently often resulting in completely unique networks. Consequently, it is very important to select the right training algorithm for a given situation.

One of the most popular and widely used training algorithms is backpropogation with gradient descent (Mann and Ayala 2009). This technique simultaneously and incrementally adjusts the weights between neurons to find a minimum on the error curve or surface. Initially, the weights in the neural networks are randomly set. The network processes an input and compares it to the output, with the initial difference between
the two being very high. The slope of the error surface is evaluated to
determine how sensitive the error is to changes in weights. This sensitivity
is used to incrementally guide the changes in the weights in the direction
of minimal error (Samarasinghe 2007). This method is called
backpropogation because the error derivative is calculated using the chain
rule through the network starting from the output layer, through the hidden
layers, and back to the input layers. Essentially, this training method works
backwards through the neural network to adjust the weights between
neurons to minimize the prediction error.

In his study of different available learning rules, Wilamowski (2003)
found that the delta or gradient descent backpropogation rule is used so
heavily because it always leads to a solution that is close to the optimum.
Fadlalla and Lin (2001) surveyed 40 finance applications of neural
networks and found that all seven authors who reported their learning rule
used some form of the delta rule. Many applications in the oil and gas
industry use the backpropogation rule as well (Mohaghegh et al 2001).
Nguyen and Cripps (2001) found that after experimenting with multitudes
of architectures, standard backpropogation was found to perform the best.

One criticism is that backpropogation takes longer than other
options to converge on a solution (Wilamowski 2003). However, with the
advances in computing power, the speed of learning isn’t a significant
problem. Additionally, the extra time is worth the generalization
capabilities that backpropogated networks provide (Mohagheh et al 1996).
Another criticism of the backpropagation delta learning rule is that it has the potential for oscillation (Wilamowki 2003). This means that it often finds local minima on the error surface. To help smooth out this process, many authors vary the learning rate or add a momentum term to the learning algorithm. The learning rate indicates how far in a given direction the weights will be adjusted each iteration. The complex error surface of multi-variable problems often necessitates a small learning rate to slowly and smoothly guide the weight shifts toward the optimum configuration. But, if the learning rate is too small, the network will take longer to train and there is more potential to get stuck in local minima. Trial and error is the only way to determine the best learning rate, although one author suggested 0.1 as a suitable starting point (Mohaghegh et al 1996).

Adding a momentum term can also help a solution to be found quickly that is close to optimal. The momentum term helps to provide stability by averaging the past weight changes and adding that to the current weight change. So, if the previous weight changes were all in the same direction as the current change, momentum accelerates the current weight change. If the previous weight changes were in the opposite direction of the current change, momentum slows the current weight change. This allows the learning algorithm to exit local minima on the error surface but to settle into the global minima in most cases. The momentum term must be determined through trial and error, but Mohaghegh et al (1996) offered 0.6 as the learning rate used in their network.
Evaluating Neural Networks

In order to gauge whether or not a neural network is successful, metrics must be put in place to test its ability to make predictions. This is done by reserving a certain portion of the known data set and not exposing the network to it during the learning process. Then, the networks prediction for the data is compared to the actual value but no adjustment is made to the network to compensate.

Many previous authors have measured the efficacy of their network by comparing its predictions to a more traditional approach, however this assumes that the traditional approach was a good model to begin with. Fadlalla and Lin (2001) surveyed many neural network applications and found that the ways neural networks are evaluated can be loosely classified into three categories: correctness, profitability, and risk. For this study, the neural network will be evaluated based on a correctness metric, more specifically the percent error between the target and prediction values.

Nguyen and Cripps (2001) suggested that both the absolute percent error and the mean absolute percent error be used to empirically evaluate a neural network. The absolute percent error is simply the percent difference between the networks prediction and the target for an individual property. The mean absolute percent error is the average of the
individual errors. Mann and Ayala (2009) considered their model to be sufficient when the percent error was within +/-10% for all data points and normally distributed about zero. However, Nguyen and Cripps (2001) used 5% absolute error as their metric based on it being generally accepted by the investment community.

In addition to percent error, crossplots are used to graphically show the prediction versus the actual, with all points falling on the unit slope line being desired. The correlation coefficient, “R”, can easily be computed to test the strength of the relationship between the prediction and target values. Mann and Ayala (2009) also used a histogram to make sure the errors were centered on zero percent and did not vary significantly.

**Artificial Neural Networks in the Oil and Gas Industry**

Artificial neural networks have been used successfully in a variety of optimization projects in the oil and gas industry, according to Mann and Ayala (2009). They noted an increase in the use of neural networks in the natural gas industry lately due to its applicability and accuracy. Mohaghegh et al (1996) asserted that the key to using ANN in oil and gas applications is to observe, recognize, and define problems so that neural networks can address them. They recognize that neural networks are not a panacea for the oil and gas industry, but can help solve problems that were not possible using conventional computing.
Conclusion

Artificial neural networking is a unique statistical modeling tool that borrows heavily from biological concepts. When compared to conventional techniques, neural networks have many benefits including their speed, ability to capture extreme values, and non-linear capabilities. If trained using a sufficient data set and model parameters, a neural network can be used for a variety of prediction and forecasting applications. In the literature, many studies exist applying this concept successfully to real estate, financial, and petroleum industries. However, no study was found applying neural networks to assess real estate purchases for the purpose of natural gas investment. For the reasons stated previously, neural networks are a natural choice for such an application and would be expected to yield a decision tool suitable to aid investors.
Chapter 3: Methodology

In this study, a neural network was created and trained based on expert evaluations of theoretical properties. Each of these theoretical properties exhibited different combinations of nine key factors that influence a property’s value. The neural network’s predictions were then compared to the expert’s opinions to assess the network’s efficacy as a decision aid. If successful, the properly trained neural network could help guide investment activity in the Marcellus shale.

Data Collection

No historical data exists publicly that could be used to train the neural network to recognize the investment potential of a given property. While an oil and gas company does have to disclose that a lease was signed and publicly file a drilling permit, no information on the dollar value of the lease or the amount of gas collected must be disclosed. So, it was necessary to collect the data in a more indirect manner. To do this, a list of theoretical properties with varying levels of the nine factors was given to the investment fund manager to score on a scale from 1 to 5. This particular expert has an in depth knowledge of the Marcellus region and
the oil and gas industry in general. The expert’s rating was a proxy for the value of the land as an investment option.

The ranges for the nine factors were selected based on research and the opinions of industry insiders. The input data is not completely random because that would create situations that do not reflect real properties or investment decisions. These properties would be difficult to evaluate fairly and would skew the data. For example, if the prospective property is adjacent to a well already in production, it is unlikely that the landmen would rate it poorly as that area is already being developed. Instead, a combination of completely random and partially random variables was used to get data that more accurately represents the current trends in the Marcellus region. The process of creating the data for each variable is described in detail in the next section.

In addition to the list of theoretical properties, the expert was also given a written description of the rating scale to be used, as shown in the Appendix. It is important to note that the rating scale is balanced and centered over a neutral or average score. A scale of 1 to 5 was chosen because it is large enough to allow the expert to differentiate between properties but still small enough to have a meaningful verbal descriptor.

**Independent Variables**

Based on research in the oil and gas industry and expert input, nine factors were chosen for this study. The combination of nine factors were
chosen due to their high perceived influence on the value of a piece of land for investment purposes. Cost per acre, other revenue sources, distance to pipelines or wells, gas price, geologist score, landsmen score, months until existing lease expires, royalty percentage, and lease terms score were selected as independent variables. The table below shows a summary of the input variables. A complete statistical analysis and histogram for all input variables is shown in the Appendix.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost ($/acre)</td>
<td>6022.3</td>
<td>1157.6</td>
<td>4000</td>
<td>8000</td>
</tr>
<tr>
<td>Other annual rev. (% of cost)</td>
<td>11.0</td>
<td>15.4</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Dist. to pipe/wells (miles)</td>
<td>5.5</td>
<td>4.6</td>
<td>0.5</td>
<td>26</td>
</tr>
<tr>
<td>Gas Price ($/mmcf)</td>
<td>6.0</td>
<td>1.4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Geologist Score</td>
<td>8.3</td>
<td>2.2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Landsmen Score</td>
<td>6.1</td>
<td>1.6</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Existing lease expires (months)</td>
<td>34.3</td>
<td>17.4</td>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td>Royalty (%)</td>
<td>16.6</td>
<td>2.7</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>Lease Terms Score</td>
<td>5.7</td>
<td>3.0</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1: Input variable summary

**Cost per Acre**

One the most significant drivers of value for investment is the initial upfront cost for acquiring the land. For the purpose of this study, it will be assumed that both mineral and surface rights will be purchased for any given parcel of land. This variable will be measured in the dollars per acre. The range of costs studied is $4,000 to $8,000 per acre, which is based on the typical prices for a piece of land within the core of the Marcellus shale at the current time.
The range of values is derived from research on informal landowner forums and word of mouth from industry insiders. The input data was generated randomly within that range and is uniform, as seen on the histogram in the Appendix. It is expected that as the cost per acre increases, the expert’s rating should decrease. This is because more expensive properties are less attractive and generate a smaller return on investment.

**Other Revenue Sources**

To help offset the upfront cost of acquiring land, the investment fund wants to consider other potential revenue sources that may be present on the property. Some examples of this are houses that could generate rent, timber, storage, or potential for pipeline easements. To consider all of these possibilities equally, any potential revenue sources will be measured as recurring annual revenue measured as a percentage of the purchase cost. For example, if the total purchase cost was $40,000 and there is a house and barn on the property that can be rented for $2,000 per month or $24,000 per year, the other revenue factor would equate to 60%.

The range studied for this variable is 0% to 100% of revenue as a percentage of purchase cost. This input data was generated randomly using a gamma distribution that was then scaled up to encompass values
within the designated range. A gamma distribution was used because it accurately models positively skewed data when random variables are greater than zero. This was necessary because most properties do not have a significant source of other revenue, so it is much more likely to be on the lower end of the scale. There are very few properties that can carry a significant portion of their own cost through other revenue sources, but enough that it was important to consider in the model. The Appendix contains a histogram and summary data that exhibits the gamma distribution. As this factor increases, it is expected that it would become more attractive to investors and be rated more highly by the expert. This is because if there is a significant amount of incoming revenue, it would decrease the cost for an investor to hold the property until it is developed for natural gas.

**Distance to Pipelines or Wells**

This factor measures the distance of the property under consideration to the nearest natural gas pipeline or well in miles. If a potential property is close to an existing pipeline or well then there will be less need for drilling companies to construct new infrastructure to get gas from the property to market. This makes it more attractive and likely to be incorporated into a production unit, as the cost to lay new pipelines is very substantial. This variable also helps ensure that the property is close to a
well-developed area where natural gas has already been found, again increasing the chances that the property will be of interest to drilling companies.

The range of interest studied is 0.5 miles to 26 miles, anything more than 26 miles starts to be cost prohibitive to drilling companies. The input data was randomly generated using a gamma distribution for two reasons. First, the investment fund wants to focus more on properties that are closer to other wells or pipelines, so it is important to have those values well represented in the data set. Also, with the increasing rate of development in the Marcellus shale, it is becoming more likely that any given property will be close to some kind of infrastructure. The gamma distribution accurately models the increased likelihood that a property under consideration will be close to infrastructure. However, the shape and scale parameters were adjusted so that there were some theoretical properties in the data set that are far from pipelines and other wells. Because the distance to pipelines and other wells is representative of the development trends and existing infrastructure, it is expected that the expert would rate properties more favorably as the distance decreases.

**Gas Price**

Another important factor is the current price of gas, as measured in dollars per thousand cubic feet (mcf) by the NYMEX natural gas futures.
As the gas price increases, so does the amount of drilling and new well completions. According to a Congressional summary, the estimated median break-even cost for a drilling company is about $6.64 per mcf (Andrews et al 2009). So, unless the gas prices are above that, it is unlikely that significant new drilling will take place. As of November 2010, the NYMEX gas futures were just under $4 per mcf, with forecasts estimating it to rise significantly when the economy picks up (Durham 2009).

![NYMEX Natural Gas](image)

For this study, prices between $4 and $8 per mcf were considered because it represents the forecasted direction of gas prices. The input values were randomly generated and exhibit a uniform distribution. As well drilling activity increases with natural gas prices, it would be expected that...
the expert would rate properties higher as the gas prices rise. If more drilling is happening, there is a higher demand for property to drill on, which increases the likelihood and speed with which any natural gas investment would be profitable.

**Geologist Score**

While there are many business related factors that affect investment decisions, the underlying purpose behind the investment is to find natural gas. Consequently, it is important to have a geologist evaluate any potential purchase to make sure the geological factors align with the business needs. A geologist would look at seismic images if available to observe subsurface characteristics to identify natural gas prospects. Also, a geologist would look at information like the shale thickness, thermal maturity, and total organic content to estimate how much gas is trapped under a given property.

With all of the available resources, it is possible for a good geologist to determine with some certainty if quality natural gas is present and rate a property on a scale from one to ten with respect to geological aptitude. However, due to the homogeneity of the Marcellus, the geology would not vary enough to effect the investment potential unless there is something significantly wrong. Essentially, unless the property sits over an empty cavern or odd fault line, it would be geologically sound to produce
gas. To represent this accurately, the input variable was generated with 10% of the data being randomly uniform between one and three and 90% of the data being randomly uniform between eight and ten. This means that 10% percent of the theoretical properties will exhibit a geologist score signifying a major issue with the property. This was done to create a data set that more accurately represents what the investment fund will encounter in the real world. As the geologist score increases, it is expected that the expert would score the property higher.

_Landsmen Score_

In addition to evaluation by a geologist, it is imperative to have landsmen weigh in on any potential land purchases. Landsmen have the benefit of being embedded within the communities of the target area in northeastern Pennsylvania. They have access to critical knowledge about the land, gas development, and people involved in any purchase. Much of this information may not be publicly available and is hence advantageous from a strategic standpoint. Landsmen can identify the trends in drilling activity, talk with landowner groups, and locate potential purchases.

Because of their unique knowledge, a score from landsman on a scale from zero to ten will help the model weigh intangible factors. However, the landsmen score is heavily dependent on other variables in the model. For example, it is unlikely that a landsmen will rate a potential property poorly if it is located adjacent to a well already in production.
Consequently, the landsmen score data was created by taking into account the distance to other wells or pipelines, gas price, and months until the existing lease expires. As described above, one strength of landsmen is that they can evaluate intangible factors that may be present. To model this, a random point value was added to each landsmen score. So, the landsmen score variable is based 90% on the three variables listed above and 10% on a random variable. This creates landsmen score data that is realistic but still somewhat variable. The relationship between the landsmen score and the three variables that comprise it is shown in the Appendix. As designed, the landsmen score increases with rising gas prices and decreases with newer leases and far away properties. Because the landsmen score indicates the speed with which a natural gas investment would be profitable, the expert would tend to rate properties higher as the landsmen score increases.

*Months Until Existing Lease Expires*

By some estimates, up to 80% of the core Marcellus shale area is already under lease with drilling companies. So, this model will assume that any land being considered for purchase is already under lease. The amount of time left on that lease, measured in months, is very important for potential investors. If a drilling company does not drill on a piece of property it has leased within the contractually stipulated time frame,
typically five years, it forfeits all rights to the property and must sign a new lease. However, if they drill in the specified time frame, they maintain the right to continue collecting gas from the existing well. If a lease is about to expire, the drilling company holding the lease must either drill, which would generate royalty payments for the landowner, or they must sign a new lease and pay a new lease bonus to the landowner. The expected profit therefore increases as existing lease comes closer to expiration.

This model will evaluate leases that are between 6 and 60 months of expiring. The input data was generated randomly using a uniform distribution and then rounded to the nearest six-month interval. Since a lease that expires soon would generate income quickly, the expert is expected to rate a property higher as the lease becomes closer to expiration.

**Royalty Percentage**

In the long term, the most profit is to be found in the form of royalty payments that are calculated based on the amount of gas that is extracted from a landowner’s property. Assuming typical production numbers at a gas price of $4.50 per mcf and a 12.5% royalty percentage, landowners can expect around $7,600 per acre per year in royalty payments. With wells producing upwards of 25 years, royalty payments generate significant long-term revenue.
As such, it is important that leases have a high royalty percentage, with values from 11% to 22% being used for this study. The royalty percentage is partially dependent on the months left on the existing lease. This is because older leases tended to favor the drilling companies as opposed to the landowner and consequently have lower royalty percentages. The royalty percentage data was generated by taking into account the months left on the existing lease and a random component. The existing lease accounts for 90% of the royalty percentage with the random factor accounting for about 10%. The relationship between royalty percentage and months until existing lease expires is shown in the Appendix. As designed, the royalty percentage is smaller with older leases that expire soon. It is expected that the expert will rate properties with a higher royalty percentage more favorably as they would generate the most long-term revenue.

**Lease Terms Score**

Since only properties with existing leases are being considered in this model, it is important to evaluate the terms of the lease. Particularly in older leases, there are many clauses and stipulations that might be unfavorable to an investor. For example, there is sometimes a renewal clause that calls out the price at which the drilling company may renew their lease after the current one expires. As the lease prices in the
Marcellus shale have increased significantly in the last five years, the property would most likely bring in more money on the open market.

To account for this in the model, a variable ranging from one to ten was added representing the favorability of a lease as evaluated by an attorney. This data was generated randomly within that range for inclusion in the model. It is expected that a lower lease terms score would result in a lower rating by the expert.

**Dependent Variable**

The dependent variable used in this model is a rating given by experts on a scale from 1 to 5. A score of 5 means that the combination of independent variables characterizes a parcel of land that is excellent for investment purposes. A score of 1 denotes a very poor investment with a 3 signifying an average investment. The score given by the expert correlates to the market value of a natural gas property. This artificially constructed variable is a necessary stand-in, as lease price information is not available publicly.

**Creating the Neural Network**

To create the artificial neural network, Matlab Neural Network Toolbox version 6 was used (Demuth et al 2008). The Neural Network
Toolbox is capable of designing, implementing, visualizing, and simulating neural networks. The user can select functions that automatically create networks with default architectures or fully customize their own network. Both graphical user interfaces and command line functionality can be used. Matlab was selected because it is readily available, easy to learn, and fully customizable.

**Creating the Program**

For this study, a program was custom developed to create a neural network utilizing best practice architectures found during the literature review. The program was created by running the graphical user interface, copying that code, and then modifying it. The initial program in text form can be seen in the Appendix. The function used to create the new network was "newff", which generates a feedforward backpropagation network. This function takes information such as the variables, number of hidden units, and training rule as arguments. According to the best practices found in the literature, the network was initially created with one hidden layer of 25 neurons and a gradient descent learning rule with momentum. Of the 300 data points, 80% was used for training, 10% was used for validation, and 10% was saved to test the neural network.
Program Output

The program was written so that it would store the network’s prediction and the network itself as a variable for analysis. This way, the network can be used to evaluate potential land acquisitions in the future. The network’s predictions were put into an Excel spreadsheet and compared to the expert’s evaluation to calculate the absolute percent error and the mean absolute percent error. This was used to evaluate the neural network, as discussed previously.

In addition to this, several plots showing the performance of the network were generated automatically. A cross plot shows the target or expert’s score versus the network’s output or prediction. It is desired that all points fall on the unit slope line, as that would mean there was no difference between the target and prediction. The correlation coefficient, “R”, is also given with this plot and measures the strength of the relationship between the target and network output. An “R” value closer to 1 means that there is a strong correlation and the network is able to predict the expert’s evaluation. Figure 3 shows the four cross plots that were generated showing the performance of the training, validation, testing, and all samples. For this example, the network learned the training set very well, did not perform too well on the validation set, and did relatively good on the test data set. One possible reason that the validation correlation coefficient was so low is that the network started far
away from the final state, so initial validation checks performed very poorly. Once the network settled into its solution, the test set was exposed to the network and performed relatively well.

![Figure 3: Example cross plot generated by Matlab](image)

Also, a training performance plot is generated showing the performance of the network on the training, validation, and testing data sets. The independent variable is the number of epochs or training iterations and the dependent variable is the mean squared error. It is desired that as the number of epochs increases, the error decreases to an acceptable level. In the example shown below, the network achieved its...
best validation performance after 5 epochs. After that, the training mean squared error (shown in blue) starts to decrease very quickly meaning that the network is starting to learn the training set too well and will likely lose the ability to make good predictions. Concurrently, the validation error (shown in green) increases quickly and that is what causes the neural network to stop iterating. This training performance plot corresponds to the network shown in Figure 3, which suggests that this network achieved a strong ability to generalize and make predictions.

Figure 4: Example training performance plot
Iterating

Throughout the literature review, the consensus was that no formula exists to calculate the best neural network architecture. Instead, an iterative process was suggested in which the experimenter varies different parameters until the neural network created yields suitable results. Further, the program used randomly assigns the initial weights within the neural network. What this means is that using the same code and parameters to create a neural network will generate a different result every time. Consequently, it is very important to use an iterative process when constructing neural networks. For this study, the most widely used architecture in the literature review was the starting point. From there, if the network did not perform as desired, other parameters were used. Each unique architecture was used to generate fifteen neural networks and the network with the lowest mean absolute percent error was selected.
Chapter 4: Results

After creating many neural networks using the initial parameters specified in the previous sections, none of the networks yielded the desired results. The best network created using that architecture had a mean absolute percent error (MAPE) of 12.5% with a maximum absolute percent error (APE) of 112%. This was not within the criteria for success specified previously of a MAPE less than 5% and an APE no larger than 10% for any single observation. Consequently, a more exhaustive search using different architectures was necessary.

Network Iterations

To help guide the exploration, a full factorial experiment was generated with four different factors. The four architecture parameters selected were the number of hidden nodes in each layer, the number of layers, the training algorithm, and the transfer function to the hidden layers.

The number of hidden nodes in each layer was varied between 25 and 50 and the number of hidden layers was varied between one and two, which falls within the range found in the literature review. If fewer hidden
units are used than this, the network is likely to be biased. If too many are used, then the network starts to lose its ability to generalize.

Two different training algorithms were used, a basic gradient descent backpropogation technique (“traingdx”) and the Levenberg-Marquardt backpropogration technique. The later technique interpolates between the Gauss-Newton algorithm and gradient descent. By doing so, it can often find a solution even if it starts far away from the minimum. This technique is very popular for curve-fitting and is the default training algorithm used by Matlab.

In total, sixteen different network architectures were specified and each one was executed fifteen times to generate a total of 240 networks. The network with the lowest MAPE was selected as the best for that given architecture. The network, outputs, MAPE, and correlation coefficients “R”, were saved for future reference.

Unfortunately, none of these networks met the original criteria to be deemed acceptable as a decision aid for natural gas investment. The lowest MAPE was 6.75%, achieved by network 16. While the MAPE is the most widely used measure of success, as seen in the literature review, it is important to look at the correlation value “R” for the test sample. The “R” value for the test sample indicates the networks ability to make predictions for properties that it has not previously been exposed to or trained on. The best network in that respect was network 5, which has a correlation coefficient of 77.2%. In addition to the MAPE and APE metrics, the
literature stresses the importance of the prediction errors being normally distributed about zero, which is achieved by both networks 5 and 16. Please see the Appendix for a probability plot and histogram of the errors for these two networks.

<table>
<thead>
<tr>
<th>Network #</th>
<th>Hidden Nodes</th>
<th>Training Algorithm</th>
<th># Layers</th>
<th>Hidden layer transfer function</th>
<th>MAPE</th>
<th>R Training</th>
<th>R Validation</th>
<th>R Test</th>
<th>R All</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>traingdx</td>
<td>1</td>
<td>logsig</td>
<td>12.53</td>
<td>0.650</td>
<td>0.394</td>
<td>0.702</td>
<td>0.633</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>traingdx</td>
<td>1</td>
<td>tansig</td>
<td>17.44</td>
<td>0.806</td>
<td>0.613</td>
<td>0.768</td>
<td>0.782</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>traingdx</td>
<td>1</td>
<td>logsig</td>
<td>18.99</td>
<td>0.760</td>
<td>0.634</td>
<td>0.678</td>
<td>0.743</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>traingdx</td>
<td>1</td>
<td>tansig</td>
<td>17.05</td>
<td>0.829</td>
<td>0.693</td>
<td>0.662</td>
<td>0.797</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>trainlm</td>
<td>1</td>
<td>logsig</td>
<td>11.63</td>
<td>0.984</td>
<td>0.499</td>
<td>0.772</td>
<td>0.906</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>trainlm</td>
<td>1</td>
<td>tansig</td>
<td>11.23</td>
<td>0.976</td>
<td>0.693</td>
<td>0.738</td>
<td>0.903</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>trainlm</td>
<td>1</td>
<td>logsig</td>
<td>10.52</td>
<td>0.999</td>
<td>0.196</td>
<td>0.725</td>
<td>0.896</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>trainlm</td>
<td>1</td>
<td>tansig</td>
<td>10.56</td>
<td>1.000</td>
<td>0.763</td>
<td>0.710</td>
<td>0.915</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>traingdx</td>
<td>2</td>
<td>logsig</td>
<td>18.53</td>
<td>0.761</td>
<td>0.711</td>
<td>0.680</td>
<td>0.744</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>traingdx</td>
<td>2</td>
<td>tansig</td>
<td>16.50</td>
<td>0.845</td>
<td>0.632</td>
<td>0.765</td>
<td>0.805</td>
</tr>
<tr>
<td>11</td>
<td>50</td>
<td>traingdx</td>
<td>2</td>
<td>logsig</td>
<td>17.63</td>
<td>0.790</td>
<td>0.644</td>
<td>0.756</td>
<td>0.774</td>
</tr>
<tr>
<td>12</td>
<td>50</td>
<td>traingdx</td>
<td>2</td>
<td>tansig</td>
<td>16.54</td>
<td>0.831</td>
<td>0.745</td>
<td>0.630</td>
<td>0.794</td>
</tr>
<tr>
<td>13</td>
<td>25</td>
<td>trainlm</td>
<td>2</td>
<td>logsig</td>
<td>10.04</td>
<td>0.979</td>
<td>0.650</td>
<td>0.771</td>
<td>0.920</td>
</tr>
<tr>
<td>14</td>
<td>25</td>
<td>trainlm</td>
<td>2</td>
<td>tansig</td>
<td>9.30</td>
<td>1.000</td>
<td>0.549</td>
<td>0.682</td>
<td>0.927</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
<td>trainlm</td>
<td>2</td>
<td>logsig</td>
<td>11.43</td>
<td>1.000</td>
<td>0.587</td>
<td>0.697</td>
<td>0.908</td>
</tr>
<tr>
<td>16</td>
<td>50</td>
<td>trainlm</td>
<td>2</td>
<td>tansig</td>
<td>6.75</td>
<td>1.000</td>
<td>0.583</td>
<td>0.614</td>
<td>0.898</td>
</tr>
</tbody>
</table>

Figure 5: Summary of Network Results
Chapter 5: Discussion

Other Neural Network Benchmarks

Although none of the networks could be considered successful based on the MAPE and APE metrics, it was possible to use other means to benchmark their performance. One way to do this was to compare the neural network to a more traditional method, regression. While much literature is available on the use of neural networks versus regression, none speaks to their capabilities for natural gas investment decisions. The basis of this study was the hypothesis that neural networks would perform the best, due to key characteristics discussed earlier. Since the network did not perform well, it is necessary to see if regression would perform better. The network with the largest correlation coefficient for the test sample was chosen for comparison.

Two regression models were created to investigate this, the first of which used all nine independent variables to predict the dependent variable. In the second regression, the independent variables with a P-value greater than 0.1 were deemed statistically insignificant and removed from the model. The following variables were removed from the model: other revenue, gas price, months until existing lease expires, and royalty percentage. For comparison purposes, the regression model was created
using all 300 samples and then used to predict the scores of 30 randomly chosen properties. This was done because 30 samples were used to calculate the correlation coefficient “R” for neural networking testing. Please see the Appendix for complete regression analysis.

The goal of both the neural network and regression models was to approximate the expert evaluation of theoretical properties. The assumption is that the expert would be able to replicate his or her own scores anytime. So, after the expert scored the initial 300 properties, 30 properties were randomly selected to be re-scored by the same expert. Those scores were then compared to the original scores to see if there was consistency in the results. Both the MAPE and correlation coefficient were calculated for those 30 samples.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAPE (%)</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network 5</td>
<td>11.63</td>
<td>77.2%</td>
</tr>
<tr>
<td>Regression</td>
<td>15.99</td>
<td>63.7%</td>
</tr>
<tr>
<td>Regression with variables removed</td>
<td>16.62</td>
<td>72.9%</td>
</tr>
<tr>
<td>Expert re-score</td>
<td>18.44</td>
<td>45.2%</td>
</tr>
</tbody>
</table>

Figure 6: Comparison of Neural Network, Regression, and Expert Re-score

It was found that Neural Network 5 had both the lowest MAPE and the highest correlation coefficient. More interesting though is that the expert re-score produced both the highest MAPE and the lowest correlation coefficient. This suggests that the expert may not be a very reliable source of data. Out of the 30 properties that were re-scored, the
expert reproduced the same initial score only 17 times, or slightly more than half.

This does not mean that the expert can’t accurately choose properties for natural gas investment, but that there are factors not considered in this model that might affect his or her decision. Since the factors considered in the model were chosen partially because the expert deemed them to be the most important, it very possible that the error is due to factors that cannot be measured. The error could be due in part to the order that the properties were presented in, environmental distractions such as noise, or many other factors that cannot be captured by the model.

The assumption in building any kind of model is that the data used to create it can be reproduced and is reliable. If that assumption is not met, the model cannot reasonably be expected to generalize correctly. In the case of the neural network and the regression in this study, the effects of having unreliable data is evident. The large percent error seen in all estimation methods may be partially or entirely due to the data given to them. Because it was not based on historical data, there is a chance that the data is biased or skewed in such a way that no technique would accurately pick out the pattern.
Neural Network Evaluation Metrics

While the MAPE is the most common measure of success for neural networks, as seen in the literature review, it is important to look at the correlation value “R” for the test sample, as that indicates the networks ability to make predictions. The MAPE takes into account the percent error all samples, including those with which the model was trained. So, 80% of the MAPE’s weight is determined by the network’s ability to copy the data that was used to create it. This was actually seen during this study several times when the MAPE was found to be around 10%. But, when the cross plots were checked, the training “R” was 100% and the validation and testing “R” were below 40%. What this means is that the network learned the training set too well and lost its ability to generalize or make predictions on new data. This can be a symptom of a network with too many hidden units. Nonetheless, the MAPE alone could not have caught this phenomena. It is very important to use the correlation coefficients or perhaps even the MAPE calculated only for the testing set. It is recommended that for future study, both the MAPE and correlation coefficients for all samples be calculated and considered when judging the abilities of a neural network.
Continuous versus Integer Values

One possible source of error in the model is the use of continuous and integer values. When the expert rated the list of theoretical properties, only integer values, or whole numbers, were used. Fractional numbers such as 3.2 were not allowed. This was done so that the ranking was more consistent, as each integer number allowed had a verbal descriptor. However, the output of the neural network and regression equations is a continuous variable. This can create large amounts of perceived error for a minor difference with the expert’s evaluations. For example, the MAPE for an expert score of 1.0 compared to a network output of 1.1 is 10%. Just that small amount of variation, if found consistently, pushes the network in the direction of being unsuccessful according to the MAPE and APE metrics.

One attempt to remedy this was to use a combination of scaling and rounding to get integer values from the network’s output. The output was first scaled linearly so that it fell within the desired range of 1-5. Then, it was rounded to the nearest whole number. For some network outputs, this slightly helped reduce the MAPE but negatively affected others. For example, if the expert gave the property a three and the neural network predicted a 3.4, rounding would make the results identical. However, if the network predicted a 3.6, it would be rounded up to a four and yield a percent error much higher than if rounding had not been used. Whether or
not the scaling and rounding helped varied between different network architectures and within them. Overall, scaling and rounding decreased the MAPE by approximately 1-3% for about two thirds of the networks and increased the MAPE by the same amount for the other third. For this reason, it was deemed ineffective and not used for final analysis.

Another way that was investigated to fix this problem was to use a completely different network type. For this study, a fitting model was used to approximate or fit an unknown relationship between the input and output variables. When this did not perform as expected, a completely different type of neural network was investigated: a pattern recognition neural network. Pattern recognition networks are widely used for classification and have the ability to sort input data into groups. The hope was that the pattern recognition network could definitively classify a property into one of five possible integer scores, similar to how the expert did. However, this option proved to be futile and yielded worse results than the fitting model originally used so it was not investigated in depth. The use of continuous versus integer values is a possible source of error in this model that could not be reconciled.

**Network Architecture**

Network architecture refers to the parameters selected to construct the network such as the training algorithm, number of hidden layers and
nodes, and hidden layer transfer function. These parameters are important because they can help make a network more robust if chosen carefully. Of the sixteen different network architectures investigated, the best neural network in terms of test correlation coefficient was Network 5. This network had one hidden layer consisting of 25 neurons using the logsig transfer function and was trained using the Levenberg-Marquardt backpropogation technique. When put in order from highest test correlation coefficient to lowest, the top five networks use only 25 hidden neurons. Beyond that, there is no discernable pattern in the network architecture as it relates to test correlation coefficient. This could be due to lack of generalization or prediction ability as the number of hidden units within a network increases. All other parameters do not seem to affect the network’s performance with regard to this metric.

When looking at the MAPE, the best network is Network 16. This network has two layers each with 50 hidden neurons, uses the tansig transfer function, and the Levenberg-Marquardt training algorithm. It is interesting that when looking at the MAPE metric, the training algorithm seems to have an important effect. Those networks trained using the Levenberg-Marquardt technique exhibit an average MAPE of 10.2% while those using a basic gradient descent backpropogation algorithm have an average MAPE of 16.9%. Additionally, all networks using the Levenberg-Marquardt algorithm outperform the other algorithm with respect to MAPE. Further, networks trained using the Levenberg-Marquardt algorithm went
through about 80% less epochs than the other training algorithm. What these pieces of information could mean is that the Levenberg-Marquardt algorithm traverses more of the error surface in a smaller number of epochs to find a less local minimum. Indeed, one author described the Levenberg-Marquardt algorithm this way:

*LM can be thought of as a combination of steepest descent and the Gauss-Newton method. When the current solution is far from the correct one, the algorithm behaves like a steepest descent method: slow, but guaranteed to converge. When the current solution is close to the correct solution, it becomes a Gauss-Newton method.* (Lourakis 2005).

This is consistent with the results found in this study. As discussed earlier, the training algorithm was not seen to affect the correlation coefficient of the test sample. This indicates that the Levenberg-Marquardt algorithm does not necessarily result in networks that make better predictions. The reason for this discrepancy can be seen in the correlation coefficients for the training samples. All of the networks trained using the Levenberg-Marquardt algorithm have R values that are very close to one, with the smallest being 0.976 for Network 6. The basic gradient descent training algorithm has correlation coefficients for the training sample on the range of 0.650 to 0.845. This suggests that the Levenberg-Marquardt learns the training sample very well but does not generalize any better than a basic gradient descent training algorithm. This again emphasizes the importance of evaluating neural networks not only based on the MAPE, but also on the correlation coefficients.
Random Weight Initialization

Before the network begins training, the weights connecting the input, hidden, and output neurons are initialized to random numbers. From there, the weights are adjusted during each epoch so as to minimize error between the network's output and the target values. With each epoch, the training algorithm traverses the error surface looking for the absolute minimum, but often gets stuck in a local minimum. The random initialization of weights dictates where on the error surface the network begins and affects the chances of finding the absolute minimum prediction error. Because of this, two networks created using the same code can be completely different. Indeed, in this study, the same results were never seen within networks with the same architecture. Creating a robust neural network then depends in part on luck and trial and error. The difference seen across all metrics between two networks created thirty seconds apart was remarkable. Although architecture decisions and a good data set affect the network as well, chance plays a part. For this reason, it is important to thoroughly vet all different network architectures many times to find the best network.
Chapter 6: Conclusion

In conclusion, the neural networks created did not meet the criteria for success initially set for this study. The networks did not exhibit a mean absolute percent error or absolute percent error that is considered by other experimenters to be satisfactory. In evaluating the neural networks, a different metric was found to be more helpful in assessing prediction ability. The correlation coefficient “R” calculated using only the data reserved for testing purely measures the network’s ability to generalize whereas the mean absolute percent error can be skewed by over-fitting.

However, neural networking should not be disregarded all together for the purpose of evaluating natural gas properties. Using the same data set as the neural network, two regressions were created and their output was compared to that of the expert. It was found that the best neural network outperformed both regression equations with respect to both the correlation coefficient and mean absolute percent error. However, when the expert was asked to re-score a sample of properties, the unreliability of the larger data set became apparent. If the data used to create it is unreliable, no model can be expected to generate accurate predictions.

While neural networking did not yield sufficient results to be deemed successful, it should not be discarded wholly. Much of the error present in the models may be due an unreliable data set. If this could be
overcome, neural networks could be evaluated more fairly. A properly trained neural network could reduce the subjectivity associated with an expert evaluating properties and be an asset to any investor.
Chapter 7: Recommendations

This study was designed to create a decision tool for natural gas investment that mimics the evaluations of experts. Neural networks were investigated for this purpose but fell short in part due to unreliable data. If the data were available, these suggestions for further study could yield a more robust and reliable neural network.

Use of Historical Lease Data

The model constructed in this study is based on evaluations provided by an expert in the field of natural gas investment. However, even experts can be wrong and incorrectly rate one property higher than another, which would in turn throw off the model. To combat this, it would be beneficial to create a model that is based purely on historical data of leases that have already been signed. The independent variables would be the same as the model in this study, but the dependent variable would be the dollar value of the lease that was signed. In this manner, the model would predict how much the drilling company is willing to pay for a real property exhibiting a set of characteristics versus how much the expert thinks the drilling company would pay for a theoretical property.
To accommodate this, a geographic information system (GIS) would need to be constructed to plot the pieces of property on a map and see how far they are from pipelines or wells. This would also aid potential investors in tracking development patterns of drilling companies. Such a system was developed and delivered to the investment fund using a database and Google Earth GIS software. The only missing information is the dollar value of historical leases, as that information is not available publicly.

Reducing Subjectivity

One goal of developing the model in this study was to reduce the subjectivity involved in making investment decisions. Investors can become overly excited or attached to an investment property, potentially leading to confirmation bias. To further reduce the influence of the effect, another study could focus on replacing the geologist and landsmen scores as factors in the model with more quantitative and tangible factors.

If historical data and a GIS system were used as described above, the geologist score could be more easily quantified. Maps detailing geological factors such as the gross isopach, thermal maturity, and percent organic content are publicly available. These maps could be overlaid on top of potential acquisition properties and the geological factors could easily be measured. So, for example, it would be possible to
see exactly how thick the shale is under and piece of property in the state of Pennsylvania.

The landsmen score would be more difficult to objectify, as landsmen specialize in rumors, people, and development patterns. Adding a factor such as the distance to a well drilled within the last three months might help capture the development trends and movement of the drilling companies. However, it would be necessary to have a very robust GIS application to implement the study of such a variable. Similarly, recent sales or leases of land within a certain proximity might indicate that an area is being targeted by drilling companies.

Comparison to Other Shale Plays

A thorough study of the other seven significant shale plays within the United States might also improve the strength of investment decisions. General trends in development, pricing, or other activity could be gleaned and applied to the Marcellus shale. This generalization could prove to be very significant as many of the drilling companies in the Marcellus shale have previous experience with the other shale plays. By better understanding the strategy behind the drilling companies, property purchases could be tailored accordingly.
Use of a Broader Rating Scale

One possible way to address the issue of continuous versus integer variables would be to use a broader rating scale. If the properties were rated from one to one hundred instead of one to five, some of the error might be reduced. For example, on the current scale, a prediction of 1.1 compared to a target of 1 is a 10% error. On a larger scale, a prediction of 90.1 compared to a target of 90 is a 0.1% error. While the percent error would be disproportionately larger for properties that fall lower on the rating scale, that could be balanced out by properties that are rated higher. Also, that effect is present in a rating scale of any size. A larger scale would also allow for more differentiation between properties, which might help the network to learn the relationships more effectively. While the smaller rating scale has the advantage of having meaningful verbal descriptors attached to each score (i.e. three means average), a scale from one to one hundred would still hold meaning and be manageable for an expert to evaluate.

Multiple Experts

To help diminish the influence of a single expert, it would be beneficial to have multiple experts rate the same set of properties. Then, the average score could be taken across the group of experts to obtain the
target score. This would reduce the bias in the model and result in a more accurate data set. In addition, the value of a property for natural gas investment is subject to market forces and opinions. In other words, it is only worth what someone is willing to pay for it. The model would be grounded by having multiple experts weigh in on the relative value of each property.
Bibliography


"Marcellus Shale Gas: New Research Results Surprise Geologists!"

*Geology.com - Earth Science News, Maps, Dictionary, Articles, Jobs.*


Appendix

### Table 3: Expert scoring rating scale

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>5</td>
</tr>
<tr>
<td>Good</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>3</td>
</tr>
<tr>
<td>Poor</td>
<td>2</td>
</tr>
<tr>
<td>Very Poor</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 7: Statistical descriptors for Cost
Figure 8: Statistical descriptors for Other Annual Revenue

Figure 9: Statistical descriptors for Distance to Pipeline/Wells
Figure 10: Statistical descriptors for Gas Price

Figure 11: Statistical descriptors for Geologist Score
Figure 12: Statistical descriptors for Landsmen Score

Figure 13: Statistical descriptors for Existing Lease Expires
Figure 14: Statistical descriptors for Royalty

Figure 15: Statistical descriptors for Lease Terms Score
\[ \text{Lands\_Score} = \left\{ 5 - \left(3 \times \left( \frac{\text{Dist\_To\_Pipe/Wells}}{25} - 1 \right) + \left(1 - \frac{\text{Gas\_Price}}{4} \right) + \left( \frac{\text{Exist\_Lease\_Exp}}{60} - 1 \right) \right) \right\} \times 1.8 + \text{rand}(-1,1) \]

Figure 16: Equation used to calculate Landsmen Score

\[ \text{Royalty\_\%} = \left( \frac{\text{Existing\_Lease\_Expires}}{15} \right) \times 2 + 12 + 2 \times \text{rand}(-1,1) \]

Figure 17: Equation used to calculate Royalty %

Figure 18: Scatterplot of Royalty (%) vs Existing lease expires (months)
Figure 19: Scatterplot of Landsmen Score vs. Existing Lease Expires

Figure 20: Scatterplot of Landsmen Score vs. Gas Price
Figure 21: Scatterplot of Landsmen Score vs. Distance to Pipeline and Wells

```
% Create Network
numHiddenNeurons = 25; % Adjust as desired
input = factors';
target = score';
net = newff(input,target,[numHiddenNeurons],{'tansig' 'purelin'},'trainlm','learngdm','mse');
net.divideParam.trainRatio = 90/100; % Adjust as desired
net.divideParam.valRatio = 5/100; % Adjust as desired
net.divideParam.testRatio = 5/100; % Adjust as desired

% Train and Apply Network
[net, tr] = train(net,input,target);
outputs = sim(net,input);

% Plot
plotperf(tr)
plotfit(net,input,target)
plotregression(target,outputs)
```

Figure 22: Initial Matlab program code
Figure 23: Probability plot of percent error for Network 5

Figure 24: Histogram of percent error for Network 5
Figure 25: Probability plot of percent error for Network 16

Figure 26: Histogram of percent error for Network 16
Regression Analysis: Score (1-5) versus Cost ($/acre, Other annual, ...)

The regression equation is
Score (1-5) = 1.97 - 0.000056 Cost ($/acre) 
+ 0.00156 Other annual rev. (% of cost) 
- 0.110 Dist. to pipe/wells (miles) - 0.0079 Gas Price ($/mmcf) 
+ 0.157 Geologist Score + 0.0951 Landsmen Score 
- 0.00259 Existing lease expires (months) + 0.0265 Royalty (%) 
+ 0.0018 Lease Terms Score

Predictor            Coef    SE Coef    T    P
Constant             1.9688    0.5746    3.43  0.001
Cost ($/acre)       -0.00005606  0.00003296  -1.70  0.090
Other annual rev. (% of cost)     0.001562    0.002481    0.63  0.530
Dist. to pipe/wells (miles) -0.110200    0.01499    -7.35  0.000
Gas Price ($/mmcf)   -0.00794    0.03718    -0.21  0.831
Geologist Score      0.15701    0.01740     9.02  0.000
Landsmen Score       0.09513    0.05530    1.72  0.086
Existing lease expires (months) -0.002594    0.004997   -0.52  0.604
Royalty (%)          0.02648    0.03089     0.86  0.392
Lease Terms Score    0.00180    0.01270    0.14  0.887

S = 0.654497    R-Sq = 56.4%    R-Sq(adj) = 55.0%

Analysis of Variance

Source            DF    SS     MS      F      P
Regression        9   160.494 17.833 41.63  0.000
Residual Error   290  124.226  0.428
Total             299  284.720

Source            DF    Seq SS
Cost ($/acre)     1     0.713
Other annual rev. (% of cost)     1     0.000
Dist. to pipe/wells (miles)       1  120.100
Gas Price ($/mmcf)    1     0.610
Geologist Score     1    37.201
Landsmen Score      1    1.489
Existing lease expires (months)    1     0.053
Royalty (%)           1     0.319
Lease Terms Score    1     0.009