The Pennsylvania State University

The Graduate School

Department of Energy and Mineral Engineering

THE IMPACT OF SAFETY PERFORMANCE MEASURES AND STRUCTURAL FACTORS ON UNDERGROUND COAL MINE PRODUCTIVITY 03-07

A Thesis in

Energy and Mineral Engineering, and Operations Research

by

Seyed Safa Eslambolchi

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2010

The thesis of Seyed Safa Eslambolchi was reviewed and approved* by the following:

R. Larry Grayson
Professor of Energy and Mineral Engineering
Thesis Advisor

Mark C. Radomsky
Director, Miner Training Program
Senior Lecturer in Energy and Mineral Engineering

Yu Zhang
Assistant Professor of Statistics.

Yaw D. Yeboah
Professor and Department Head of Energy and Mineral Engineering

*Signatures are on file in the Graduate School

ABSTRACT

Safety in the underground coal mine industry has always been a concern. The Mine Safety and Health Administration (MSHA) enforces safety rules and regulations by periodically inspecting the mines and recording the violation(s) and issuing citation(s) for them. Violations can impact the production of a mine depending on their severity. Some violations can result in temporary mine closure which significantly impacts the production, or permanent mine closure, in which case the production stops forever. After three disasters in 2006, the “MINER Act of 2006” was passed which strengthened the “Federal Mine Safety and Health Act of 1977” by mandating new laws and safety regulations. The MINER Act increased the chance of getting citations and consequently a significant increase in the penalty amounts in 2006 compared to the previous years. These issues could impact the production of the mines until.

In this thesis, MSHA databases were used to study the mine-size transitions and the productivity. In the mine-size transition study, the MSHA address/employment database was used to observe the mine-size transitions of 454 underground coal mines in the 5-year period of 2003-2007. This study showed that Very Large and Large mines were the most faithful to their size category. In the second study, MSHA address/employment, accident/injury, and inspection databases were used to study the impact of safety measures and structural factors on the productivity of underground coal mines in the period of 2003-2007. For this study, mines were categorized in three categories of Very Large/Large, Medium, and Small/Very Small. Using Forward and Backward Stepwise regression methods a robust multi-linear regression model was generated for each mine-size category to find the significant factors impacting the productivity of the mines in each category.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Dr. R. Larry Grayson, who has been a great mentor during my study and I have learnt a lot from. Without his valuable support, comments, patience, and encouragement I would have not been able to finish this study. I wish to express my warm and sincere thanks to Dr. Mark Radomsky for his great and thorough comments on my thesis and helping me to provide quality thesis. I am also indebted to Dr. Yu Zhang for his valuable comments and guidance during my work.

I would also like to thanks my wife, Fatemeh, for her persistent support and encouragement. Without her support I would not have finished the degree. My especial gratitude goes to my mother, Rayhaneh, whose unconditional love and support always made my dreams come true. She was the one who taught me the first English word and always encouraged me to learn the English language. Last but not least, I would like to thank my father, Seyed Ahmad, whose belief in me never wavered, and whose dreams for me were boundless.

Chapter 1

INTRODUCTION

Background of the Problem

Safety in underground coal mines has been one the most important concerns in this industry. Strict regulations govern the coal mining industry, yet accidents happen resulting in injuries and/or fatalities. Such incidents can have very high losses associated with them, such as worker compensation claims, medical expenses, non-compliance penalties, legal fees, rehiring costs, and production interruption. Some mine operators are putting a great amount of effort to achieve the zero-accident goal by educating and training employees and improving their safety culture.

Changes in the mine health and safety regulations by the passage of the MINER Act of 2006, which was stimulated by three mine disasters in 2006,have greatly affected the mining industry. Penalties for non-compliance with the health and safety rules and regulations increased significantly with the new legislation. Understanding and adopting the new rules and regulations and training the employees takes time, and those mines that could not adapt to the new requirements, such as installation of more breathing apparatus, in a short period of time incurred losses from penalties and were challenged to stay in business. In some case the penalties increased 3-4 times more than the previous years (Kinilakodi, 2009) (Kinilakodi, 2009). This too can impact the overall performance or productivity of a mine and even drive it out of business.

The decreasing trend of productivity of underground coal mines from 2003 triggers the incentive of studying the productivity of these mines and searching for the factors impacting them in this period. In the ACI Joint Summit in St. Louis, Grayson [2009] presented a graph showing a persistent decrease in the productivity of underground coal mines, while the employee hours worked and average number of employees went up. Although the employee hours worked have an inverse effect on productivity, it is not the only factor affecting productivity.

Problem Statement

Productivity is an important measure of efficiency used in many industries such as mining. Figure 1-1 depicts the productivity of U.S. underground coal mines in tons per employee hour worked from 1993 through 2009. This figure shows a persistent decrease in the productivity of these mines from 2003 forward. This phenomenon can be due to many factors including safety issues. Accidents in mines can interrupt the production. Mine disasters such as explosions can lead to fatalities and hazardous conditions, which after result in mine closure, sometimes permanently. This was the case for Sago mine, West Virginia, where 12 miners died due to an explosion in 2006; the mine ceased production in March of 2007 and later on December 12 of 2008, International Coal Group Inc., announced the permanent closure of the mines (International Coal Group, Inc., 2008). Depending on the severity of the accident, production can be stopped due to temporary mine closure, meaning no production would occur until the hazardous conditions are eliminated, or it can be slowed down for losing skilled workers. The hypothesis tested in this study is whether any of the proposed safety performance measures and structural factors impacted the productivity of underground coal mines. Also if any of them did, how? Findings from this investigation could shed light on the direction of finding the root cause(s) of the problems and possibly eliminating them.

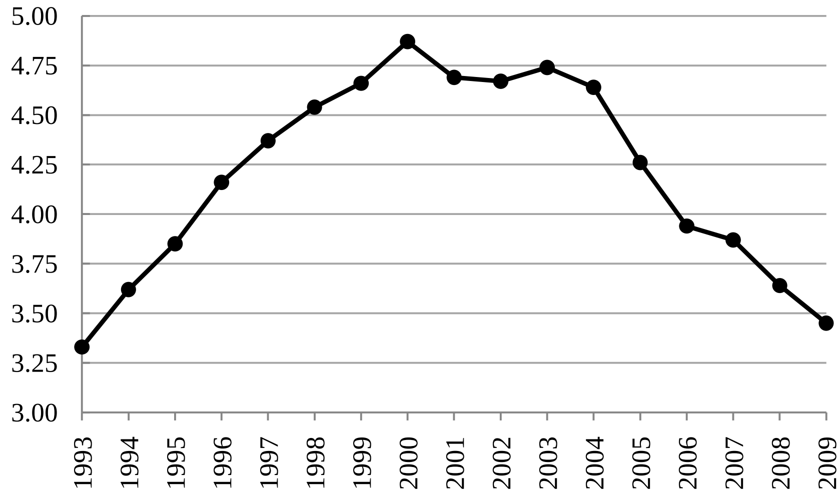


Figure 1- 1: Productivity of the U.S. underground coal mine operators (data source: MSHA's mine injury and worktime quarterly statistics coal data, 2010)

Objective and Scope of Research

This study has two main branches. First, the mine-size transitions of 454 mines of different sizes are studied in the period of 2003-2007. The objective of this part is to capture the trend of mine-size transitions during this 5-year period and also compare these transitions year by year to identify any unusual trend(s) for further investigation. This study does not scrutinize in more depth the reasons behind the unusual trends. The second part of this thesis examines the relationship between productivity and safety performance measures, and structural factors on a sample of U.S. underground coal mines. The objective is to find the most significant variables, from a list of proposed variables, apparently impacting the productivity of these mines through a multi-linear regression analysis. This study does not seek to generate models to calculate productivity from a series of variables, but to illustrate how the productivity changed by influence of the selected variables.

Research Methodology

To study the trend of mine-size transitions, sizes of the mines are identified in each year of 2003 through 2007. The total number of mines in each min-size category is derived for each year. These numbers are then converted to rates, then changes in these transition rates are studied in four one-year periods (2003-2004, 2004-2005, 2005-2006, and 2006-2007). In the productivity impact study, three multi-linear regression models are generated for the three mine-size categories, Very Large/Large, Medium, and Small/Very Small using Forward and Backward Stepwise regression methods in Minitab 15. The goal is to generate models with high R^2 and meaningful (significant) variables in them. Regression diagnosis is run by eliminating the data points with standard residuals of greater than 3.0 to reach the "best" models explaining the productivity versus the proposed variables.

Contribution to Knowledge

Many studies have been conducted on either productivity or safety in underground coal mines; however, little attention has been given to the combination of the two. This study aims to identify the safety measures impacting the productivity of underground coal mines to shed light on the relationship between productivity and safety measures. Having identified these influencing factors, the industry can target them to decrease or eliminate them or improve their productivity and profitability, yet keeping their workplaces safer for the miners.

Structure of the Thesis

Chapter 1 presents a brief background on the U.S. underground coal mining industry and its recent disasters and changes in the health and safety rules and regulations. The problem statement, the research methodology, and the contribution of this study to knowledge are further highlighted in this chapter.

Chapter 2 presents the past research conducted on the safety in underground coal mines using different safety measures. Some studies on the productivity in the mining industry are also critically reviewed. Finally the summary of existing literature on the intersection of the two topics, safety measures and productivity, are presented.

In Chapter 3 the methodologies used in this study are highlighted. First the data-gathering process is presented followed by the creation of the samples. Second, how to study the size transitions of the U.S. underground coal mines is fully explained. Third, some safety measure variables representing safety measures in underground coal mines are introduced and finally their relationship with productivity is presented in a general multi-linear regression model.

Chapter 4 presents the results of the mine-size transition study followed by the result of the productivity impact analysis using multi-linear regression analysis. Finally, in Chapter 5 summary and conclusions are presented followed by study limitations and recommendations for future work.

Chapter 2

LITERATURE REVIEW

Mine Safety and Health Administration

The Mine Safety and Health Administration (MSHA), a part of the Department of Labor in the United States, is a federal enforcement agency with the main mission of providing the nation's miners with health and safety in their workplaces. The Federal Mine Safety Health Act of 1977 (Mine Act) gives MSHA the authority to enforce health and safety standards for the U.S. mines (Weeks, 1991). Mine operators must comply with these standards or they are penalized. An accident that stops the activities for 30 minutes or more, or causes any accident or fatality must be reported to MSHA (Coleman, et al., 2007). All mine operators in the U.S. are required to report any injury under Part 50, Title 30, of the Code of Federal Regulations (Passmore, et al., 1985).

MSHA Databases

MSHA maintains a very detailed and thorough database on mines' employment, accidents and injuries. It has been collecting mine accident data for a long time and in 1983 they provided public access to these data (Coleman, et al., 2007). This database is by far the most comprehensive database available on mining accidents and injuries (Grayson, 2001). These data can be analyzed to shed light on the existing problems in the mining industry especially in the safety aspects. These databases can be categorized as Address/Employment, Accident/Injury/Illness, and the Data Retrieval System, each of which are briefly explained below.

Address/Employment Files

This database contains general employment and address information such as mine ID, company name, mine name, street address, commodity type, mine type, mine status, total production, and average number of employees for all the mines in the United States. These data are available to the public free in database format files, assorted by year.

Accident/Injury/Illness Files

This database provides detailed information on the accidents and injuries occurring in the U.S. mines, recognized by their MSHA Mine ID. These files contain the following information: date, time, state, county, commodity, accident type, sex, age, days lost, statutory days, restricted days, and a narrative description of each incident. These files are available to the public for free in database format files, assorted by year.

Data Retrieval System (DRS)

MSHA gives access to the detailed information on the mines' inspections, accidents, and violations through the Data Retrieval System website at http://www.msha.gov/drs/drshome.htm. In this website, mines can be searched in different ways, such as by Mine ID, Operator or Mine Name, Contractor ID or Name. For each mining operator, "Overview" section gives overall information on the number of injuries, injury rates, "hours worked," total coal production, number of citations, orders and safeguards, and penalties for each year. The "inspection" section contains information such as the type, begin and end date of the inspections as well as the inspection hours. The "accidents" section contains information on accidents and injuries including "accident date," "degree of injury," "classification," and "occupation activity". The “violations” section gives detailed information on the type of violations and the penalties associated with them. These information can be viewed for any period of time by specifying a “beginning date” and “ending date” for that period. (MSHA, 2010)

Coal Mining

Mines are generally divided into three main categories, i.e. metal, nonmetal, and coal. “Coal mines are further subdivided into surface and underground mining” (Coleman, et al., 2007). Surface coal mines have many advantages over underground coal mines, such as thicker coal seams, “wider choice of technologies,” and lower cost of extraction (Kulshreshtha, et al., 2001). Moreover, surface coal mines enjoy better safety performance and less hazardous working conditions compared to underground coal mines. This results in more efficient operations due to lower accident costs and higher productivity.

Underground Coal Mining

In underground coal mines, coal is extracted with different methods. Continuous mining and longwall mining are today’s most common underground coal mining methods (Weeks, 1991). The most productive mines employ both methods, for example, in a large mine, two to three longwall and as many as 12 continuous mining machines can operate in separate sections of the mines at the same time (Weeks, 1991). Longwall mining is generally more productive than continuous mining, which involves significant “idle” time. Weeks [1991] states that idle time in a continuous mining section is about “more than half the time of a workshift.” It is due to moving from one face to another or waiting for the shuttle car to come back from the conveyer belt for another load. On the other hand, longwall mining is more efficient since there is no need to leave any coal pillars for roof support and nearly all the coal is mined out; moreover, the machinery is not as frequently moved around, and finally with a conveyer belt at the face it eliminates the waiting time for the shuttle cars (Weeks, 1991).

Mine Size

It is very important to recognize the size of the mines due to the differences in the type of technology used, workers' skills, and the management system, all of which can influence the productivity of a mine. The literature categorizes mines by size differently. Grayson et. al [2009] uses the average number of employees as an indication of the mine size. He categorizes underground coal mines in 5 major categories, Very Large, Large, Medium, Small and Very Small. Larger mines use more robust methods and technologies such as longwall mining machinery. Szwilski [1987] indicates that the size and location of the coal reserve dictates the size of a coal mine. This means that the size of a mine can change over time, e.g., a depleting mine does not have much coal to extract and may not need many employees working there.

A coal mine or a section of it can be closed due to different reasons, such as depletion of the coal reserve, extreme violations and catastrophic events as well as extreme changes in the price of coal and regulations. Any coal reserve depletes eventually despite the size of the mine; however, larger mines take longer to deplete whereas smaller mines may deplete sooner. Extreme violations and catastrophic events can also happen in any mine despite the size of the mine. Szwilski [1987] suggests that small mines may not be able to survive extreme changes in the coal industry. He further indicates that not using the best mining method results in higher costs for the mine operator and eventually in a permanent mine closure.

Safety in Underground Coal Mines

The health and safety of miners was improved significantly in the last four decades by the Federal Coal Mine Health and Safety Act of 1969, generally referred to as the Coal Act. New requirements such as minimum air quality and quantity standards, dust control, methane concentration control were the major successes of this Act that affected mine planning, engineering and practice (Ramani, et al., 1999). The Coal Act was later amended by the Federal Mine Safety and Health Act of 1977, whose major provisions in underground coal mining industry were the creation of the Mine Safety and Health Administration, requiring four annual inspections for underground coal mines, mandating of miner training, and requiring mine rescue teams for all underground coal mines. In 2006, President George W. Bush signed the Mine Improvement and New Emergency Response Act of 2006 (MINER Act), which amends the Mine Safety and Health Act of 1977 with a number of provisions to improve the safety and health of the U.S. mines. The main impact of the MINER Act of 2006 was a significant increase in the citation penalty amounts, which can especially impact the overall performance of the more sensitive underground mines, smaller mines (Szwilski, 1987).

Safety Measures in Underground Coal Mines

After the passage of the Coal Act the rate of fatalities decreased dramatically (Kniesner, et al., 2004). However, despite all the safety rules, regulations and provisions, fatalities and non-fatal injuries continue to occur. Kucuker [2006] believes that the underground coal mining industry has one of the highest occupational fatality rates throughout the world. This is due to the confined environment of the underground coal mines and the exposure to the potential hazardous conditions such as bad air, methane, dust, and roof falls. Szwilski [1987] believes that large mines generally have better safety records compared to smaller mines, because they takes advantage of more professional engineers and employ better management systems. On the other hand, he also acknowledges the advantage of closer supervision in the small mines. Safety in underground mines is measured in different ways, some of which are presented in this part of the literature review.

Non-Fatal Days Lost Incident Rate

One of the most common safety measures widely used by industry is the Non-Fatal Days Lost Incident Rate (NFDL IR). These rates for each mine and the average of the industry is available on MSHA's Data Retrieval System website under the "overview" section. Ramani & Mutmansky [1999] believe that the non-fatal injury rate is a useful statistic that can be studied to identify and reduce "hazards and hazard realization potential."

Lost Workdays

Another indication of safety in underground coal mines is the number of lost workdays due to accidents, injuries, and illnesses. MSHA presents data on three different types of workdays lost indices, i.e., " … the number of workdays an injured miner takes before returning to work, the number of days of restricted work activity after returning to work, and the number of statutory days lost" (Coleman, et al., 2007). As mentioned earlier these data can be found in the Accident/Injury/ Illness files. "The average values of lost workdays are useful measures of overall safety performance in a mine, since they can reflect factors such as the use of personal protective equipment, the effective use of first aid and rapid access to medical care, or company policies concerning return-to-work practices" (Coleman, et al., 2007). Coleman et al. [2007] used lost worktime measures as an alternative metric to evaluate the safety and health performance of a job. They first examined the distribution and summary statistics of all the injuries reported to MSHA in the period of 1983-2004, during which 31,515,368 workdays were lost due to mining injuries. They used a beta distribution to model the lost workdays and compared underground coal mines with metal/non-metal mines from 2000 to 2004. The results show that the probability of having 10 or more lost workdays in underground coal mines was 0.52, whereas in metal/non-metal mines it was 0.35. They also found that the ratio of average losses from the injuries, which involved continuous mining machines in 2003-2004 was 1.08 of that in 2001-2002, suggesting a risk increase in this type of operation. Finally, they conclude that their analysis helps the industry to identify operations with higher risk and evaluate their safety improvement plans more realistically to reduce injury risks.

Severity Measure

Severity Measure is a less used indication of safety that presents the severity of injuries and illnesses in the underground coal mining industry. Coleman et al. [2007] used not only lost workdays and restricted days, but also statutory days to calculate the severity measure for a mine. Moreover, they believe that statutory days serve as a surrogate measure of diminished working lifetimes. Using MSHA accident data for 1996, Grayson [2001] concluded that the larger the mines the higher its severity measure is. He further calculated the correlation between severity measure and various variables, i.e., seam height, average employment, age of miner, experience in job, experience in mine, and total mining experience to examine the possibility of generating a multi-linear regression model regressing severity measure on the proposed variables. The correlations, however, were very low and some illogical.

Inspections

To ensure the compliance of the mine operators with the safety rules and regulations MSHA does not solely rely on the incident reports received from the mine operators, they conducts frequent inspections, which are quarterly for underground mines and semiannually for surface mines (Weeks, 1991). Other than these regular inspections, workers may request an inspection or MSHA may initiate one for a special reason, such as “modification in ventilation or roof control plans” (Weeks, 1991). Generally the regular quarterly inspections take a long time. Depending on the size and condition of a mine, it can take up to several weeks during which MSHA inspector(s) spend hundreds of hours to examine the safety and integrity of the equipment, infrastructure, different sections of the mine, mine roofs, and ventilation systems. The inspections are either announced or unannounced. Mines are encouraged to remove mine hazards before the inspections or forced to rectify hazards found during the inspections (Kniesner, et al., 2004). Applying an econometrics approach, Kniesner et al,[2004] conclude that the MSHA inspections have a modest impact on improving the safety of underground coal mines. They believe that the cost of these inspections outweigh their benefits.

Penalties

If an MSHA inspector encounters a violation during an inspection, he records it and must issue the proper citation or order. Depending on the severity of the violation, the amount of the penalty can vary. Generally, penalties for non-compliance with safety regulations in underground coal mines are more substantial compared to surface mines. When a penalty is imposed, the chance of follow-up inspection is 100% and if the problem is not fixed then not only the fine escalates but also a withdrawal order is issued, meaning the miners cannot go back "… to work until an MSHA inspector verifies that the hazard has been corrected" (Kniesner, et al., 2004).

"MSHA's Office of Assessments proposes the penalties. Currently, a violation not reasonably likely to cause reasonably serious injury that is corrected promptly is assessed a $60 penalty. Violations that are likely to cause reasonably serious injury (so-called "significant and substantial" violations) are assessed according to a formula that considers six factors: 1) history of previous violations; 2) size of the operator's business; 3) any negligence by the operator; 4) gravity of the violation; 5) the operator's good faith in trying to correct the violation promptly; and 6) effect of the penalty on the operator's ability to stay in business. These factors are determined from the inspector's findings, MSHA records, and information supplied by the operator. In some cases where there are fatalities, serious injuries, or high levels of negligence, the regular penalty formula does not yield an appropriate penalty. In these cases, MSHA will make an enhanced civil penalty proposal, known as a special assessment, based on the same six factors" (MSHA, 2006). Underground coal mines not only have more frequent inspections but also stricter safety laws and regulations, which results in more substantial penalties for non-compliance with them (Kniesner, et al., 2004).

Productivity in Underground Coal Mines

Generally productivity is defined as the ratio of the production of coal in tons to the employee hours worked, which is known as labor productivity. Fieldler et al. [1984], Weeks [1991], and Grayson [2001] use the definition of productivity as tons of coal produced per man hour worked in their analyses. Szwilski [1988] believes that measuring the productivity of a coal mine as labor productivity is a poor indication of the actual productivity. He suggests that Total Factor Productivity (TFP), which can be calculated using Equation (2-1), is a much more accurate indication of the productivity. Szwilski [1988] states that the TFP not only accounts for labor productivity, but also for other principal resource factors such as capital investment, new technology, and mining equipment. Figure 2-1 depicts the "key ingredients" of production, which should be applied to increase the efficiency of the coal production (Szwilski, 1988).

$$TFP = Q/(aL + bK) \qquad (2\text{-}1)$$

Where

- Q : real product
- L : labor input (does not consider the quality of labor, training and education)
- K : capital input (does not consider R&D)
- a, b : percentage shares of L and K.

He further suggests using the modified Cobb-Douglas production formula as in Equation (2-2) with the inclusion of a factor representing the quality of coal reserve, or land, and to utilize this formula for mine production (Szwilski, 1988).

$$\ln\left(\frac{Q}{L}\right) = \ln C + \ln\left(\frac{K}{L}\right) + \ln L + \sum \delta Z \qquad (2\text{-}2)$$

where

- Q: real product
- L: labor input
- Q/L : labor productivity
- K : capital
- Z: control variable, such as management performance rating, time (years), size of mine/production rate, technology adjustment to capital, and skill adjustment on labor.

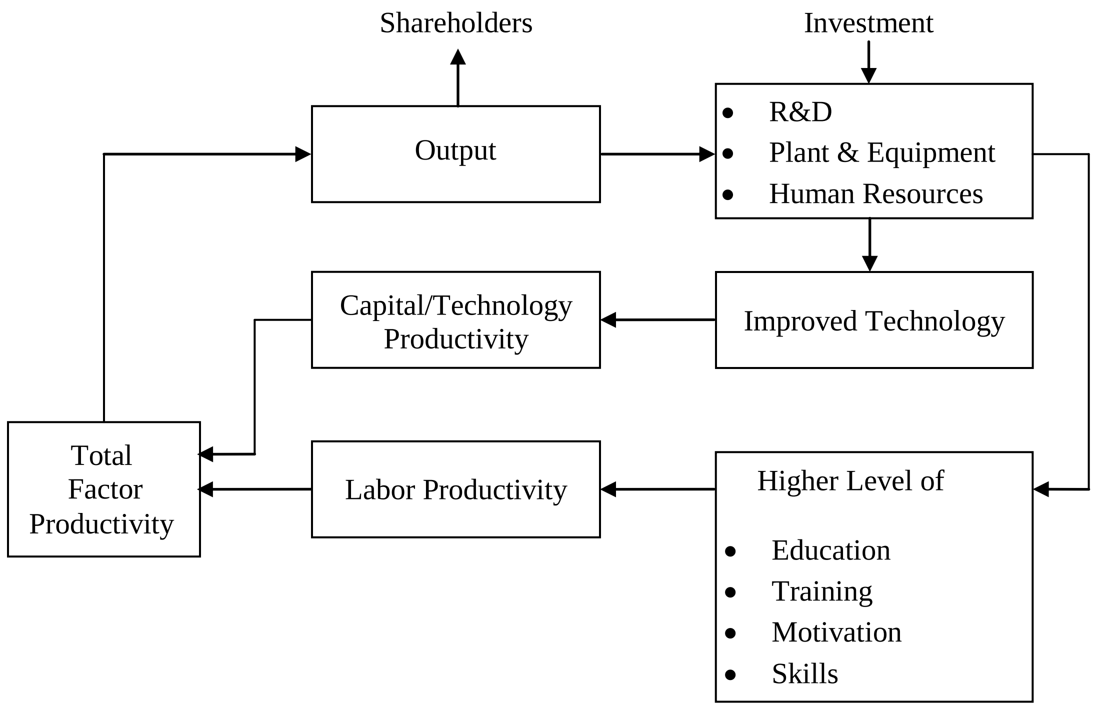


Figure 2-1: Significance of key inputs for total factor productivity (source: Szwilski, 1988)

Studying the MSHA data for 1996, Grayson [2001] investigates the productivity trend of 583 underground coal mines for five size categories, i.e., Very Large, Large, Medium, Small, and Very Small. The Large mines with an average productivity of 5.15 were the most productive mines followed by Very Large mines with an average productivity of 4.60. Moreover, Medium, Small, and Very Small mines placed afterward with an average productivity of 4.47, 4.32, and 3.59, respectively. Finally, he concludes that generally larger mines are more productive ones.

Many of the studies done on productivity in coal mines are historical, i.e., from the pre technology era. During those days labor was the most important part of the productivity studies. Today, with new technology and automation, especially in larger mines, more attention should be given to other factors interrupting production.

Factors Influencing Productivity of Underground Coal Mines

Underground coal mines generally have a lower productivity when compared to surface coal mines. This is due to the very different nature of these two types of mines, such as confined conditions in the underground coal mines and limitations in using machines and equipment compared to surface coal mines. The lower height of the coal seam is also a factor.

Grayson [2001] examines the relationship between the productivity of underground coal mines and some safety measure using MSHA accident data for year 1996. He examines the correlation between productivity and "severity measure" for the five mine-size categories, i.e., Very Large (-0.432), Large (-0.357), Medium (-0.262), Small (+0.003), and Very Small (+0.0001). Using Fisher's Z-transformation approximation, the significance of the correlations were tested at level of significance of α=2.5%. The results of these tests show that Very Large, Large, and Medium mines were statistically significant with a 95% level of confidence, whereas in the case of Small and Very Small mines the null hypothesis could not be rejected and thus their correlations were not significant. He further split the mines in each category to "high-productivity" and "low-productivity" ones to test the validity of the conclusion from the first tests. Again in the case of Very Large, Large, and Medium mines, the more productive mines were found to be the safer ones, whereas in Small and Very Small mines the more productive mines were the less safe ones.

Kulshreshtha et al. [2001] analyze the trend of productivity by studying the total factor productivity (TFP), and capital and labor productivity of an Indian coal company and its major subsidiaries to identify the major factors influencing productivity. They believe that technical abilities play a major role in determining the level of productivity of the underground coal companies, especially in the "Board and Pillar" method, in which the efficiency of the miner is reduced by poor ventilation and lighting, and "congestion of galleries with equipment." They find that factors such as difficult mining conditions and low skill levels lowered the productivity of the underground coal mines (Kulshreshtha, et al., 2001).

Szwilski [1987] suggests that coal reserves, capital, and labor are the principal factors in coal production. He states that productivity is a function of investment and sees the future long-term profit depending on an increase in investments not in cutting costs "… by reducing personnel and inventories" (Szwilski, 1987). Szwilski [1988] acknowledges that the labor costs are usually the largest cost component in the coal industry; however, he believes that the productivity can be improved by other means rather than just cutting the labor cost. He suggests that improving mining equipment, machinery, the skills of the workers and increasing motivation, morale, and pride with a better management can improve productivity.

Safety rules and regulations can also impact the productivity of coal mines. Szwilski [1988] believes that the Federal Coal Mine Health and Safety Act of 1969 significantly affected the productivity of underground coal mines. On the other hand, Weeks [1991] states that "… safety practices promoted by the Mine Act are consistent with management practices that promote increased productivity." Fieldler et al. [1984] investigate the effectiveness of Organization Development and Structured Management Training (SMT) programs in two underground mines. Safety increased substantially in both mines, while productivity also increased substantially in the mine which adopted the SMT.

One of the factors that can impact the efficiency of a mine is the costs it incurs due to accidents. In addition to the penalties that operators must pay due to injuries and/or fatalities resulting from accidents, they also need to pay for the cost of work-related illnesses, injuries and compensations. Camm et al. [2004] believe that the cost of injuries and fatalities to the industry are enormous and they should be considered in determining the efficiency of the mine. Depending on the severity of the accidents, the employer incurs some indirect "… costs associated with additional hiring and re-training, time delays due to the disruption of work processes, and the effects of workplace injury, exposure, or fatality on the productivity of coworkers who see themselves at heightened risk" (Camm, et al., 2004). These costs can result in an economically unviable situation and put the mine operator out of business.

In sum, to run a safe and productive underground mine, the first step is to identify the significant factors impacting the productivity of these mines. By scrutinizing these significant factors we can then proceed to fix the problem(s) and improve the productivity. The next chapter uses the insights from this literature review to identify the factors impacting productivity.

Chapter 3

METHODOLOGY

Introduction to MSHA Databases

Every year, the Mine Safety and Health Administration (MSHA) gathers data from registered mines all over the United States and generates data files containing information such as name, address, employment, production (coal), accidents, and injuries. For ease of use, the National Institute for Occupational Safety and Health (NIOSH) presents these raw data files in SPSS and dBase IV file formats, which can conveniently be used in spreadsheet and database software for further analyses. These files are available to the public for free on the NIOSH website[1]. These files are categorized in two categories, “Address/Employment Files (AE)” and “Accident/Injury/Illness Files (AI).” The former contains annual information on the mine’s name, address, type, status, total production (coal) and average number of employees as well as the total employee hours worked, whereas the latter provides detailed information on the time, type, and description of accidents, injuries and illnesses that occurred in the mines as well as the number of Statutory, Restricted and Lost Work days for each accident.

[1] http://www.cdc.gov/NIOSH/Mining/data/

Creation of Study Database and Sample

Creation of Database

"Address/Employment" (AE) and "Accident/Injury/Illness" (AI) files are the main sources of data for this study. Each of these files contains information on more than 25,000 mines of different types in the U.S. To create the database for this study only parts of these files, relating to underground coal mines, were required. Therefore, each file needed to be filtered such that excess information was eliminated and only information on underground coal mines remained. Since conducting the analysis on all of the underground coal mines is very time consuming and beyond the scope of this study, an appropriate sample was collected from the population of underground coal mines for the five-year period of 2003 through 2007. This sample needed to be designed so that it represents the mine-size structure of the underground coal mining industry as truly as possible. The next step was to introduce some variables that represent safety measures and structural factors of underground coal mines and then collect and/or calculate values for these variables for each of the mines in the sample. Microsoft Excel was used to organize, manage and store the data for further analysis. This section demonstrates how the database was created and introduces the variables used in this study.

Criteria for Filtering the Database

The first step in creating the database was to collect a sample of underground coal mines that reasonably represented the mine-size structure of the underground coal mine industry and had a significant share of annual production of coal from underground coal mines. To obtain this sample, four criteria were used, i.e., Mine Type, Mine ID, Mine Size and Total Production. Mine type was used to identify underground coal mines. Mine ID was the most appropriate piece of information to identify these mines, especially to track their mine-size transition over time. Mine Size was used to categorize the mines to conduct a separate analysis for each size category due to the similarities in characteristics of each size category. Finally, to study the productivity of underground coal mines, the Total Production of the mines was essential. In this section, these criteria are explained in details.

Mine Type

The only mine type of interest for this study is ‘underground coal mine’. The types of the mines are specified in the eighth column of the “Address/Employment Files,“ under “MINETYPE” by a number ranging from 1 to 14, where the number “11” represents “Underground-Coal” mines (MSHA, 1993). The “Filter” feature of Microsoft Excel was used to eliminate mines of other types and their corresponding information. The “Number Filters” was set equal to 11 to only show underground coal mines. This reduced the long list of more than 25,000 mines in each year, largely operations of different types, to approximately 1,000 registered underground coal mines per annum (Table 3-1).

Table 3- 1: Number of registered mines and registered underground coal mines, 2003-2007

Mine type	Year				
	2003	2004	2005	2006	2007
All mines	26,598	26,043	25,992	27,501	29,731
Underground coal mines	1,040	1,015	1,014	986	940

Mine ID

The Mine ID is a 7-digit number assigned to each mine by an MSHA district office. The first two digits of this number represent the state in which the mine is located. These numbers are unique, meaning once a number is assigned to a mine it cannot be used for another mine, regardless of the changes in the status or ownership of the mine (MSHA, 1993). This is a valuable property and the reason that Mine ID was chosen to address the mines in this study is because it is a unique identification, while a mine name, which can be shared by two or more mines is subjected to change. Moreover, having the ID of a mine, detailed information of that mine can be easily accessed on the MSHA website[2]. Mine IDs are located in the first column of both "Address/Employment " and "Accident/Injury/Illness" files, the major sources of data for this study.

To observe the mine-size transition of a mine over the 5 years of 2003 through 2007, it was necessary to have the information of that mine in all those years. This required careful investigation to identify these mines. Notice that a mine may have no records available for some years. Such mines were eliminated from the database. This reduced the number of registered underground coal mines to a total of 604. In other words, of approximately 1,000 registered underground coal mines, 604 had their records available from 2003 through 2007.

Mine Size

Mines of different sizes have different operating environments, conditions, equipment and technologies, and manpower structures as well as different cultures and supervision systems. In mines with advanced technologies and equipment, production is much higher than those with

[2] www.msha.gov

conventional technology. Generally mines with more advanced technology have more production and are larger sized. In this study, it was assumed that mines within the same size category are similar to each other; therefore, they were studied in their own size categories, instead of comparing mines from all mine-size categories. Separate models were generated for the mine-size categories, instead of generating one model trying to explain the productivity for all the mines of different sizes.

Mine size is a structural factor of a mine and it can be defined based on several criteria, e.g., total volume of coal reserve, total production, or average number of employees. In this study, the average number of employees was used since it can be easily derived from "Address/Employment" files, under the "AVENEMP" column. Grayson et al. [2009] also uses these criteria to categorize the mine size of underground coal mines. They introduced five categories, very large, large, medium, small and very small based on the average number of employees (Table 3-2). Mines with no employees over all the five years were eliminated from the database since they do not provide any information; however, those with an average number of employees greater than zero in at least one year remained in the database. This reduced the number of mines in the database from 604 to 454.

To clearly identify the size of a mine, all of its corresponding cells in the main spreadsheet were assigned a color based on the size of that mine. Blue, purple, red, orange and yellow represent very large, large, medium, small and very small mines, respectively, while black represents mines with no employees (Table 3-2). The number of employees of a mine can change for different reasons such as layoffs, so it is possible that a mine changes its size over time. For example a Large mine can become a Very large mine if its number of employees exceeds 250; on the other hand it can reduce in size to a Medium mine, if the number of employees decreased to 100 or less but not less than or equal to 50. This phenomenon is clearly observable by having the cells corresponding to mines colored according to the scheme in Table 3-2. In this study, a mine which kept its size over the period of 2003-2007 is called a “Faithful” mine. All the mines used in the regression analysis part of this study were of this type.

Table 3- 2: Mine-size categories based on average number of employees

Mine size	Color code	Average number of employees (EMP)
Very Large	Blue	EMP > 250
Large	Purple	100 < EMP ≤ 250
Medium	Red	50 < EMP ≤ 100
Small	Orange	20 < EMP ≤ 50
Very Small	Yellow	0 < EMP ≤ 20
No-Employee	Black	EMP=0

Total Production

Total Production is a very important piece of information sine it is used in the calculation of productivity. It can be found in “Address/Employment“ files under the “TOTPROD” column. These values were extracted from each year’s “Address/Employment“ files and added to the main Excel sheet. A coal mine can have no production in a particular year due to reasons such as being permanently or temporarily closed. A mine is permanently closed after it is depleted or when, according to MSHA regulations, it is not a safe mine to operate any more. A mine can be temporarily closed due to construction activities or when resolving serious hazardous conditions found by MSHA inspectors. A closed mine can have employees working in it, meaning the average number of employees in a mine with no production is not necessarily zero. This adds two more categories of “Small with no production” and “Very Small with no production” to consider in addition to the categories shown in Table 3-2. In the regression analysis part of this study, only

“Productive” mines, defined as those with a total production of at least 10,000 tons per annum, were used. If a mine remains Productive over the 5 years, it was called “Consistently Productive”.

Creation of the Samples

Two separate samples were generated for this study. The first sample was generated to study how the sizes of the underground coal mines changed over the period of 2003-2007. This required having information on each of the mines for all the 5 years of the study. Mines with no employees and no production were eliminated from this sample. This analysis consisted of the two recently added categories: “Small with no production” and “Very Small with no production” in addition to the six categories initially introduced in Table 3-2. The second sample was created for the productivity impact analysis. For this part of the study, Consistently Productive mines, which were also Faithful during the period of 2003-2007 were used. These limitations reduced the number of mines eligible for this sample. The five mine-size categories of Very Large, Large, Medium, Small and Very Small were initially considered for this part; however, necessary adjustments may be possible.

Mine-Size Transition

The average number of employees working in a mine can change from one year to another for different reasons: for example, a mine expansion creates more jobs; on the contrary, a depleting mine or a temporarily closed mine does not need as many employees. Changes in the average number of employees of a mine can result in a change of its mine-size category. Eight mine-size categories used in this analysis are Very Large, Large, Medium, Small, Small with no production, Very Small, Very Small with no production, and No-Employee, all of which have been defined in previous sections. For the period of 2003-2007, the number of mines in each category can be easily visualized using color codes, as defined in Table 3-2 and the Total Production column in the main spreadsheet. When the Mine ID column of a particular year is filtered by a specific cell color, only cells with that color are shown in the spreadsheet and the number of those cells is shown at the bottom left side of the spreadsheet window. To get the number of Very Large, Large, Medium, and No Employee mines the cells are filtered by colors blue, purple, red, and black, respectively. To get the number of Small and Very Small mines in each year, first, the MINE ID column of that year is filtered by colors orange and yellow, respectively, and second, the Total Production column of that year is filtered to exclude mines with no production. This can be done by removing the check mark next to the value zero. To get the number of Small and Very Small mines with no production in each year, again, the Mine ID columns are filtered by colors orange and yellow, respectively, and then the Total Production columns are filtered to show only the mines with zero production. Following this procedure the number of mines in each category can be derived for each year; however, this doesn't provide any information on the number of mines transitioning from one mine-size category to another one.

Assuming one-year-long transition periods, four periods of 2003-2004, 2004-2005, 2005-2006, and 2006-2007 needed to be investigated. In each period, the number of mines transitioning from category i to category j was determined and organized in a matrix, whose rows represent the categories of the mines in the starting year of the period and the columns represent the categories of the mines in the ending year of the period. Table 3-3 illustrates a one-step transition matrix from year k to k+1, where k=2003-2006. The middle elements of this matrix, a_{ij}s, represent the number of mines transitioning from category (state) i to category j, where i, j=1-8 stand for the mine-size categories of VL to NE (Table 3-3). The last column of Table 3-3, Row Sum, gives the number of mines in each mine-size category in year k, whereas the last row, Column Sum, represents the number of mines in each category in the next year.

To get a better sense of these numbers, these transitions can be expressed as rates. Equation 3.1 is used to calculate the transition rates.

$$r_{ij} = \frac{a_{ij}}{S_{k,i}} \tag{3.1}$$

Where:

r_{ij} is the rate of mine transitioning from size i to size j in a one-step transition (i, j=1-8),

a_{ij} is the number of mines transitioning from size i to size j in a one-step transition, and

$S_{k,j}$ is the total number of mines in size category j in year k (the starting year of the period).

Table 3-4 illustrates these transition rates for the period of year k to k+1, where k=2003-2006. The last row of this table, Percent Change, gives the percentage of change in each mine-size category in each period. Equation 3.2 is used to calculate these Percent Change values. A positive percent change for a category means an increase in the number of mines in that category, whereas a negative percent change shows a decrease in the number of mines in that category. A zero percent change in a category means that the total number of mines in that category didn't change during the transition period. This does not necessarily mean that no mine made a transition to or from that category.

$$\Delta_j = \frac{S_{k+1,j} - S_{k,i}}{S_{k,i}} \times 100 \qquad (3.2)$$

Where:

Δ_j is the percent change in mine-size category j from year k to year k+1,

$S_{k,i}$ is the total number of mines in category i in year k (the starting year of the period), and

$S_{k+1,j}$ is the total number of mines in the same category (j) in year k+1 (the ending year of the period.)

Table 3- 3: Mine-size transition matrix for the period of k-k+1 (in percent)

Year		k+1								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	Row Sum
k	VL	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}	$S_{k,1}$
	L	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	a_{27}	a_{28}	$S_{k,2}$
	M	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	a_{37}	a_{38}	$S_{k,3}$
	S	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	a_{47}	a_{48}	$S_{k,4}$
	S-NP	a_{51}	a_{52}	a_{53}	a_{54}	a_{55}	a_{56}	a_{57}	a_{58}	$S_{k,5}$
	VS	a_{61}	a_{62}	a_{63}	a_{64}	a_{65}	a_{66}	a_{67}	a_{68}	$S_{k,6}$
	VS-NP	a_{71}	a_{72}	a_{73}	a_{74}	a_{75}	a_{76}	a_{77}	a_{78}	$S_{k,7}$
	NE	a_{81}	a_{82}	a_{83}	a_{84}	a_{85}	a_{86}	a_{87}	a_{88}	$S_{k,8}$
Column Sum		$S_{k+1,1}$	$S_{k+1,2}$	$S_{k+1,3}$	$S_{k+1,4}$	$S_{k+1,5}$	$S_{k+1,6}$	$S_{k+1,7}$	$S_{k+1,8}$	S_t

Table 3- 4: Mine size transition rate matrix for the period of k-k+1 (in percent)

Year	Mine Size	k+1 VL	L	M	S	S-NP	VS	VS-NP	NE	No of Mines
k	VL	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	$S_{k,1}$
	L	r_{21}	r_{22}	r_{23}	r_{24}	r_{25}	r_{26}	r_{27}	r_{28}	$S_{k,2}$
	M	r_{31}	r_{32}	r_{33}	r_{34}	r_{35}	r_{36}	r_{37}	r_{38}	$S_{k,3}$
	S	r_{41}	r_{42}	r_{43}	r_{44}	r_{45}	r_{46}	r_{47}	r_{48}	$S_{k,4}$
	S-NP	r_{51}	r_{52}	r_{53}	r_{54}	r_{55}	r_{56}	r_{57}	r_{58}	$S_{k,5}$
	VS	r_{61}	r_{62}	r_{63}	r_{64}	r_{65}	r_{66}	r_{67}	r_{68}	$S_{k,6}$
	VS-NP	r_{71}	r_{72}	r_{73}	r_{74}	r_{75}	r_{76}	r_{77}	r_{78}	$S_{k,7}$
	NE	r_{81}	r_{82}	r_{83}	r_{84}	r_{85}	r_{86}	r_{87}	r_{88}	$S_{k,8}$
Percent Change		Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	———

Productivity Impact Analysis

Introduction

Different analytical techniques can be used to study the relationship between two or more variables. One of the most common methods is Linear Regression Analysis, in which the objective is to find a linear relationship between a dependent variable and one or more independent variables. In this part of the study the relationship of the Productivity of underground coal mines is studied against a number of independent variables representing safety measures and structural factors of underground coal mines that could impact productivity.

Multi-linear Regression Analysis

Researchers frequently use Multi-linear Regression (MLR) as an analytical method to find the relationship between some variables, which are mainly of two types, dependent and independent variables. In an MLR, there is usually one dependent variable whose linear relationship with two or more independent variables is sought. In other words, the objective is to generate a model that predicts the dependent variable values for different values of the independent variables. The word “Multi” refers to the case where there is more than one Independent variable in the model. If there is only one independent variable in the model, it is called Simple Linear Regression, which is basically the simple version of the Multi-linear Regression. The word “Linear” means that there is only a linear relationship between the variables and there is no multiplication of independent variables in the model. In other words, the dependent-variable value is the summation of independent variables, each of which is multiplied by a corresponding coefficient as shown in Equation 3.3.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad ; \quad i = 1, 2, \ldots, n \qquad (3.3)$$

Where:

Y_i is the dependent variable (response) of the ith observation,

$X_{i1}, X_{i2} \ldots and\ X_{i,p-1}$ are independent variables (predictors),

$\beta_0, \beta_1, \beta_2, \ldots and\ \beta_{p-1}$ are the regression coefficients, and

ε_i's are the error terms, which are assumed to be independent and normally distributed with a mean of zero.

Description of Variables:

Two types of variables, dependent and independent, are used in any linear regression analysis. In this study, Productivity plays the role of the dependent variable, while a few variables, representing safety measures and structural factors, play the role of the Independent variables. In this section, all the variables used in this analysis are defined and fully explained.

Productivity

Productivity is widely used as a performance indicator of many industry sectors, e.g., coal industry and construction industry. To calculate this variable researchers often divide some value representing output, i.e., production, by some input value, e.g., total hours or total manpower. In this study, productivity of a mine per annum is defined as the ratio of the Total Production of coal in tons to the Total Employee Hours worked per annum. Thus Productivity is measured in units of “Tons/Employee hour.” The values of Total Production and Total Employee Hours, found in “Address/Employment “ files under “TOTPROD” and “TOTHRS” columns, were added to the main spreadsheet in two new columns called Total Production and Total Worked Hours. Using Equation 3.4, Productivity values were calculated to two decimal places and presented in the main spreadsheet in the column Productivity.

$$Productivity = \frac{Total\ coal\ production}{Total\ employee\ worked\ Hours} \quad \left(\frac{Tons}{Hr}\right) \qquad (3.4)$$

Inspection Days

To keep underground coal mine work places safe and to ensure the compliance of mine operators with safety rules and regulations, MSHA conducts different types of inspections. The most comprehensive of these inspections is the “Regular Safety and Health Inspection,” conducted on a quarterly basis. MSHA inspectors thoroughly inspect all the sections of a mine and the equipment based on their safety rules and regulations (MSHA, 2009). This type of inspection usually takes a long time, and depending on the size and condition of the mine and its compliance with the safety rules and regulations, it can take up to hundreds of hours. In addition to these quarterly inspections, MSHA conducts other announced and unannounced inspections to assure the safety of these mines. For example, if a fatality occurs in a mine, an MSHA inspector is sent to the mine for further investigation; an inspector may also be sent for a non-fatal or non-injury event, depending on the seriousness of the incident. For the present study, total number of inspection days is derived from the quarterly regular safety and health inspections and other inspection types, as listed in Table 3-4.

Inspection Data can be found on MSHA’s Data Retrieval System website. Having a Mine ID and specifying a time period, the "Inspections Summary Report” on that mine during the specified period is shown in a table. The first column of this table lists the inspection’s “Event Number.” The second column of this table contains links providing detailed information on the

"Inspection Activities Conducted." Inspection durations are recorded in units of hours. For this study, instead of taking into account the "Total Hours Spent" for the inspection, only "On-Site Hours Spent" were considered. "Begin Date" and "End Date" of each inspection event is also shown in this table. An inspection event can start near the end of one year and continue to the next year. In such a case, proper adjustments need to be made to have the correct inspection duration recorded for that event. Following is the step-by-step procedure used to get the total inspection hours for each mine per annum.

1. Select a mine from the main spreadsheet and copy its MSHA Mine ID.
2. Go to MSHA's "Mine Data Retrieval System" webpage at http://www.msha.gov/drs/drshome.htm.
3. Paste the Mine ID in the "MSHA Mine ID" box under "Mine Information Search," "Mine Identification Number (ID) Search" and then click on the red "Search" button.
4. Specify the time frame by entering the first and the last day of the desired year in the "Beginning Date" and "Ending Date" boxes, respectively.
5. Select "Inspection" and click on "Get Report" button.
6. A table appears showing detailed information on each inspection event conducted in that particular year. Copy this table, "Inspection Summary Report," in a spreadsheet; change the title of the second and third columns to "On Site Inspection Hours" and "Ratio," respectively.
7. Click on the "Details" button, located next to the event number to get the details on each inspection event.
8. Copy the value listed in front of "On-Site Hours Spent."
9. Paste this value in the "Inspection Summary Report" table next to the corresponding event number, under column "On site Inspection."
10. Repeat this process for the entire events listed in the "Inspection Summary Report" table.

11. After finishing with one year go back to step 4 and repeat this procedure for the next year, until the “On-Site Hours Spent” for the entire events in the 5-year-period are collected. At this point the spreadsheet should contain five “Inspection Summary Report” tables, for years 2003-2007.
12. Check the “Begin Date” and “End Date” of each inspection event in each year to see whether the inspection activities have taken place completely within that year; if yes; put value “1” in the cell next to inspection hour, under the “Ratio” column; on the other hand if an overlap exists, take one of the following actions assigned to the inspection types listed in Table 3- 5.
    a. In case (1), if the “Begin Date” of the inspection is early in the year, check the current year and the previous year and assign that inspection to the year with a missing regular inspection. On the other hand, if the “Begin Date” is very close to the end of the year check the current year and the next year to see which one requires an additional “Regular Inspection” to make the total of four in that year.
    b. In cases (2)-(11), assign the total number of inspection hours to the year in which the inspection began, since almost most of the hours spent on such inspection types occur in the very first days of the inspections.
    c. Cases (12)-(21) are all technical investigations, which continuously carry over the period of inspection. If this is the case, apportion the inspection hours to each year based on the ratio of the number of inspection days in each year.
    d. Finally, in the case (22), ignore the inspection hours spent.
13. Now that the “Ratio” column is complete, multiply the inspection hours by the ratio values and sum up all these values for each year to get the total inspection hours.
14. Repeat steps 1-15 for all the mines.

15. Add a new column, called “Inspection Hours,” to the “Main” spreadsheet and insert the total inspection hours of each mine in each year accordingly.

Table 3- 5: Inspection types and appropriate action regarding inspection hour apportionment

Case no	Inspection Type	Action
1	Regular Inspection	(a)
2	Impoundment Spot Inspection	(b)
3	Mine Emergency Operations	(b)
4	Fatal Accident Investigation	(b)
5	Non-Fatal Accident Investigation	(b)
6	Non-Injury Accident Investigation	(b)
7	Non-Chargeable Accident Investigation	(b)
8	Part 50 Audits	(b)
9	Shaft, Slope or Major Construction Spot Inspection	(b)
10	Verbal Hazard Complaint Inspections	(b)
11	103 (g) Written Notification Hazard Complaint Inspection	(b)
12	Ventilation Technical Investigation	(c)
13	Water/Supply Technical Inspection	(c)
14	Electrical Inspection	(c)
15	Haulage Technical Inspection	(c)
16	Health Technical Investigation	(c)
17	Respiratory Dust Technical Investigation	(c)
18	Roof Control Technical Investigation	(c)
19	Noise Technical Investigation	(c)
20	Other Technical Investigation	(c)
21	103(I) Spot (5-day, 10-day and 15-day) Inspection	(c)
22	Petition for Modification Investigation	(d)

Finally, to calculate the number of Inspection Days for each mine, Total Inspection Hours are divided by five. “Any remainder amount increases the number of inspection days by one” (MSHA, 2007). Equation (3.3) can be used to calculate the number of inspection days in a new column called “Inspection Days.”

$$Inspection\ Days = -INT\left(-\frac{Inspection\ Hours}{5}\right) \tag{3.5}$$

Dollars per Citation

If an MSHA inspector encounters a violation during a mine inspection, he issues a citation to the mine operator. The penalty amount of the citation varies based on the severity of the violation. If a violation is reasonably likely to cause a reasonably serious injury or illness, it is considered a "significant and substantial" (or S&S) violation. MSHA inspectors determine whether a violation is S&S or not. S&S violations have elevated costs for mine operators.

There are different ways to address the severity of violations in a mine. "Dollar per citation" is introduced as an indicator of the severity of the violations in a mine. It is defined as the dollar amount per citation and is calculated by dividing the total amount of "proposed penalty" by the number of citations per annum (Equation 3.6). The higher the average dollars per citation, the more elevated citations the mine has; on the other hand the smaller this value the fewer elevated citations occurred at the mine. If this value is low for a mine in a particular period, it means that severity of the violations is low; on the other hand a high-dollar value for citations indicates that more severe violations, i.e. Significant and Substantial (S&S) types or Orders exist.

$$Dollar\ per\ Citation = \frac{Proposed\ Penalty}{Number\ of\ Citations, Orders, and\ Safeguards} \qquad (3.6)$$

The MSHA Mine ID can be used to obtain the "Proposed Penalty" and the "Number of Citations, Orders, and Safeguards" from the "overview" report section of the MSHA's Data Retrieval System website. These data are listed in the "Citations, Orders, and Safeguards" table under the "overview" report. The first column of this table shows the year, while columns 2-7 illustrate different types of citations, orders and safeguards, the sum of which gives the total number of citations per annum. Moreover, the 8^{th} column contains the amounts of the penalties in dollars, proposed in each year. The total number of citations, orders and safeguards, and the proposed penalties for each year are extracted from these tables and added to the main spreadsheet. Finally, a new column called "Dollar per Citation" is added to the main spreadsheet to accommodate the dollar amounts per citation, calculated using Equation 3.6.

Number of Citations per Inspection Day

Another safety indicator used in this study is the number of citations a mine receives per an inspection day. A larger number of citations per day in a mine indicates more violations of the health and safety regulations in that mine. A mine operation can be interrupted or stopped as a result of these violations, particularly if they are S&S or higher in severity, i.e., Orders. This intuitively results in lowering the productivity of the mine. This variable is calculated in a new column called "citations per inspection day" using Equation 3.7.

$$No.of\ Citations\ per\ Inspection\ Day = \frac{Number\ of\ citations, orders, and\ safeguards}{Number\ of\ inspection\ days} \qquad (3.7)$$

Reliability

Reliability in underground coal mines is defined as the probability of operating a mine without a citation in the period t and is calculated as:

$$R = e^{-\lambda t} \qquad (3.8)$$

Where:

λ is the "Failure Rate", which, in this case, is the average number of citations per inspection day calculated in the previous part.

t is the period of time, in which reliability is being measured, equals to one day for this study.

As the number of citations per inspection day increases, reliability decreases (Figure 3-1). In other words, if a mine receives more citations during an inspection day, it has violated more safety regulations thus it is a less safe and reliable mine, particularly regarding serious violations.

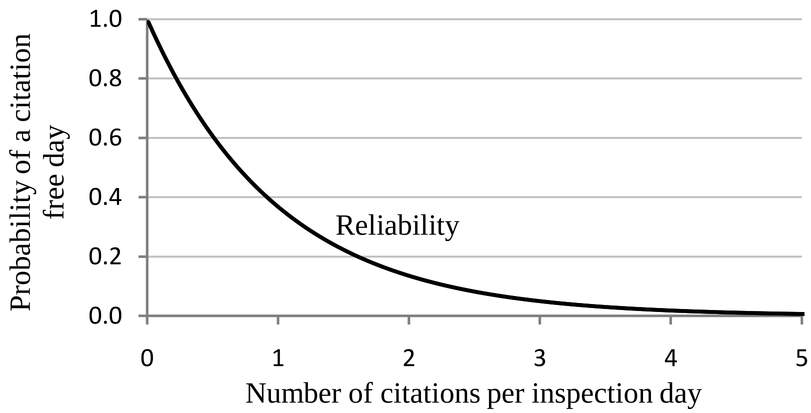


Figure 3- 1: Reliability versus number of citations per inspection day

Non-Fatal Days Lost Injury Rate (NFDL IR)

Non-Fatal Days Lost is another safety measure, which illustrates the number of days lost due to non-fatal accidents/incidents. Losing a work day can impact the productivity. In this study, however, instead of using NFDL, injury rates for 100 full-time employees are used. Assuming each employee works 40 hours a week and 50 weeks a year, he works 2,000 hours per annum. The average total hours worked for 100 full-time employees is thus 200,000 hours. Furthermore, NFDLIRs are calculated in a new column called NFDLIR using Equation 3.9. NFDLs are extracted from “Injuries, Hours Worked, and Production Totals“ tables in the “Overview“ report on MSHA's Data Retrieval System website. Total worked hours already exist in the main spreadsheet from the productivity calculation.

$$NFDL\ IR = \frac{NFDL}{Total\ employee\ worked\ hours} \times 200{,}000 \qquad (3.9)$$

Severity Measure

The severity measure is the sum of the number of Lost workdays and Restricted workdays divided by the total number of hours worked per 200,000 employee-hours, rounded to whole numbers (Hartman, 1992). Lost workdays are the days that an injured miner is away from work. Restricted workdays include the days that an injured miner works at an assigned temporary job, works less than full-time on a permanent job, or works at a permanently assigned job but not being able to perform all the duties normally expected of that position. The number of restricted workdays and the number of lost workdays are listed, under the columns "DAYSREST" and "DAYSLOST" in the "Accident/Injury" files. These data are added to the main spreadsheet in two columns of "Restricted Days" and "Lost Days," while another column accommodates severity measures calculated using Equation 3.10.

$$SM = \frac{Number\ of\ restricted\ days + Number\ of\ lost\ days}{Total\ employee\ worked\ hours} \times 200{,}000 \qquad (3.10)$$

Seam Height

So far all the independent variables discussed are safety measures. Seam height represents the thickness of the coal seam being mined, or the seam thickness plus the thickness of an amount of roof material being taken, which is another structural factor of a coal mine. The thicker the seam height, the faster the coal can be extracted with less idle time for moving equipment; hence more productivity is expected. Seam height is measured in inches and can be found in "Address/Employee" files under the column "SEAMHITE." These values are extracted from the "Address/Employee" files and added to the main spreadsheet under a column called Seam Height.

Year

Mines can have different values for the dependent and independent variables in different years due to changes in internal factors such as the conditions in the mine, or external factors, e.g., safety rules and regulations. A variable can have a significant impact on the productivity of a mine in a specific year. To clearly address this issue, four dummy variables, Y07, Y06, Y05 and Y04, were introduced. Combinations of these four variables represent different years of the period of 2003-2007 (Table 3-6). Moreover, the interaction of the year (dummy variables) and the other independent variables were used to illustrate how a specific variable impacts the productivity of the underground coal mines in a particular year.

Table 3- 6: Dummy variables for year

	Dummy Variable			
Year	Y07	Y06	Y05	Y04
2007	1	0	0	0
2006	0	1	0	0
2005	0	0	1	0
2004	0	0	0	1
2003	0	0	0	0

Longwall (dummy variable)

Longwall mining is a nearly automated form of underground coal mining applied in larger mines. This technique is feasible in relatively flat-lying, thick, and generally uniform coal beds and realizes high extraction rates. A high-powered cutting machine (a shearer) runs along the long face of the coal bed and shears the coal away on a conveyor system to haul it away. The size of a longwall coal panel can vary over time; however, in 2008, it averaged about 900 feet wide, 8,000 feet long and 7 feet high.

Since longwall mines have high total production rates, they are generally more productive than those using other mining techniques, e.g., room and pillar. Longwall mining method is introduced as another structural factor in this study. Another dummy variable, LW, was introduced to address these types of mines. All the longwall mines in the sample are identified from "United States Longwall Mining Statistics 1989-2007" by Weir International, Inc. LW equal to “1” represents a longwall mine, while LW equal to “0” represents mines using other extraction techniques (Table 3- 7).

Table 3- 7: Dummy variables for longwall mines

	Dummy Variable
Mining Technique	LW
Longwall	1
Others	0

Regression Model and Analytical Method

A multi-linear regression model was generated for each mine-size category to illustrate the impact of the safety measures and the structural factors on the productivity of U.S. underground coal mines. Instead of using the complete name of the variables in the models, their abbreviations, listed in Table 3-8, were used. The general form of the models consists of 49 independent variables and looks like Equation 3.1, where the β_is are the coefficients of the regression and ε is the error term. This model contains seven independent variables, seven squared terms of the independent variables 35 interaction terms, 28 of which are interactions of seven independent variables with the dummy variable year, and seven of them are the interactions between Longwall mining and the seven independent variables. Since none of the Small and Very Small mines have longwall mining sections, the latter interaction terms were eliminated. To avoid difficulty in the interpretation of the models, use of other interaction terms was avoided.

Table 3- 8: Summary of the variables used in the regression model

Variable Name	Variable Type	Variable Abbreviation
Productivity	Dependent	PTVT
Inspection Days	Independent	INS_D
Dollars per Citation	Independent	USD_CTN
Number of Citations per Inspection Day	Independent	CTN_D
Reliability	Independent	RL
Non-Fatal Days Lost Injury Rate	Independent	NFDL_IR
Severity Measure	Independent	SM
Seam Height	Independent	SH
Year	Independent Dummy	Y07
	Independent Dummy	Y06
	Independent Dummy	Y05
	Independent Dummy	Y04
Longwall	Independent Dummy	LW

Forward and Backward Stepwise regression method in Minitab 15 are used to find the significant independent variables in the model. After identifying these variables, PTVT is regressed versus these variables. The general form of this multiple regression model looks like Equation 3.11.

$$PTVT = \beta_0 + \beta_1.INS_D + \beta_2.USD_{CTN} + \beta_3.CTN_D + \beta_4.RL + \beta_5.IR + \beta_6.SM + \beta_7.SH + \beta_8.{INS_D}^2 + \beta_9.{USD_{CTN}}^2 + \beta_{10}.{CTN_D}^2 + \beta_{11}.RL^2 + \beta_{12}.IR^2 + \beta_{13}.SM^2 + \beta_{14}.SH^2 + \beta_{15}.Y07.INS_D + \beta_{16}.Y06.INS_D + \beta_{17}.Y05.INS_D + \beta_{18}.Y04.INS_D + \beta_{19}.Y07.USD_{CTN} + \beta_{20}.Y06.USD_{CTN} + \beta_{21}.Y05.USD_{CTN} + \beta_{22}.Y04.USD_{CTN} + \beta_{23}.Y07.CTN_D + \beta_{24}.Y06.CTN_D + \beta_{25}.Y05.CTN_D + \beta_{26}.Y04.CTN_D + \beta_{27}.Y07.RL + \beta_{28}.Y06.RL + \beta_{29}.Y05.RL + \beta_{30}.Y04.RL + \beta_{31}.Y07.IR + \beta_{32}.Y06.IR + \beta_{33}.Y05.IR + \beta_{34}.Y04.IR + \beta_{35}.Y07.SM + \beta_{36}.Y06.SM + \beta_{37}.Y05.SM + \beta_{38}.Y04.SM + \beta_{39}.Y07.SH + \beta_{40}.Y06.SH + \beta_{41}.Y05.SH + \beta_{42}.Y04.SH + \beta_{43}.LW.INS_D + \beta_{44}.LW.USD_{CTN} + \beta_{45}.LW.CTN_D + \beta_{46}.LW.RL + \beta_{47}.LW.IR + \beta_{48}.LW.SM + \beta_{49}.LW.SH + \varepsilon \tag{3-11}$$

Regression Diagnostics

After the model is generated, diagnosis is run on the regression model by examining the residuals. Residual plots are examined to find any curvature, trend, or funnel-shape. If any of these are observed appropriate transformation is done to fix the problem. Standardized residuals are also examined to find residuals equal to or greater than 3.0. Data points with standard residuals of equal to or greater than 3.0 are removed one at a time from the sample and a new regression model is generated with the new sample. This procedure is repeated until a model with high R^2, scattered residual plot, and standardized residuals of less than 3.0 is found. This model is called the “best” regression model.

Chapter 4

RESULTS AND DISCUSSIONS

In this chapter, the result of the analysis on the Mine Size Transition and the Productivity Impact are presented separately. First overall size transitions of 454 mines by size are presented followed by the detailed mine size transitions by year in four one-year periods from 2003 through 2007. The second part of this chapter presents the result for the Productivity Impact analysis, conducted on three mine size categories. A separate multi-linear regression model is generated for each mine size category to identify the significant variables impacting the productivity of the mines in each size category.

Mine Size Transition Analysis

The mine-size transitions of 454 mines of different sizes were studied for the 5-year period of 2003-2007. First the size category of each mine was specified using Table 3-2 and the total amount of coal production of the mines. Then, the number of mines in each size category for each year was obtained using the “Filter” feature of the Microsoft Excel. Table 4-1 illustrates the number of mines in the eight size categories, defined in the previous chapter, from 2003 through 2007. Each row of this table gives information on a particular mine size category. The first column of this table contains the eight size categories defined in the previous chapter. The number of mine(s) in each size category in each year is listed under the corresponding Year, ranging from 2003 through 2007.

Table 4- 1: Mine-size categories based on average number of employees and total production, 2003-2007

Mine Size Category	Mine Size Abbreviation	Year				
		2003	2004	2005	2006	2007
Very Large	VL	30	35	40	42	41
Large	L	48	47	45	49	48
Medium	M	51	59	70	70	57
Small	S	133	146	146	127	104
Small w. no production	S-NP	1	1	2	2	2
Very Small	VS	109	97	88	100	81
Very Small w. no production	VS-NP	33	32	30	27	45
No Employee	NE	49	37	33	37	76
Sum		454	454	454	454	454

Mine-Size Transitions by Size

A quick investigation showed that the number of mines in each category does not remain the same and it changes from one year to another (Table 4-1). In the periods of 2003-2004 and 2004-2005, the number of Very Large mines increased with an increment of 5 mines per year from 30 to 35 and from 35 to 40, respectively; whereas in the period of 2005-2006 only 2 mines joined this size category and finally in the period of 2006-2007 one mine left this category. The minimum (30) and maximum (42) number of Very Large mines occurred in 2003 and 2006, respectively.

The number of Large mines decreased from 48, in 2003, to 47, in 2004, and then to 45, in 2005. In 2006, it reached the maximum number of 49 mines; however, in the next year, 2007, it dropped to 48 mines. Moreover, the minimum (45) and maximum (49) number of Large mines occurred in 2005 and 2006, respectively.

The number of Medium mines increased from 51, in 2003, to 59, in 2004, and then to 70, in 2005. In 2006, it stayed the same as in 2005 and then dropped to 57 mines in 2007. The minimum (51) and maximum (70) number of Medium mines occurred in 2003 and 2005, and 2006, respectively.

Small mines had the largest population among all mine sizes. The number of Small mines jumped from 133 in 2003 to 146 in 2004. In 2005, it stayed the same and later in 2006, it dramatically dropped to 127. In year 2007, things did not get any better for Small mines and again the number of mines in this category dropped significantly to 104. The minimum (104) and maximum (146) number of Small mines occurred in 2007 and 2004, and 2005, respectively. Not many Small mines with no production existed in the sample. In 2003 and 2004, only one mine belonged to this category. This later increased to 2 in 2005 and stayed the same for the next two years of 2006 and 2007.

The Very Small mine size was the second largest size category in the period of 2003-2007. The number of Very Small mines decreased from 109, in 2003, to 97, in 2004, and then to 88 in 2005. In 2006, it increase to 100; however, in 2007, it substantially dropped to 81. The "Very Small with no Production" (VS-NP) mine-size category was a quite crowded category unlike the "Small with no Production." Starting with 33 mines in 2003, the number of Very Small mines with no production decreased slightly to 32, 30, and 27 in 2004, 2005, and 2006, respectively. In 2007, it noticeably boosted to 45. The minimum (27) and maximum (45) number of Very Small mines belonged to 2006 and 2007, respectively. The last mine-size category, No Employee, was also a crowded one. With 49 mines in 2003, the number of mines in this category decreased to 37 and 33 in 2004 and 2005. In 2006, it reached 37 again and significantly jumped to 76 in 2007. The minimum (33) and maximum (76) occurred in 2005 and 2007, respectively.

Comparison of Mine-Size Categories

The mean and standard deviation of the number of mines in each size-category over the 5-year period was calculated. The latter parameter shows how the number of mines in each category is scattered from the average value; in other words it indicates how variable the number of mines in each category is. A larger standard deviation for a size-category means more variation in the number of mines in that size category. Table 4-2 illustrates the means, standard deviations, and variation rankings of the mines. The rankings are based on the standard deviations. The No Employee mines with the STDEV of 17.6 showed the highest variation in the number of mines in a mine-size category during the 5 years. Small and Very Small mines with standard deviations of 17.3 and 10.8, respectively, placed as the second and third most variable mine-size category. On the other hand Large and Very Large mines were found to be the least variable ones with standard deviations of 1.5 and 5.0, respectively (Table 4-2).

Table 4- 2: More statistic on the mines by size category

Mine Size	Mean	STDEV	Variation Ranking
VL	38	5.0	6
L	47	1.5	7
M	61	8.4	4
S	131	17.3	2
S-NP	2	0.5	8
VS	95	10.8	3
VS-NP	33	6.9	5
NE	46	17.6	1

Notice that the standard deviations were calculated using the total number of mines in each category. Mines can still transition between categories, yet leaving the total number of mines in that category unchanged from the previous year. If a mine-size category has less variation in the total number of mines over the five years, it does not necessarily mean that the mine-size category is more stable and less variable. However, these rankings based on the standard deviations can be used as a hypothesis to identify the stability of the mine-size categories. This issue is scrutinized later by tracking individual mines, and the results from that section are compared to the results derived here.

Mine-Size Transition Rates by Year

Since the total number of mines in this study is constant (454) and the same mines are studied in the period of 5 years, it is clear that some mines transitioned from one size category to another; however, Table 4-1 does not provide detailed information on these transitions. To observe these phenomena, the size of the mines in each year were identified and recorded. The size transitions of the mines were tracked during the four one-year periods of 2003-2004, 2004-2005, 2005-2006 and 2006-2007, and recorded in Table B- 1 through Table B- 4 in Appendix B. These tables illustrate the actual numbers of mines transiting from one mine-size category to another. Using Equation (3.1), the transition rates of the mines in each period were calculated. Table 4-3 through Table 4-6 present the transition rates in percentages for the four periods. The diagonal elements of these tables are the most informative pieces of information giving the percentage of the mines staying in the same mine-size category as they were in the preceding year. The closer these value to 100, the less transition occurs, thus the more stable that mine-size category. A diagonal value of 100 for a mine-size category means no mine transitioned to or from that category, meaning the category was perfectly stable, e.g., Very Large category in the period of 2003-2004 (Table 4-3).

Period of 2003-2004

In the first period, 2003-2004, the 30 Very Large mines were perfectly stable, meaning none of them transitioned to the inferior size categories. Out of 48 Large mines, in 2003, 83.3% remained Large in 2004, while 10.4% upgraded to Very Large, and 4.2% and 2.1% downgraded to Medium and Small mines, respectively. Medium mines were not as stable as Large mines with only 74.5% of them remaining Medium in 2004, while 11.8% upgraded to the Large category and 11.8% and 2.0% downgraded to Small and Very Small, respectively.

Small mines were slightly more stable than Medium mines with 76.7% of them remaining Small in 2004, while 10.5% upgraded to Medium and 8.3%, 2.3% and 2.3% downgraded to Very Small, Very Small with no Production and No-Employee categories, respectively. In 2003, there was only one Small mine with no production which remained in that category in the following year. Very Small mines were less stable than Small mines with 64.2% of them remaining Very Small; however, this category showed more improvement than the Small mine-size category with 17.4% and 2.8% upgrading to Small and Medium, respectively. In 2004, 7.3% of the Very Small mines stopped producing coal and 8.3% downgraded to the No-Employee category. Very Small with no Production mines were the second to the last unstable mine-size category in this period with only 48.5% remaining in that category, while 15.2% of them that had no production in 2003 started producing coal in 2004, and 9.1% and 3.0% upgraded to Small and Medium, respectively. The largest transition from this category was to No-Employee mines with the rate of 24.2%. Finally, No-Employee mines were found to be the most unstable ones in this period with only 34.7% of them being faithful to the category, while 10.2% transitioned to the Very Small with no Production category and 20.4% upgraded to Very Small. The largest transition rate from the NE category was to the Small category with the rate of 30.6%, while 2.0% upgraded to each of the Medium and Large categories. These transitions resulted in the increase/decrease in the number of mines in each category as compared to the preceding year. The last row of Table 4-3 shows the Percent Change in the number of mines in each category during the period of 2003-2004. The Percent Change of -24.5% for No-Employee mines means that 24.5% of the NE mines in 2003 transitioned to the other size categories in 2004. On the other hand, the number of Very Large mines increased by 16.7% in 2004 (Table 4-3).

Table 4- 3: Mine-size transition rates for the period of 2003-2004 (in percent)

Year		2004								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	No of Mines
2003	VL	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30
	L	10.4	83.3	4.2	2.1	0.0	0.0	0.0	0.0	48
	M	0.0	11.8	74.5	11.8	0.0	2.0	0.0	0.0	51
	S	0.0	0.0	10.5	76.7	0.0	8.3	2.3	2.3	133
	S-NP	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	1
	VS	0.0	0.0	2.8	17.4	0.0	64.2	7.3	8.3	109
	VS-NP	0.0	0.0	3.0	9.1	0.0	15.2	48.5	24.2	33
	NE	0.0	2.0	2.0	30.6	0.0	20.4	10.2	34.7	49
Percent Change		+16.7	-2.1	+15.7	+9.8	0.0	-11.0	-3.0	-24.5	--

Period of 2004-2005

In the period of 2004-2005, all the 35 Very Large mines again remained in that category, and none of them transitioned to other categories. Large mines, in this period, were almost similar to the previous period with 83% remaining in the category while 10.6% upgraded to Very Large mines. On the other hand, 6.4% downgraded to Medium, which is more than the previous period; however, no Large mines downgraded to a Small mine as in the 2003-2004 period. With 84.7% of the Medium mines remaining in the category, they were more faithful in this period compared to the previous period. This resulted in the reduction in the transition rates to Large and Small mines. From the 59 Medium mines, 8.5% transitioned to Large, while 6.8% of Medium mines joined the Small size category.

Small mines had almost the same transition rates as in the previous period, with 76.7% of them remaining in that category, that is exactly the same as it was in the 2003-2004 period. Out of the 146 Small mines in 2004, 9.6% and 0.7% upgraded to Medium and Large categories, respectively, while 8.9%, 0.7% and 3.4% transitioned to Very Small, Very Small with no Production and No-Employee, respectively. The only mine in the Small with no Production category in 2004 remained in that category in 2005. Very Small mines, like Small mines, were very similar to the previous period, with 62.9% remaining in the category, while 21.6% and 2.1% transitioned to Small and Medium categories. Out of 97 Very Small mines in 2004, 9.3% stopped their production in 2005 and 4.1% downgraded to No-Employee category. Very Small with no Production was the most unstable mine-size category in this period with only half of its mines remaining in the category, while 18.8% of them started producing in 2005, and 18.8% and 3.1% upgraded to Small and Medium categories, respectively. On the other hand, 9.4% of these mines joined the No-Employee category. The second to last unstable mine-size category was No-Employee, with the rate of 56.8% being faithful to the category. This is higher compared to the previous year, meaning the rate of the mines transitioning to higher categories was less and out of 37 mines, 10.8%, 21.6%, 2.7%, and 8.1% transitioned to VS-NP, VS, S-NP, and S categories, respectively. Moreover, the number of mines in the S-NP category doubled in 2005 while the number of Medium and Very Large mines increased by 18.6% and 14.3%, respectively. There was no change in the number of Small mines during this period, while the number of Large, Very Small with no Production, Very Small, and No-Employee mines reduced by 4.3%, 6.3%, 9.3%, and 10.8%, respectively. Overall the period of 2004-2005 was very similar to 2003-2004 with the exception of some significant differences in the No-Employee category that were discussed above (Table 4-4).

Table 4- 4: Mine-size transition rates for the period of 2004-2005 (in percent)

Year		2005								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	No of Mines
2004	VL	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	35
	L	10.6	83.0	6.4	0.0	0.0	0.0	0.0	0.0	47
	M	0.0	8.5	84.7	6.8	0.0	0.0	0.0	0.0	59
	S	0.0	0.7	9.6	76.7	0.0	8.9	0.7	3.4	146
	S-NP	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	1
	VS	0.0	0.0	2.1	21.6	0.0	62.9	9.3	4.1	97
	VS-NP	0.0	0.0	3.1	18.8	0.0	18.8	50.0	9.4	32
	NE	0.0	0.0	0.0	8.1	2.7	21.6	10.8	56.8	37
Percent Change		+14.3	-4.3	+18.6	0.0	+100.0	-9.3	-6.3	-10.8	---

Period of 2005-2006

Table 4-5 illustrates the mine-size transition rates in the period of 2005-2006. In this period, 97.5% of the Very Large mines remained in that category, while the remaining 2.5% downgraded to Large mines. Large mines, in this period, showed slightly more faithfulness compared to the previous periods with 86.7% of them remaining in the category. The rest of the Large mines transitioned to Very Large and Medium size-categories each with the transition rate of 6.7%. Medium mines showed more improvement with 12.9% upgrading to Large mines, and only 7.1% downgrading to Small, while 64.3% of them remained Medium.

The ratio of the Small mines remaining in their initial category was 76%, while 8.2% upgraded to Medium size and 11.0%, 1.4%, and 3.4% downgraded to VS, VS-NP, and NE, respectively. In 2005, there were only 2 Small with no Production mines, one of which started to produce coal in 2006 while the other one remained in the S-NP category. Very Small mines were quite different in this period compared to the previous periods. They became more faithful to their initial size-category with the rate of 85.2% remaining in the category; while 5.7% of them upgraded to Small size by the rate of 5.7%, which is much less than in the previous periods. Moreover, 3.4% of the Very Small mines in 2005 stopped producing coal in 2006, while 5.7% of them joined the No-Employee category. The rate of VS-NP mines remaining in the category increased to 63.3%, which is higher compared to the previous periods. Ten percent of the Very Small mines started producing coal in 2006, while 3.3% had an increase in the number of employees that resulted in an upgrade to the Small size-category, yet with no production. Moreover, 6.7% of the VS-NP mines joined the Small size-category. The remaining 16.7% of VS-NP mines lost their employees and joined the No-Employee mine-size category. The No-Employee mines remained faithful with the rate of 66.7%, which is higher compared to the first and second periods; however, the transition rate to VS-NP and VS reduced to 6.1% and 18.2%, respectively; while 9.1% of NE mines transitioned to Small mines. In sum, the number of Very Large mines increased in 2006 by 5.0%, which is a lower rate compared to the previous periods; on the other hand, the number of Large mines increased by 8.9%, the first positive change. The number of Medium and Small with no Production mines remained unchanged. The number of Small and Very Small with no Production mines decreased by 13.0% and 10.0%, respectively. Finally, despite the previous periods, the number of Very Small and No-Employee mines increased by 13.6% and 12.1%, respectively (Table 4-5).

Table 4- 5: Mine-size transition rates for the period of 2005-2006 (in percent)

Year		2006								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	No of Mines
2005	VL	97.5	2.5	0.0	0.0	0.0	0.0	0.0	0.0	40
	L	6.7	86.7	6.7	0.0	0.0	0.0	0.0	0.0	45
	M	0.0	12.9	78.6	7.1	0.0	0.0	1.4	0.0	70
	S	0.0	0.0	8.2	76.0	0.0	11.0	1.4	3.4	146
	S-NP	0.0	0.0	0.0	50.0	50.0	0.0	0.0	0.0	2
	VS	0.0	0.0	0.0	5.7	0.0	85.2	3.4	5.7	88
	VS-NP	0.0	0.0	0.0	6.7	3.3	10.0	63.3	16.7	30
	NE	0.0	0.0	0.0	9.1	0.0	18.2	6.1	66.7	33
Percent Change		+5.0	+8.9	0.0	-13.0	0.0	+13.6	-10.0	+12.1	---

Period of 2006-2007

Table 4-6 illustrates the mine-size transition rates for the last period of this study, 2006-2007, when many unusual transitions happened. Very Large mines were less faithful in this period compared to the previous periods with 92.9% remaining in that category, while 7.1% of them downgraded to Large mines. Large mines also remained in their category with a lower rate of 81.6%, which is less than in the previous periods. The rate of the Large mines transitioning to Very Large and Medium size both dropped to 4.1%, which is also smaller compared to the previous periods; moreover, 2% of the Large mines downgraded to Small mines, which is similar to the first period. The latter accounts for only one mine that was abandoned in early 2007. Finally, some unusual transitions that had not happened in the previous periods occurred when some Large mines transitioned to Very Small with no Production and No-Employee categories with the rate of 6.1% and 2.0%, respectively. The former accounts for 3 mines, two of which were “temporarily closed” due to “roof fall” and the other one was abandoned, while the latter accounts for one mine that also happened to be “temporarily closed” due to “roof fall.” The Medium mines were also less faithful to their size-category with the rate of 64.3%, which is the least compared to the previous periods. The transition rate of upgrading from Medium to Large also decreased to 7.1%, the lowest value among all the four periods. All in all, more Medium mines downgraded to inferior categories, i.e., Small, Very Small, and Very Small with no Production mines, with the rate of 17.1%, 2.9%, and 2.9%, respectively; while 5.7% of them joined the No-Employee mine-size category.

Small mines also became less faithful with 61.4% of them remaining in the category, which is more than 15% lower compared to the previous periods. The transition rate to Medium size reduced to 7.1%, while the rest of the mines in this category downgraded to Very Small, Very Small with no Production, and No-Employee with the rate of 2.9%, 2.9%, and 5.7%, respectively. S-NP category with only 2 mines turned out to be perfectly faithful with both of the mines remaining in the category. Very Small mines, however, with only 57% of them remaining in that category, were the least stable mine-size category in this period. The transition rates from VS to S mines improved to 11.0%, while 1.0% of the VS mines also upgraded to Medium size. On the other hand, 9.0% of the 100 Very Small mines in this period stopped producing coal, while surprisingly the remaining 22.0% transitioned to the NE category. The VS-NP mines remained in that category with the rate of 85.2%, which is the highest among the four periods, meaning once a Very Small mine stopped its production, it remained non-producing with the likelihood of 85.2%. In fact, only 3.7% of the VS mines resumed their productions in 2007. An upgrade of 3.7% to Small size was observed; whereas the remaining 7.4% transitioned to the No-Employee category. Another unexpected phenomenon was the very high rate of 86.5% for the No-Employee mines remaining in that category. Since mostly Small and Very Small mines transition to this category, this high rate shows that, in this period, once a Small or Very Small mine transitioned to this category the chance of getting back to business and retrieving the production of coal was less than 15%. That being said, some mines still transitioned to VS-NP, VS, and S with the rate of 5.4%, 5.4%, and 2.7%, respectively; however, all of these upgrade transition rates were smaller compared to the previous periods. Overall, in this period, Very Large, Large, Medium, Small and Very Small mines had the lowest rates for remaining faithful to their own mine-size categories; whereas Small with no Production, Very Small with no Production, and No-Employee mines were the most faithful to their categories compared to the previous periods. Also the overall tendency of the mines transitioning to inferior mine-size categories, in this period, were higher compared to the previous periods. This phenomenon was discovered by finding larger transition rate values for the above diagonal elements in Table 4-6 in comparison to Tables 4-3 through 4-5. Another unique phenomenon in this period was the significant increase in the number of VS-NP and NE mines by 66.7% and 105.4%, respectively.

This again shows how dramatically mines were impacted in this period. These dramatic changes are important facts, needing to be scrutinized to find the reasons behind them. This is beyond the scope of this study and is left to the interested readers for further investigations.

Table 4- 6: Mine-size transition rates for the period of 2006-2007 (in percent)

Year		2007								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	No of Mines
2006	VL	92.9	7.1	0.0	0.0	0.0	0.0	0.0	0.0	42
	L	4.1	81.6	4.1	2.0	0.0	0.0	6.1	2.0	49
	M	0.0	7.1	64.3	17.1	0.0	2.9	2.9	5.7	70
	S	0.0	0.0	7.1	61.4	0.0	15.0	4.7	11.8	127
	S-NP	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	2
	VS	0.0	0.0	1.0	11.0	0.0	57.0	9.0	22.0	100
	VS-NP	0.0	0.0	0.0	3.7	0.0	3.7	85.2	7.4	27
	NE	0.0	0.0	0.0	2.7	0.0	5.4	5.4	86.5	37
Percent Change		-2.4	-2.0	-18.6	-18.1	0.0	-19.0	+66.7	+105.4	---

Summary of one-year transition rate for the period of 2003-2007

Table B-5, average one-year mine-size transition matrix for the period of 2003-2007, located in Appendix B of this document was used to generate Table 4-7 which illustrates one-year average transition rates in the period of 2003-2007. This table gives an overall image of how mines of different sizes tended to change their categories from one year to the next over the period of 2003-2007. Among all the mine-size categories, Very Large followed by Large were found to be the most faithful ones with the rate of 97.3% and 83.6%, respectively, remaining in their initial mine-size categories. After the S-NP category with average number of two mines remaining in that category with the rate of 83.3%, Medium turned out to be the next most faithful mine-size category with 75.2% of its mines remaining in the Medium category. Moreover, Small and Very Small mines ranked the 5^{th} and 6^{th}, respectively, with 73.0% and 66.8% rates of remaining faithful. Finally, the VS-NP category ranked 7^{th}, or second to the last, with a 60.7% rate of mine-size category faithfulness, followed by the least faithful category, the NE, having the faithfulness rate of 59.0%. Furthermore, the VS-NP, VS, and S mines were found to have the highest transition rates to the NE category. This could be due to the quick depletion of these mines due to having smaller amounts of coal reserves or more violations, hazardous events, incidents, or accidents which resulted in shutting down these mines. More in-depth investigations in the history of production, violations, accidents, and citations are required to accurately judge these higher rates. This is beyond the scope of this project and is left to the interested readers for further investigation.

Table 4- 7: Average one-year mine-size transition rates for the period of 2003-2007

		k+1								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	No of Mines
k	VL	97.3	2.7	0.0	0.0	0.0	0.0	0.0	0.0	37
	L	7.9	83.6	5.3	1.1	0.0	0.0	1.6	0.5	47
	M	0.0	10.0	75.2	10.8	0.0	1.2	1.2	1.6	62
	S	0.0	0.2	8.9	73.0	0.0	10.7	2.2	5.1	138
	S-NP	0.0	0.0	0.0	16.7	83.3	0.0	0.0	0.0	2
	VS	0.0	0.0	1.5	14.2	0.0	66.8	7.4	10.2	99
	VS-NP	0.0	0.0	1.6	9.8	0.8	12.3	60.7	14.8	30
	NE	0.0	0.6	0.6	14.1	0.6	16.7	8.3	59.0	39
Percent Change		7.5	0.0	2.4	-5.3	16.7	-7.1	9.8	17.3	---

Regression Analysis on Productivity Impact

Introduction

In this section the impact of safety measures and structural factors on the productivity of Consistently Productive underground mines was studied. The objective was to identify important factors impacting the productivity of the mines. This was done by regressing the variables, introduced in Chapter 3, against the productivity of these mines. Due to the differences in the nature of mines of different sizes, separate models were generated for each mine-size category. The sample used in this part of the study was a reduced version of what was used in the Mine Size Transition study.

Generating the Sample

The first step in this analysis was to generate an appropriate sample by filtering the database using the criteria discussed in Chapter 3. Mines were initially categorized in 5 categories of Very Large, Large, Medium, Small, and Very Small. The "consistently productive" mines in the period of 2003-2007 were identified and the rest were eliminated from the sample. Table 4-8 shows the distribution of these mines in the five mine-size categories. Similar to what was observed in the Mine Size Transition section previously, not all the mines were faithful to their mine-size categories over time and they were switching from one category to another. Using MSHA's mine IDs "Faithful" mines were identified. Adding this criterion narrowed down the total number of consistently productive mines from 247 to 125. The latter sample consists of 28 Very Large, 22 Large, 15 Medium, 42 Small, and 18 Very Small mines (Figure 4-1).

Table 4- 8: Distribution of consistently productive underground coal mines, 2003-2007

Mine Size	Year				
	2003	2004	2005	2006	2007
Very Large	29	34	39	42	41
Large	42	39	38	40	43
Medium	42	49	54	55	45
Small	91	94	85	80	79
Very Small	43	31	31	30	39
Sum	247	247	247	247	247

Since it was intended to have at least 30 mines in each mine-size category, this sample needed to be modified to get more mines in each category. Due to the similarities between Very Large and Large mines, and Small and Very Small mines especially in the adopted technology, environment and culture, these four categories were reduced to two categories of Very Large/Large and Small/Very Small. Medium mines stayed intact; however, their upper bound of the Average Number of Employees changed from 100 to 150. Table 4-9 illustrates the boundaries of the Average Number of Employees for each the new mine-size categories.

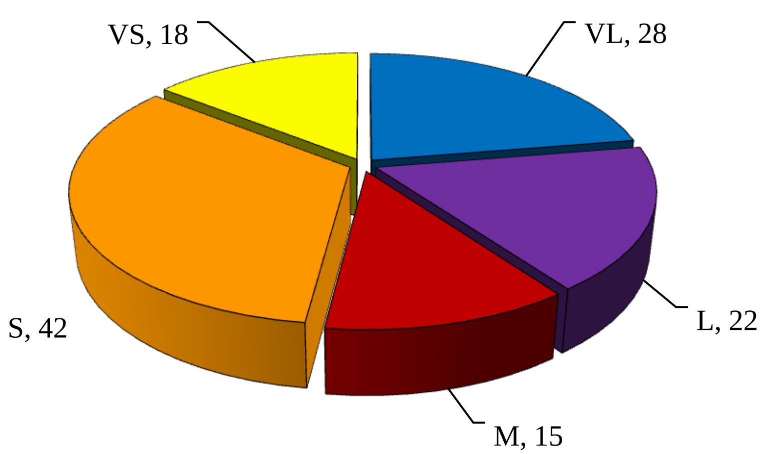


Figure 4- 1: Mine-size allocation of the consistently productive and faithful mines

Table 4- 9: Modified mine-size categories based on the average number of employees

Mine Size	Abbreviation	Average Number of Employees (EMP)
Very Large/Large	VL/L	EMP > 150
Medium	M	50 < EMP ≤ 150
Small/Very Small	S/VS	0 < EMP ≤ 50

The 247 "consistently productive" mines were categorized using the modified mine-size categories (Table-4-9). The number of mines in each of the three new categories, Very Large/Large, Medium, and Small/Very Small, increased compared to the initial categories; however, changes in the number of mines in each category from one year to the next clearly showed that there were still a number of unfaithful mines in each category (Table 4-10). Further investigation revealed that, with the new categorization, out of 247 "consistently productive" mines 178 were "faithful" during the period of 2003-2007. These 178 mines consist of 48 Very Large/Large, 31 Medium, and 99 Small/Very Small "consistently productive" and "faithful" mines (Figure 4-2). Thirty-one mines were randomly selected from each category to study productivity impact over the 5-year period of 2003-2007.

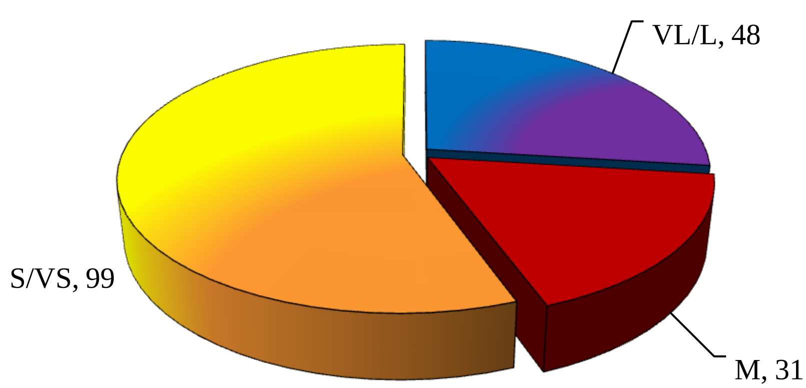


Figure 4- 2: Modified mine-size allocation of the consistently productive and faithful mines

Table 4- 10: Distribution of consistently productive underground coal mines, 2003-2007

	Year				
Mine Size	2003	2004	2005	2006	2007
Very Large/Large	57	56	57	63	61
Medium	56	66	74	74	68
Small/Very Small	134	125	116	110	118
Sum	247	247	247	247	247

Splitting the Sample into Three Categories

Due to the differences in the nature, culture, equipment, and technology of mines of different sizes, each mine-size category was studied separately. Each mine-size category consisted of data for 31 mines over the 5-year period of 2003-2007. This resulted in 155 data points for each size category. All the required information to calculate the variables of the regression models were acquired and stored in the main spreadsheet. Using these variables, a separate multi-linear regression model was generated for each mine-size category.

Very Large/Large Mines

Using both Forward and Backward Stepwise techniques in Minitab 15, a regression model was developed to find the significant variables impacting the productivity of Very Large/Large mines. The sample included 155 data points, which were the data for 31 Very Large/Large mines over the 5-year period of 2003-2007. Productivity was regressed against 49 independent variables, out of which only six were found to be significant. This part discusses the result of this regression model.

Discussions of Correlation Among Variables

The Pearson correlations of the variables were calculated using the "Correlation" feature of the Minitab software under "Basic Statistics." Among the six selected independent variables, SH^2 and SH were the most correlated ones with the response PTVT with the correlation coefficients of 0.754 and 0.727, respectively (Table 4-11). The SH^2 with the highest correlation coefficient with the PTVT was expected to be the first variable to enter into the model. No further predictions could be made regarding the order of the other variables entering into the model.

Table 4- 11: Correlation matrix for Very Large/Large mines

Variable	PTVT	SH	SH^2	SH.Y07	USD_CTN .Y06	SM
SH	0.727					
SH^2	0.754	0.993				
SH.Y07	-0.087	0.158	0.152			
USD_CTN.Y06	-0.185	-0.078	-0.073	-0.196		
SM	-0.164	-0.317	-0.337	-0.157	-0.016	
RL.Y07	-0.101	0.061	0.053	0.939	-0.111	-0.196

Estimated "Best" Regression Function

Forward and Backward Stepwise regressions were run and the 6 variables of Seam Height squared (SH^2), Seam Height of the VL/L mines in Year 2007 (SH.Y07), Seam Height (SH), the amount of Dollars per Citation in Year 2006 (USD_CTN.Y06), the Reliability of the VL/L mines in Year 2007 (RL.Y07), and Severity Measure (SM) were significant. Table 4-12, derived from the Minitab output, illustrates the coefficients, and standard errors of the coefficients of these significant variables.

Table 4- 12: Regression coefficient statistics for Very Large/Large mines

Predictor	Coefficient	Standard Error Coefficient	T-Value	P-Value	VIF
Constant	10.2120	2.0580	4.96	0.000	---
SH	-0.2360	0.0521	-4.53	0.000	73.883
SH^2	0.0021	0.0003	6.66	0.000	75.048
SH.Y07	-0.0555	0.0115	-4.84	0.000	9.310
USD_CTN.Y06	-0.0021	0.0006	-3.70	0.000	1.049
SM	0.0014	0.0007	2.03	0.045	1.179
RL.Y07	6.7330	1.9480	3.46	0.001	9.122

Equation (4-1) shows the regression function for the Very Large/Large mines. Since some of the variables cannot be zero, e.g., SH, the constant term alone does not have any meaning and cannot be interpreted as the predicted mean for productivity of the mines. Moreover it is a very high value for the average of the Productivity of the VL/L mines in 2003. This high value, however, is reduced by the negative impact of the combination of SH and SH^2. The value of SH ranges from 48” to 120” for VL/L mines. The combined term “-0.2360.SH+0.0021.SH^2“ is an increasing function of SH when SH is greater than 57 inches. Moreover for SH greater than and equal to 114 inches this sum becomes positive. This means that productivity of the VL/L mines decreases when the SH increases from 48 to 57, and then it increases when the SH is greater than 57 inches (Figure 4-3). Overall, the thickness of the coal seam had a positive impact in the productivity of these mines. With every $1 increase in the amount of Dollar per Citation of Very Large/Large mines in 2006, the mean of the productivity decreased by 0.0021 while keeping the other variables constant. Moreover, the mean of productivity of the VL/L mines increased by 0.0014, with every unit increase in the Severity Measure of these mines, assuming other variables are held constant. This can be due to the fact that in the VL/L mines the focus was on high production, which increases the pace of the activities, hence increasing the risk of exposure to accidents, thus the Severity Measure increased. Finally, the equation suggests that with every one unit increase in the Reliability of the VL/L mines in 2007, their Productivity increased by 6.7330.

$$PTVT = 10.20 - 0.2360.SH + 0.0021.SH^2 - 0.0555.SH.Y07 - 0.0021.USD\_CTN.Y06 + 0.0014.SM + 6.7330.RL.Y07 \quad (4\text{-}1)$$

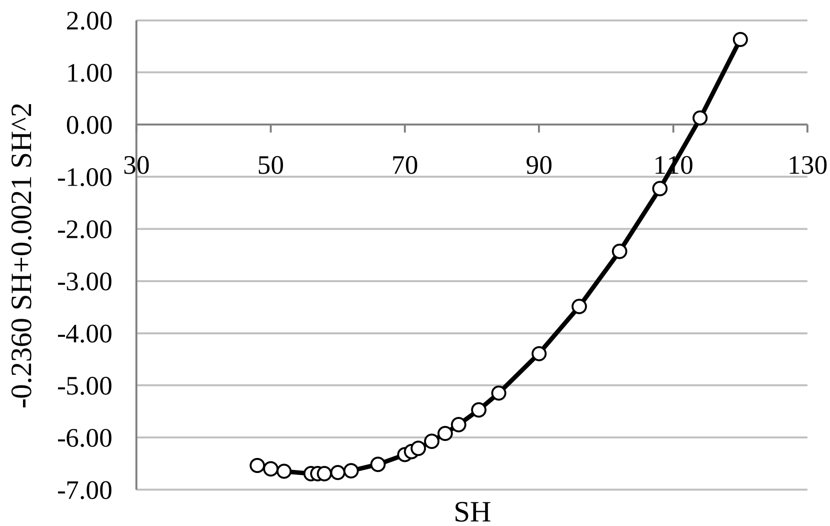


Figure 4- 3: Impact of the changes in the seam height on the productivity of Very Large/Large mines

Analysis of Variance

Table 4-13 illustrates the results for the Analysis of Variance (ANOVA) of the regression model. This table was derived from the Minitab result, in Appendix C. The regression model had degrees of freedom of 154. The SSR was calculated as 834.47, while the SSTO was 1181.75. The MSE was also calculated as 2.35. The P-value of zero suggested the regression model with the six variables was significant. Sequential sum square values, shown in Table 4-14, were also derived from the Minitab results, located in Appendix C. “SH“ had the largest contribution in the model with a sequential sum of squares of 598.10, which accounts for 50.61% of the SSTO. The rest of the independent variables each contributed less than 10% (Table 4-14).

Table 4- 13: Analysis of variance for Very Large/Large mines 2003-2007

Source	DF	SS	MS	F	P
Regression	6	834.47	139.08	59.27	0.000
Residual Error	148	347.28	2.35		
Total	154	1181.75			

Table 4- 14: Sequential sum of squares of the regression model for Very Large/ Large mines 2003-2007

Source	DF	Sequential SS	% of SSTO
SH	1	598.10	50.61%
SH^2	1	103.68	8.77%
SH.Y07	1	53.23	4.50%
USD_CTN.Y06	1	32.09	2.72%
SM	1	31.94	2.70%
RL.Y07	1	15.43	1.31%

Coefficient of Multiple Determination (R^2)

The R^2 value of this regression model was 70.61%, meaning 70.61% of the variations were explained by the model in Equation 4-1. Furthermore, the adjusted R^2 turned out to be 69.42%, which is, as expected, obviously less than R^2.

Examinations of the Residual Plots

The upper-right graph of Figure 4-4 depicts the standardized residuals versus fitted values of the response. This plot does not show any curvature or funnel shape while all the residuals are smaller than 3.0. The upper-left graph, Normal Probability Plot, shows how well the residuals are laying on a straight line, which proves the normality of the residuals. The lower-left graph, Histogram, also shows a nice bell shape of the distribution of the residuals, which again represents the normality of the residuals. The lower-right graph, "Versus Order," depicts how well the residuals are spread up and down. The standard deviation of the residuals was calculated as S=1.53183.

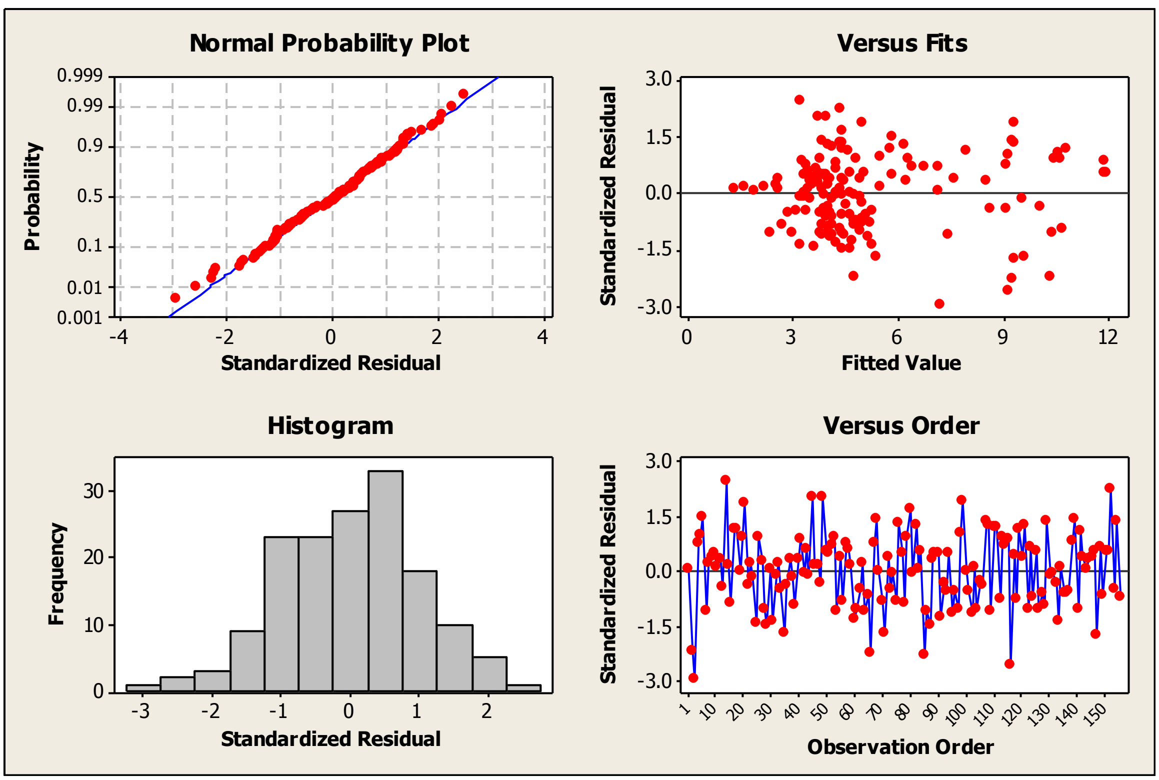


Figure 4- 4: Residual plots for Very Large/Large mines

Consideration of Multicollinearity

Some multicollinearity was observed between independent variables in this model. The independent variables SH and SH^2 were highly correlated with each other with the correlation coefficient of 0.993, while RL.Y07 and SH.Y07 also turned out to be highly correlated with the correlation coefficient of 0.939 (Table 4-11). Checking the Variance Inflation Factors (VIF) of the variables and using the "rule of thumb" of VIF>10, only SH and SH^2 were found to be highly correlated with the VIFs of 73.883 and 75.048 (Table 4-11). However, since the objective of this study was not to predict the Productivity using the independent variables, it is not necessary to fix the multicollinearity.

Conclusion

The productivities of the 31 randomly selected Very Large/Large mines were regressed against 49 variables indicating the safety measures and structural factors of these mines. After the analysis six variables turned out to be significant, i.e., SH^2, SH, SM, SH.Y07, RL.Y07, and USD_CTN.Y06. Seam Height, overall, had a positive impact in the productivity of the VL/L mines, meaning when Seam Height increased the productivity of the mines increased. There was however an exception for the Year 2007, when the productivity decreased slightly with an increase in the Seam Height. This impact was expressed in the model with the negative coefficient of the variable SH.Y07. Moreover, the Severity Measure surprisingly had a positive impact on the productivity. Although this positive impact was very minor, it still brings up a flag. The question is, did the exposure to risk increase due to the focus on the production? The amount of Dollars per Citation, which is representative of the significance of violations, had a negative impact on the productivity of these mines in 2006. This can reflect the impact of the increases in the citations in the latter part of 2006. Finally, in 2007, Reliability of the mines had a positive impact on the productivity of these mines. Thus mines with lower probability for incurring citations had higher productivity.

Medium Mines

The objective was to find the model that “best” explains the impact of safety measures and structural factors on the productivity of Medium-size mines. First Forward and Backward Stepwise regression methods were used in Minitab 15 to find the significant variables out of the 49 variables explained in Chapter 3, then the PTVT was regressed on the selected variables. The number of data points was 155, which accounts for 31 mines in the five years of 2003-2007. In the process of finding the “best” model for Medium mines, after finding the significant variables, standard residuals were examined to identify any data point(s) with a standard residual of greater than 3.0. The “best” model was selected after 5 iterations, during which Medium mines number 134, 103, 72, and 54 were removed from the sample because they had standard residuals more than 3.0.

Discussions of Correlation among Variables

The Pearson correlations of the variables were calculated using the “Correlation” feature of the Minitab software under “Basic Statistics.” Having the highest correlation coefficient of 0.697 with the PTVT, the independent variable RL.LW entered the model first (Table 4-15). Several independent variables were also found to be correlated with each other. This issue is discussed in the section on multicollinearity (p.79).

Table 4- 15: Correlation matrix for Medium mines

	PTVT	RL.LW	RL^2	CTN_D .LW	NFDL_IR .LW	SH^2	USD_CTN .LW	SH	SM.Y06	SH.LW	SH.Y07
RL.LW	0.697										
RL^2	0.277	-0.036									
CTN_D.LW	0.536	0.803	-0.125								
NFDL_IR.LW	0.695	0.856	-0.078	0.899							
SH^2	0.204	-0.037	0.156	-0.010	-0.025						
USD_CTN.LW	0.282	0.531	-0.124	0.859	0.699	-0.015					
SH	0.192	-0.015	0.135	0.016	-0.002	0.990	0.005				
SM.Y06	-0.144	-0.071	-0.087	-0.067	-0.065	0.052	-0.052	0.055			
SH.LW	0.609	0.930	-0.090	0.946	0.900	-0.009	0.700	0.019	-0.070		
SH.Y07	-0.196	-0.040	-0.181	0.069	0.021	0.144	0.168	0.148	-0.166	0.004	
RL.Y05	0.002	0.028	0.252	-0.014	-0.029	0.063	-0.044	0.062	-0.160	0.008	-0.223

Estimated "Best" Regression Function

The "best" model was obtained after five iterations. In each iteration, using Forward and Backward Stepwise regression, PTVT was regressed against 49 independent variables to find the significant variables. After examining the residuals and eliminating data points with residuals greater than 3, the "best" model was found to have 11 significant variables, i.e., Reliability of Longwall mining (RL.LW), Reliability squared (RL^2), amount of Dollars per Citation for Longwall mines (USD_CTN.LW), the NFDL_IR of Longwall mines (NFDL_IR.LW), Seam Height squared (SH^2), the Number of Citations per Inspection Day of the Longwall mines (CTN_D.LW), Seam Height (SH), Severity Measure in Year 2006 (SM.Y06), the Seam Height of Longwall mines (SH.LW), Seam Height in Year 2007 (SH.Y07), and the Reliability in Year 2005 (RL.Y05). Table 4-16 illustrates these variables with their coefficients and other statistics. The detailed Minitab regression output for all the iterations for Medium mines can be found in Appendix C-2.

Table 4- 16: Regression coefficient statistics for Medium mines

Predictor	Coefficient	Standard Error Coefficient	T-Value	P-Value	VIF
Constant	4.6820	0.7092	6.60	0.000	---
RL.LW	16.6410	2.2510	7.39	0.000	18.39
RL^2	2.0497	0.3355	6.11	0.000	1.18
USD_CTN.LW	-0.0041	0.0007	-5.64	0.000	7.97
NFDL_IR.LW	0.4662	0.0948	4.92	0.000	8.32
SH^2	0.0006	0.0001	4.43	0.000	50.87
CTN_D.LW	10.9220	2.1600	5.06	0.000	66.56
SH	-0.0707	0.0205	-3.44	0.001	50.80
SM.Y06	-0.0013	0.0003	-3.75	0.000	1.11
SH.LW	-0.1912	0.0308	-6.20	0.000	69.74
SH.Y07	-0.0120	0.0023	-5.23	0.000	1.24
RL.Y05	-1.0799	0.2797	-3.86	0.000	1.17

Equation 4-2 shows the regression model for the Medium-size mines for the period of 2003-2007. The constant term, 4.6820, is very close to the average productivity of the Medium mines in 2003 (4.40); however, since there are some variables in the model that cannot be equal to zero, this constant term is not meaningful. Similar to Very Large/Large mines, the productivity of the Medium mines were also impacted by SH and SH^2. The Seam Height of the sample Medium mines ranges from 33 to 108 inches. The combined term -0.0707 SH+0.0006 SH^2 is a decreasing and increasing function of SH in this range. Moreover, when the Seam Height increases from 33 to 59 inches the productivity decreases; on the other hand when the Seam Height increases from 60 to 108 inches the productivity increases (Figure 4-5). In 2007, the productivity of the mines decreases by 0.0120 with every one inch increase in the Seam Height, keeping the other variables constant. The positive coefficient of the RL^2 suggests that the productivity of the mines increases when the Reliability increases. One unit increase in the Severity Measure of the mines in 2006, decreases the productivity of the mines by 0.0013, when the other variables are held constant. Also in 2005, every one percent increase in the Reliability of the mines decreases the productivity of the mines by 0.010799.

There were five Longwall Medium mines, one of which had a longwall operation in the 5 years, meaning that overall there were nine observations that had a longwall operation. While holding the other variables constant, five interaction terms containing the dummy variable LW explained the impact on the productivity for Longwall Medium mines as follows: every 1 percent increase in the Reliability increases productivity by 0.166; every $1 increase in the amount of dollars per citation decreases productivity by 0.0041; one unit increase in the non-fatal days lost incident rate increases productivity by 0.4662; every one unit increase in the number of citations per day increases productivity by 10.922; and every one inch increase in Seam Height decreases productivity by 0.1912.

$$PTVT = 4.6820 + 16.6410.RL.LW + 2.0497.RL\string^2 - 0.0041.USD\_CTN.LW + 0.4662.IR.LW + 0.0006.SH\string^2 + 10.9220.CTN\_D.LW - 0.0707.SH - 0.0013.SM.Y06 - 0.1912.SH.LW - 0.0120.SH.Y07 - 1.0799.RL.Y05 \quad (4\text{-}2)$$

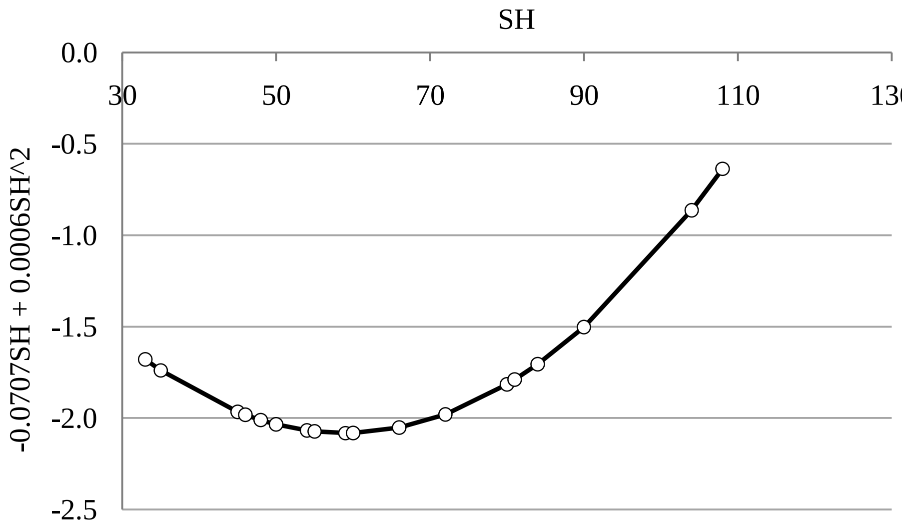


Figure 4- 5: Impact of the changes in the seam height on the productivity of Medium mines

Analysis of Variance

Regressing the 11 significant variables against PTVT, the model was found to be significant with a P-value equal to zero and a large F-value of 59.950. This model had 150 degrees of freedom, due to the removal of three influential data points. The SSR and SSE were calculated as 328.305 and 69.204, summing up to 397.509 for the SSTO value. The estimation of the model's variance was calculated as 0.498, which is a very small value for variance (Table 4-17). The RL.LW with the sequential sum of squares of 19.2951 had the largest share of 58.77% in explaining the variations in the model. The second largest contributor was RL^2 with a sequential

SS of 36.393, which explains 11.09% of the variations in the model. The rest of the variables, listed in Table 4-18, had a share of less than 10%.

Table 4- 17: Analysis of variance for Medium mines 2003-2007

Source	DF	SS	MS	F	P
Regression	11	328.305	29.846	59.950	0.000
Residual Error	139	69.204	0.498		
Total	150	397.509			

Coefficient of Multiple Determination (R^2)

The R^2 value of the regression model for Medium mines was calculated as 82.59%, meaning the 11 independent variables in the model explain 82.59% of the variation in the model. Moreover, to consider the effect of the number of variables in the model, the adjusted R^2 was calculated to be 81.21%, which is slightly smaller than R^2.

Table 4- 18: Sequential sum of squares of the regression model for Very Large/ Large mines 2003-2007

Source	DF	Sequential SS	% of SSTO
RL.LW	1	192.951	58.77%
RL^2	1	36.393	11.09%
USD_CTN.LW	1	20.312	6.19%
NFDL_IR.LW	1	19.143	5.83%
SH^2	1	13.252	4.04%
CTN_D.LW	1	11.737	3.58%
SH	1	9.189	2.80%
SM.Y06	1	6.994	2.13%
SH.LW	1	6.694	2.04%
SH.Y07	1	6.527	1.99%
RL.Y05	1	5.114	1.56%

Examinations of the Residual Plots

The upper-right graph of Figure 4-6 depicts the standardized residuals versus the fitted values of the response. This plot does not show any curvature, trend, or funnel shape; however, the data suggests that Medium mines no 138, 84, 114, and 144 with fitted values of 8.67, 10.68, 11.29, and 12.41, respectively, have large leverages. Moreover, all the data points have the standardized residual of less than 3.0. The upper-left graph, the Normal Probability Plot, shows an almost linear trend, indicating the normality of the residuals. The lower-left graph, Histogram, also shows a bell-shape histogram of the distribution of the residuals, which again represents the normality of the residuals. The lower-right graph, "Versus Order," depicts how well the residuals are spread up and down. Finally the standard deviation of the residuals was calculated to be S=0.705597.

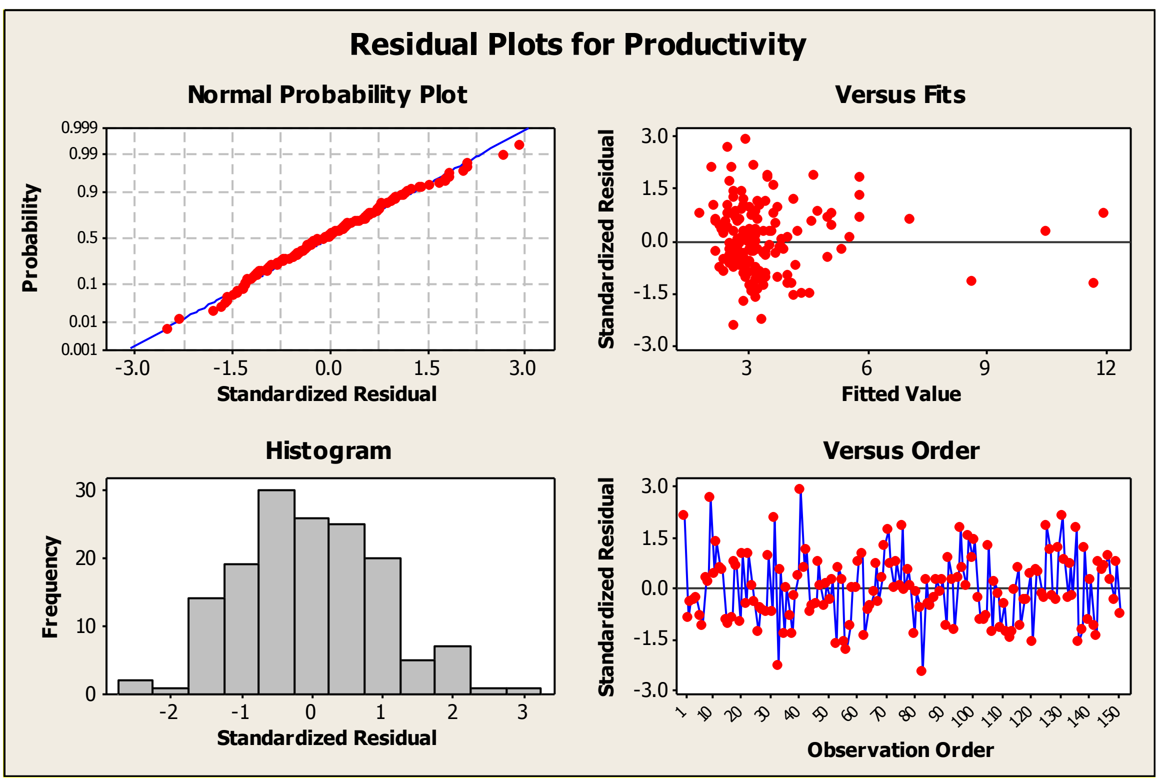


Figure 4- 6: Residual plots for Medium mines

Consideration of Multicollinearity

Several of the independent variables in the model were highly correlated with each other. These variables can be categorized in two categories, the first category consists of all the interaction terms in the model that contain the dummy variable “LW,” and the second consists of SH and SH^2, which was also the case in Very Large/Large mines. The former variables accounted for five interaction terms of RL.LW, CTN_D.LW, NFDL_IR.LW, USD_CTN.LW, and SH.LW, which were highly correlated in pairs. The correlation coefficients of these pairs are listed in Table 4-15. High VIF values for these variables also indicate the existence of multicollinearity in the model (Table 4-15). The variance inflation factor values of 50.80 and 50.87 for the variables SH and SH^2 indicate a strong collinearity between these two variables. This can be easily fixed with normalizing these variables by subtracting the mean and dividing by the standard deviation. The variables SH.LW and CTN_D.LW also showed strong collinearity with VIFs of 69.74 and 66.56, followed by RL.LW, NFDL_IR.LW, and USD_CTN.LW with VIFs of 18.39, 8.32, and 7.97 (Table 4-16). Fixing the multicollinearity is beyond the scope of this study.

Conclusion

Productivity of the 31 randomly selected Medium mines were regressed against 49 variables indicating the safety measures and structural factors of these mines. After five iterations of regressing the PTVT against these variables and examining the residuals the model yielded eleven significant variables of RL.LW, RL^2, USD_CTN.LW, NFDL_IR.LW, SH^2, CTN_D.LW, SH, SM.Y06, SH.LW, SH.Y07, and RL.Y05. Increases in the Reliability of the Medium mines improves their productivity. The productivity of the Medium non-longwall mines increases when their Seam Height increases from 33 to 60 inches, whereas when the Seam Height decreases from 60 to 108 inches the productivity increases. Moreover, Medium mines showed to be less productive in 2007. The impact on the productivity of the longwall mines was indicated by the interaction terms containing the dummy variables LW. Longwall mines with thicker Seam Height were less productive; these mines even had a lower productivity in 2007. The model indicated that longwall Medium mines with higher Reliability had a higher productivity. On the other hand, a higher amount of Dollars per Citation for Longwall mines decreased the productivity of these mines. Moreover, an increase in the number of citations per inspection day and NFDL_IR of longwall mines had a positive impact on their productivity. Some effects by Year were also observed in this size category. In 2005, with an increase in the Reliability of the Medium mines the productivity decreased. In 2006, the Severity Measure had a negative impact on the productivity. In 2007, SH had a negative impact on the productivity. Overall, this model with 11 independent variables achieved the high R-squared value of 82.6%; however, the explanation of the variables in the model was not as straight forward as in the Very Large/Large mines.

Small/Very Small Mines

Forward and Backward Stepwise techniques were used in Minitab 15 to find the significant independent variables impacting the productivity of Small/Very Small mines. The Small/Very Small mines sample also included 155 data points, which were the data for 31 Small/Very Small mines over the 5-year period of 2003-2004. Since longwall mining is not used in this size of underground coal mines, the interaction terms containing the dummy variable “LW” were eliminated from the sample, thus Productivity was regressed against 42 independent variables. Regression diagnosis was done to obtain the “best” model. After 7 iterations and removing 14 data points due to their high standardized residual values, the best model with five significant independent variables was selected.

Discussions of Correlation among Variables

Table 4-19 illustrates Pearson correlation coefficients of the variables in this model. The NFDL_IR^2, with correlation coefficient of 0.556, was the most correlated variable with the response, entering the model first (Table 4-19). No high correlations among the independent variables in the model were observed.

Estimated "Best" Regression Function

After seven iterations for finding the "best" model, the one with five variables of NFDL IR squared (NFDL_IR^2), Severity Measure (SM), Dollar per Citations in 2005 (USD_CTN.Y05), Severity Measure in 2005 (SM.Y05), and Seam Height (SH) turned out to be significant. Table 4-20, derived from the Minitab output in Appendix C-3, illustrates the coefficients and the standard error of the coefficients of these variables. Equation 4-2 shows the "best" regression function explaining the relationship between safety measures and structural factors of Small/Very Small mines with their productivity. This section solely discusses this model in details. The detailed results of the other iterations can be found in Appendix C-3.

Table 4- 19: Correlation matrix for Small/Very Small mines

	PTVT	NFDL_IR^2	SM	USD_CTN.Y05	SM.Y05
NFDL_IR^2	0.556				
SM	0.522	0.445			
USD_CTN.Y05	0.266	0.035	0.01		
SM.Y05	0.453	0.463	0.211	0.148	
SH	0.213	0.014	-0.054	0.019	-0.041

Table 4- 20: Regression coefficient statistics for Small/Very Small mines

Predictor	Coefficient	Standard Error Coefficient	T-Value	P-Value	VIF
Constant	1.7069	0.2358	7.24	0.000	---
NFDL_IR^2	0.001953	0.000492	3.97	0.000	1.523
SM	0.000217	3.88E-05	5.59	0.000	1.252
USD_CTN.Y05	0.000751	0.000204	3.69	0.000	1.025
SM.Y05	0.000558	0.000164	3.41	0.001	1.305
SH	0.016472	0.004067	4.05	0.000	1.008

The constant term in the model cannot be interpreted since SH cannot be equal to zero. While holding the other variables constant, the other terms in the model are interpreted as follows: the mean PTVT increases by 0.00195 for every unit increase in NFDL_IR^2; the mean of PTVT increases by 0.000217 when the Severity Measure increases by 1 unit; ; the mean of PTVT increases 0.000751 when the amount of dollar per citation in 2005 increase by $1; the mean of PTVT increases by 0.00558 when the Severity Measure in 2005 increases by 1 unit; the mean of PTVT increases by 0.0165 when the Seam Height increases by 1 inch.

$$PTVT = 1.71 + 0.00195.NFDL\_IR^2 + 0.000217.SM + 0.000751.USD\_CTN.Y05 + 0.00558.SM.Y05 + 0.0165.SH \quad (4\text{-}3)$$

Analysis of Variance

The model in Equation 4-3 was found to be significant with a P-value equal to zero. After removing 14 influential data points the degrees of freedom of the model reduced to 140. The SSR and SSE were calculated as 233.216 and 188.603, summing up to 421.819, which is equal to the SSTO value. Since the SRR is not much larger than SSE a large R-squared value is not expected for this model. The estimation of the model's variance was calculated as 1.397, which is not a very small value for variance (Table 4-21). The NFDL_IR^2, with the sequential sum of squares of 130.481, had the largest share of 55.95% in explaining the variations in the model. The second largest contributor was SM with the a Sequential SS of 39.618, which explains 16.99% of the variations in the model. The USD_CTN.Y05, SH, and SM.Y05 had the Sequential SS of 26.041, 22.912, and 14.165, respectively (Table 4-22)

Table 4- 21: Analysis of variance for Small/Very Small mines 2003-2007

Source	DF	SS	MS	F	P
Regression	5	233.216	46.643	33.39	0.000
Residual Error	135	188.603	1.397		
Total	140	421.819			

Table 4- 22: Sequential sum of squares of the regression model for Small/Very Small mines 2003-2007

Source	DF	Sequential SS	% of SSTO
NFDL_IR^2	1	130.481	55.95%
SM	1	39.618	16.99%
USD_CTN.Y05	1	26.041	11.17%
SM.Y05	1	14.165	6.07%
SH	1	22.912	9.82%

Coefficient of Multiple Determination (R^2)

The R^2 value of the regression model for Medium mines was calculated as 55.29%, meaning the five independent variables in the model explain 55.29% of the variation in the model. Moreover, to consider the effect of the number of variables in the model, the adjusted R^2 was calculated to be 53.63%, which is slightly smaller than R^2.

Examinations of the Residual Plots

The upper-right graph of Figure 4-7 depicts the standardized residuals versus the fitted values of the response. This plot does not show any curvature, trend, or funnel shape; however, the data suggests that Small/Very Small mines no 84, 55, 140, 72 with fitted values of 11.2910, 10.2876, 8.8146, and 8.0344, respectively, have large leverages (Appendix C-3). Moreover, all the data points have the standardized residual of less than 3.0. The upper-left graph, the Normal Probability Plot, does not show a perfectly linear shape; however, it is acceptable. The lower-left graph, Histogram, also shows a slightly skewed bell-shape histogram of the distribution of the residuals, which again represents the normality of the residuals. The lower-right graph, "Versus Order," depicts how well the residuals are spread up and down. Finally the standard deviation of the residuals was calculated to be S=1.18197.

Consideration of Multicollinearity

The Pearson correlation coefficients of less than 0.50 for the variables shows that no two of them are highly correlated (Table 4-19). Also by looking at the VIF values of the independent variables, no VIF larger than 10 was found, meaning no multicollinearity exists in the model (Table 4-20).

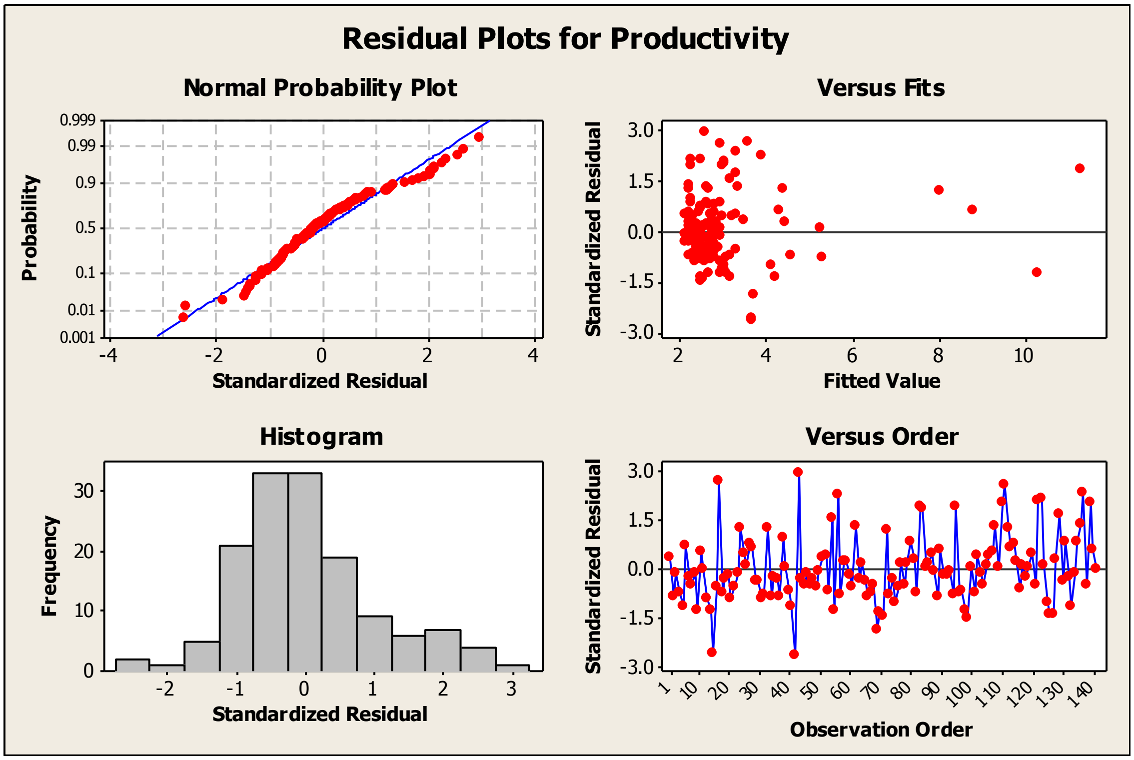


Figure 4- 7: Residual plots for Small/Very Small mines

Conclusion

The productivity of 31 randomly selected Small/Very Small mines were regressed against 42 variables indicating the safety measures and structural factors of these mines. After conducting a regression diagnosis, 14 data points were removed from the sample and finally the sample with 141 data points was significant with 5 variables, i.e., NFDL_IR^2, SM, USD_CTN.Y05, SM.Y05, and SH. Increases in the NFDL_IR^2 improves the productivity of the S/VS mines, while SM also has a positive impact on enhancing the productivity. In 2005, a higher Severity Measure had an additional impact on the productivity of the S/VS mines. Seam height again shows a positive impact on productivity; however, the impact was not as strong as in Very

Large/Large mines. Finally the model suggests that the larger the amount of Dollars per Citation in 2005 the more productive the S/VS mines were in that year; in other words, more serious violations in S/VS mines increases the productivity of these mines. Moreover the residual plots for this size category was not as clean as the other two size categories and needed a lot of cleaning to get an acceptable model. This significant variation can be due to the different nature of these mines, especially when it comes to the culture and supervision.

Chapter 5

SUMMARY AND CONCLUSIONS

Summary

The Mine Safety and Health Administration (MSHA) maintains very thorough databases on the employment, accident/injuries/illnesses, inspections, and citations of all the mines in the U.S. These data can be accessed in different ways, such as by year or by mine, which makes the identification of data relatively easy. The industry, government, labors, and researchers can use these databases to observe the production and safety trends and analyze problems in the mining industry. By comparing year-by-year outputs from the mines, influencing events such as incidents in a particular year or changes in the rules and regulations can be identified and scrutinized in more details.

Three mines disasters in the USA in 2006 were the main stimulus for the passage of the MINER Act of 2006, which resulted in stricter safety regulations to improve health, safety, preparedness, and emergency response primarily in underground coal mines (Congress, 2006). A significant increase in the penalty amounts was another outcome of this legislation. The changes and the constant decrease in the productivity of the U.S. underground coal mines in the period of 2003-2007 were the incentive to study the productivity of these mines and identify the impacting factors.

MSHA databases were the main source of data for this study, in which only underground coal mines were of interest. The mines' data were gathered for the 5-year period of 2003-2007. Using the MSHA mine ID, the large database was reduced to a smaller database which included

604 mines whose data were available for all the 5 years. Moreover, 150 of these mines were eliminated since they had no production and employees, over the period of 2003-2007. This reduced the number of mines to 454. The average number of employees was used as the criterion to categorize the mines by size as: Very Large, Large, Medium, Small, and Very Small (Grayson, 2001). Three more categories (Small with no Production, Very Small with no Production, and No-Employee) were added in addition to what Grayson et al. [2001] had used to identify mines with no production.

In the mine-size transition study, the size transition of 454 mines from the eight aforementioned size categories were studied by size and by year. In this analysis, changes in the number of mines in each mine-size category were investigated by tracking the number of mines in each size category from 2003 through 2007. Using more thorough analysis, transition rates (in percentage) of each mine-size category to the other size categories were investigated during four one-year periods of 2003-2004, 2004-2005, 2005-2006, and 2006-2007. Transition rates of the mine-size categories for the four one-year periods were presented in tables 4-3 through 4-6. Finally, a summary of these transition rates was presented in a summary table (Table 4-7) showing average values of the transition rates in the four one-year periods.

In the productivity impact analysis, only mines with the coal production of 10,000 tons over the 5-year period were used. Due to the similarity of some categories and shortage of enough number of consistently productive and faithful mines, the number of mine-size categories was reduced to three, i.e., Very Large/Large, Medium, and Small/Very Small. From a population of 178 Very Large/Large, Medium, and Small/Very Small consistently productive and faithful mines, 93 mines, 31 from each aforementioned category, were randomly selected. The required information for these mines were extracted from MSHA databases for the 5-year period of 2003-2007 to calculate the dependent variable, Productivity, and some independent variables, i.e., Inspection Days (INS_D), Dollars per Citation (USD_CTN), Number of Citations per Inspection

Day, (CTN_D) Reliability (RL), Non-Fatal Days Lost Injury Rate (NFDL_IR), Severity Measure (SM), Seam Height (SH), dummy variable Longwall (LW), and four dummy variables for Year 04-07 (Y04, Y05, Y06, and Y07), representing 5-years of 2003-2007. Interaction terms including a dummy variable for each year were used to target the impact of the variable(s) in specific years. The impact of longwall mines on other variables was also considered by using interaction terms including the dummy variable Longwall. Forward and Backward regression methods were used in Minitab 15 to generate separate multi-linear regression models regressing Productivity on the proposed independent variables for each mine-size category. Regression diagnosis was run for each model aiming at a reasonably high R^2 value and standardized residuals of less than 3.0. The "best" models contained the significant variables impacting the productivity of the mines for each size category.

Conclusions

In the period of 2003-2007, the number of mines in each category did not remain the same and kept changing from one year to another (Table 4-1). The Small mine-size category with an average of 131 mines in that category during the 5-year period had the largest number of mines, followed by Very Small with 95 mines. No-Employee mine-size category with a standard deviation of 17.6 showed the highest variation in the number of mines followed by Small and Very Small size -categories with a standard deviation of 17.3 and 10.8, respectively. On the other hand, Large mine-size category followed by Very Large showed the lowest variability in the number of mines in each year with standard deviations of 1.5 and 5.0, respectively.

Averaging the transition rates of the mine-size categories over the five years of 2003-2007, Very Large mines, followed by Large mines, were found to be the least variable mines with averages of 97.3% and 83.6%, respectively, of them staying in their initial mine-size categories.

On the other hand, No-Employee and Very Small with no Employee mines with 59.0% and 60.7%, respectively, remaining in their initial size categories were the most variable mine size categories. These results are not exactly in the line with the result obtained using the standard deviations of the number of mines in each category; however, these results are more accurate and reliable.

The best multi-linear regression model generated for Very Large/Large mines included six variables of squared term of seam height (SH^2), seam height (SH), severity measure (SM), seam height in 2007 (SH.Y07), reliability in 2007 (RL.Y07), and amount of penalty per citation in 2006 (USD_CTN.Y06). The R^2 value was calculated as 70.61%, meaning 70.61% of the variations in the productivity were explained by the aforementioned variables. Seam Height, overall, had a strongly positive impact on the productivity of the VL/L mines, meaning when Seam Height increased the productivity of the mines increased; however in 2007 the productivity dropped with respect to Seam Heigh. Moreover, the Severity Measure surprisingly had a positive impact on the productivity, although very minor. An interesting question to pose is, “did the exposure to risk increase due to the focus on production?” The amount of Dollars per Citation, which is representative of the significance of violations, had a negative impact on the productivity of these mines in 2006. This can reflect the impact of the increase in the amount of penalties in the latter part of 2006, when the MINER Act was amended which made the safety regulations stricter and increased the amount of penalties for violations. This a very important phenomenon reflecting the impact of safety regulations as a structural factor on the productivity of underground coal mines. Finally, in 2007, Reliability of the mines had a positive impact on the productivity of these mines. Thus mines with lower probability for incurring citations had higher productivity. This could show that after one year the mines got more familiar with the new regulation (MINER Act) and received fewer citations which improved their Reliability.

In the case of Medium mines, after five iterations of regressing Productivity (PTVT) against the 49 proposed variables and examining the residuals, the model with 11 variables, i.e., Reliability of Longwall mines (RL.LW), squared term of Reliability (RL^2), amount of Dollars per Citation in Longwall mines (USD_CTN.LW), Non-Fatal Days Lost Incident Rate of Longwall mines (NFDL_IR.LW), squared term of Seam Height (SH^2), number of Citations per Day in Longwall mines (CTN_D.LW), Seam Height (SH), Severity Measure in 2006 (SM.Y06), Seam Height of Longwall mines (SH.LW), Seam Height in 2007 (SH.Y07), and Reliability in 2005 (RL.Y05) were found to comprise the "best" model. These variables explained 82.6% of the variation in the productivity. The model suggests that an increase in the Reliability of the Medium mines improves their productivity. The productivity of the Medium non-longwall mines decreases when their Seam Height increases from 33 to 60 inches, whereas when the Seam Height increases from 60 to 108 inches the productivity increases. Moreover, Medium mines showed to be less productive with respect to increase in their Seam Height in 2007. Longwall mines with a thicker Seam Height were less productive; these mines even had a lower productivity in 2007. The model indicated that longwall Medium mines with higher Reliability had a higher productivity. On the other hand, a higher amount of Dollars per Citation for Longwall mines decreased the productivity of these mines. Moreover, an increase in the number of citations per inspection day and NFDL_IR of longwall mines showed a positive relationship with their productivity. In 2005, with an increase in the Reliability of the Medium mines the productivity decreased. In 2006, the Severity Measure had a negative impact on the productivity. In 2007, SH had a negative impact on the productivity. Overall, less lost work days lost due to non-compliance with the safety rules and regulations improved the productivity of Medium mines. This size category did not provide a very clear model as in the Very Large/Large mine-size category. This could be because of the nine longwall mines that existed in this size category.

The "best" multi-linear regression model for Small/Very Small mines was obtained after eliminating 14 data points, whose standardized residuals were greater than 3.0. In this model, five variables, i.e., squared term of Non-Fatal Days Lost Incident Rate (NFDL_IR^2), Severity Measure (SM), Dollar amount of penalty per Citation in 2005 (USD_CTN.Y05), Severity Measure in 2005 (SM.Y05), and Seam Height (SH) explained 55.29% of the variation in the productivity. Seam Height, square term of Non-Fatal Days Lost Incident Rate, and Severity Measure showed to have positive relationship with the productivity of S/VS mines, meaning a mine with a higher value on any of these variables would get a higher productivity. In 2005, Severity Measure and the amount of Dollars per Citation also had additional positive relationship with productivity of the S/VS mines. Furthermore, the model suggests that the larger the amount of Dollars per Citation in 2005 the more productive the S/VS mines were in that year; in other words, more serious violations in S/VS mines increased the productivity of these mines in 2005. These two 2005 year effects on S/VS mines can be more scrutinized to see what happened in 2005 that caused these phenomena. Overall, in Small/Very Small mines, mines with more incidents resulting in loosing work days were the more productive mines. All in all, this size category showed a lot more variation compared to the other two size categories. This significant variation may be due to the different nature of these mines, especially skills, culture, and type of supervision.

Recommendations from Analysis

Mine operators, agencies and researchers can make use of the extensive database on the production, injury incidents, inspections, and citations provided by MSHA to study and scrutinize the factors impacting the productivity of individual mines and possibly use the results to mitigate apparent problems. Severity Measure as a significant factor in all mine sizes must be studied further to find the causes of the lost days. Reliability was an improving factor for the productivity of the Very Large/Large and Medium mines. The mines with a lower probability of receiving a citation were more productive. Thus to have a more productive mine, attention should be given to the events and incidents that result in receiving citations and proactively try to avoid them.

Limitations of the Study

In this study, only mines which could be tracked during the 5-year period of 2003-2007 were studied. This reduced the number of mines in the productivity impact analysis sample dramatically. Moreover, filtering these mines further to consistently productive and faithful mines eliminated many mines from the sample and reduced its size even more. Using the same sample of 454 mines, which was used for the mine size transition study, could give a truer representation of the industry. Results could be used in identifying the factors that significantly impact productivity and drop the production of the coal below 10,000 tons per year. In this study the interaction terms of the non-dummy variables were not used to avoid complication of the interpretation of the model; using the interaction terms could result in a better model with its variables explaining a larger proportion of the variance in the productivity.

Future Work

The dramatic changes in the transition rates of some mine-size categories in the period of 2006-2007 deserves to be more thoroughly scrutinized to see if the mines that downgraded or went out of business were impacted by the passage of MINER Act of 2006. Mine transitions also can be studied using mine status instead of mine size. This could give a picture of how mines change their status over time. To have a more detailed analysis, mine status transitions could be studied for each mine-size category separately. In this way, more familiarity with the trends in each size category may be achieved.

As mentioned before, Severity Measure as a significant variable in the three models deserves more attention. In this study only Lost Days and Restricted Days were considered in the calculation of this variable. For future research, Statutory Days could also be considered in the calculation of the Severity Measure to see whether any difference in the result is observed.

Instead of generating a separate model for each-mine size category, a model may be generated including all the mines by introducing some dummy variables indicating the size of the mines. The result of such an analysis may be then compared to the result derived from the productivity impact part of this study.

REFERENCES

Camm T. and Girard-Dwyer J. Economic Consequences of Mining Injuries [Journal] // Mining Engineering. - [s.l.] : Society for Mining, Metallurgy, and Exploration Inc., 2004. - pp. 89-92.

Coleman Patrick J. and Kerkering John C. Measuring mining safety with injury statistics: Lost workdays as indicators of risk [Journal] // Journal of Safety Research. - [s.l.] : Elsevier Ltd, 2007. - 5 : Vol. 38. - pp. 523-533.

Congress The U.S. Mine Improvement and New Emergency Response Act of 2006 (MINER Act) [Report]. - 2006. - Available at http://www.msha.gov/mineract/mineractsinglesource.asp.. - Pub. L. No. 109-236 (S 2803). .

Crowson Philip Mine size and the structure of costs [Journal] // Resources Policy. - [s.l.] : Elsevier Ltd, 2003. - 1-2 : Vol. 29. - pp. 15-36.

Grayson R. L. Planning a Balanced National Mine Safety and Health Research Program Using Risk Analysis and Stakeholder Input [Journal] // Transactions-Society for Mining, Metallurgy, and Exploration, Inc.. - [s.l.] : Society for Mining, Metallurgy, and Exploration, Inc., 2001. - Vol. 310. - pp. 55-62.

Grayson R. Larry Mine Safety Technology and Training Commision Update, State of Coal, Mine Safety, and What's Happend [Conference] // Presentation at: ACI Joint Summit. - St. Louis : [s.n.], 2009.

Grayson R. Larry, Kinilakodi Harsisha and Kecojevic Vladislav Pilot sample risk analysis for underground coal mine fires and explosions using MSHA citation data [Journal] // Safety Science. - University Park : Elsevier, 2009.

Grayson R.L. Safety vs. productivity and other factors in US underground coal mines [Journal] // Mining Engineering. - 2001. - pp. 41-44.

Hartman Howard L. SME Mining Engineering Handbook [Book]. - 1992. - 2nd : Vol. 1.

International Coal Group, Inc. Wolf Run Mining Company to Permanently Close Sago Mine [Online] // International Coal Group, Inc. 2008 Press Release Archives. - International Coal Group, Inc., 12 12, 2008. - 7 22, 2010. - http://www.intlcoal.com/pages/news/2008/20081212.pdf.

Kinilakodi Harisha Analysis of Major Hazard Risk Impact on Underground Coal Mine Safety Performence [Report] : M.S. Thesis. - University Park : Penn State, 2009.

Kniesner Thomas J. and Leeth John D. Data Mining Mining Data: MSHA Enforcement Efforts, Underground Coal Mine Safety, and New Health Policy Implications [Journal] // The Journal of Risk and Uncertainty. - [s.l.] : Springer Netherlands, 2004. - Vol. 29. - pp. 83–111.

Kucuker Hudaverdi Occupational fatalities among coal mine workers in Zonguldak, Turkey, 1994–2003 [Journal] // Occupational Medicine. - [s.l.] : Oxford University Press, 2006. - Vol. 56. - pp. 144-146.

Kulshreshtha Mudit and Parikh Jyoti K. A study of productivity in the Indian coal sector [Journal] // Energy Policy. - [s.l.] : Butterworth-Heinemann, 2001. - Vol. 29. - pp. 701-713.

Li Z. and Topuz E. Evaluating mine size and mine life - an objective approach. [Journal] // Mining science & technology. - January 1988. - 2 : Vol. 6. - pp. 117-124.

MSHA History of previous violations per inspection day (VPID) repeat violations per inspection day (RPID) [Online] // Mine Data Retrival System. - 2007. - 2010. - http://www.msha.gov/drs/ASP/MineAction.asp.

MSHA Inspection day inclusions [Online] // Data Retrieval System. - 2010. - 3 15, 2010. - http://www.msha.gov/drs/InspectionDayInclusions.asp.

MSHA Mine Data Retrieval System [Online] // MSHA. - 2010. - 7 8, 2010. - http://www.msha.gov/drs/drshome.htm.

MSHA Mine Injury and Worktime Quarterly Statistics-Coal Data [Online] // MSHA. - 2010. - 7 7, 2010. - http://www.msha.gov/ACCINJ/ALLCOAL.HTM.

MSHA Mine Safety and Health Enforcement [Online] // MSHA. - 2006. - 2010. - http://www.msha.gov/MSHAINFO/FactSheets/MSHAFCT4.HTM.

MSHA Part 50 data user's handbook [Book]. - 1993.

MSHA Title 30 Code of Federal Regulations Part 50 [Online] // Title 30 CFR. - 7 1, 2009. - 7 22, 2010. - http://www.msha.gov/30CFR/CFRINTRO.HTM.

Passmore D. L. and Bennett J. D. Confirmatory factor analysis model of the reliability of a measure of the severity of coal [Conference] // AIME Transactions. - Littleton : Society of Mining Engineers of AIME, 1985. - Vol. 280. - pp. 2111-2114.

Ramani Raja V. and Mutmansky Jan M. Mine Health and Safety at the Turn of the Millenium [Journal] // Mining Engineering. - [s.l.] : Society for Mining Metallurgy and Exploration, 1999. - Vol. 51. - pp. 25-30.

Szwilski A.B. Economic Environment of Coal Mining Operations in Appalachia, United States [Journal] // Mining Science and Technology. - [s.l.] : Elsevier Science Publishers B.V., May 1987. - 1 : Vol. 5. - pp. 1-10.

Szwilski A.B. Significance and Measurement of Coal Mine Productivity [Journal] // Mining Science and Technolog. - Amesterdam : Elsevier Science Publishers B.V., 1988. - 6. - pp. 221-231.

Weeks James L. Occupational Health and Safety Regulation in the Coal Mining Industry: Public Health at the Workplace [Journal] // Annual Review of Public Health. - [s.l.] : ANNUAL REVIEWS INC, 4139 EL CAMINO WAY, PO BOX 10139, PALO ALTO, CA 94303-0139, 1991. - Vol. 12. - pp. 195-207.

Weir International Inc. United States Longwall Mining Statistics 1989-2007 [Report]. - [s.l.] : Weir International, Inc. Mining, Geology and Energy Consultants, 2007.

APPENDICES

Appendix A: MSHA Sorted Database- CD

The MSHA database files are available in a CD in the Department of Energy and Mineral Engineering at The Pennsylvania State University, University Park. They can also be obtained by sending me a request email at: safa.es@gmail.com.

Appendix B: Mine-size transition matrices

Table B- 1: Mine-size transition matrix for the period of 2003-2004

		2004								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	Row Sum
2003	VL	30	0	0	0	0	0	0	0	30
	L	5	40	2	1	0	0	0	0	48
	M	0	6	38	6	0	1	0	0	51
	S	0	0	14	102	0	11	3	3	133
	S-NP	0	0	0	0	1	0	0	0	1
	VS	0	0	3	19	0	70	8	9	109
	VS-NP	0	0	1	3	0	5	16	8	33
	NE	0	1	1	15	0	10	5	17	49
Column Sum		35	47	59	146	1	97	32	37	454

Table B- 2: Mine-size transition matrix for the period of 2004-2005

		2005								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	Row Sum
2004	VL	35	0	0	0	0	0	0	0	35
	L	5	39	3	0	0	0	0	0	47
	M	0	5	50	4	0	0	0	0	59
	S	0	1	14	112	0	13	1	5	146
	S-NP	0	0	0	0	1	0	0	0	1
	VS	0	0	2	21	0	61	9	4	97
	VS-NP	0	0	1	6	0	6	16	3	32
	NE	0	0	0	3	1	8	4	21	37
Column Sum		40	45	70	146	2	88	30	33	454

Table B- 3: Mine-size transition matrix for the period of 2005-2006

		2006								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	Row Sum
2005	VL	39	1	0	0	0	0	0	0	40
	L	3	39	3	0	0	0	0	0	45
	M	0	9	55	5	0	0	1	0	70
	S	0	0	12	111	0	16	2	5	146
	S-NP	0	0	0	1	1	0	0	0	2
	VS	0	0	0	5	0	75	3	5	88
	VS-NP	0	0	0	2	1	3	19	5	30
	NE	0	0	0	3	0	6	2	22	33
Column Sum		42	49	70	127	2	100	27	37	454

Table B- 4: Mine-size transition matrix for the period of 2006-2007

		2007								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	Row Sum
2006	VL	39	3	0	0	0	0	0	0	42
	L	2	40	2	1	0	0	3	1	49
	M	0	5	45	12	0	2	2	4	70
	S	0	0	9	78	0	19	6	15	127
	S-NP	0	0	0	0	2	0	0	0	2
	VS	0	0	1	11	0	57	9	22	100
	VS-NP	0	0	0	1	0	1	23	2	27
	NE	0	0	0	1	0	2	2	32	37
Column Sum		41	48	57	104	2	81	45	76	454

Table B- 5: Average mine-size transition matrix for the period of 2003-2007

		k+1								
	Mine Size	VL	L	M	S	S-NP	VS	VS-NP	NE	Row Sum
k	VL	35.8	1.0	0.0	0.0	0.0	0.0	0.0	0.0	36.8
	L	3.8	39.5	2.5	0.5	0.0	0.0	0.8	0.3	47.3
	M	0.0	6.3	47.0	6.8	0.0	0.8	0.8	1.0	62.5
	S	0.0	0.3	12.3	100.8	0.0	14.8	3.0	7.0	138.0
	S-NP	0.0	0.0	0.0	0.3	1.3	0.0	0.0	0.0	1.5
	VS	0.0	0.0	1.5	14.0	0.0	65.8	7.3	10.0	98.5
	VS-NP	0.0	0.0	0.5	3.0	0.3	3.8	18.5	4.5	30.5
	NE	0.0	0.3	0.3	5.5	0.3	6.5	3.3	23.0	39.0
Column Sum		39.5	47.3	64.0	130.8	1.8	91.5	33.5	45.8	454.0

Appendix C-1: Minitab output for Very Large/Large mines

Iteration 1: All Very Large/Large mines in

Stepwise Regression: PTVT versus 49 independent variables

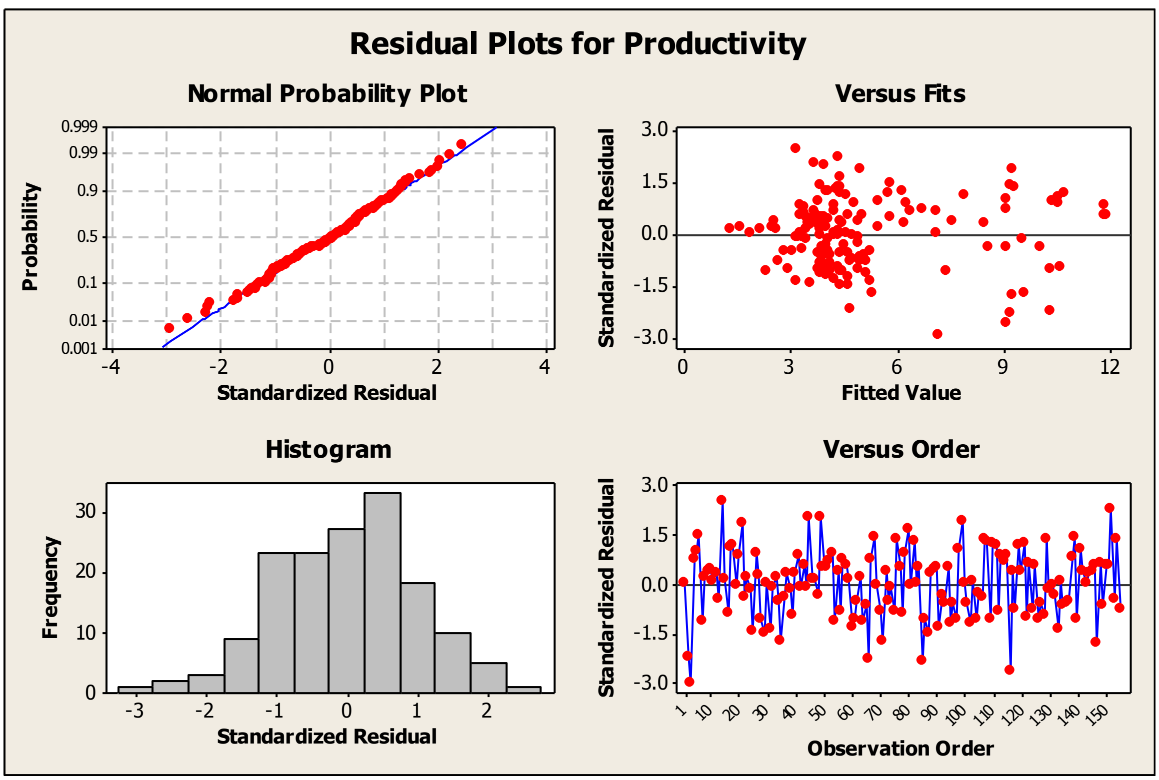
```
Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05


Response is PTVT on 49 predictors, with N = 155

Step                  1        2        3         4         5         6
Constant          1.280    1.394    9.471    10.051    10.096    10.212

SH^2            0.00061  0.00063  0.00183   0.00187   0.00195   0.00208
T-Value           14.22    15.35     5.52      5.87      6.32      6.66
P-Value           0.000    0.000    0.000     0.000     0.000     0.000

SH.Y07                   -0.0174  -0.0165   -0.0193   -0.0568   -0.0555
T-Value                    -4.03    -3.97     -4.78     -4.92     -4.84
P-Value                    0.000    0.000     0.000     0.000     0.000

SH                                 -0.204    -0.211    -0.219    -0.236
T-Value                             -3.63     -3.93     -4.21     -4.53
P-Value                             0.000     0.000     0.000     0.000

USD_CTN.Y06                                -0.00227  -0.00218  -0.00211
T-Value                                       -3.82     -3.79     -3.70
P-Value                                       0.000     0.000     0.000

RL.Y07                                                    6.8       6.7
T-Value                                                  3.45      3.46
P-Value                                                 0.001     0.001

SM                                                              0.00137
T-Value                                                            2.03
P-Value                                                           0.045

S                  1.82     1.74     1.67      1.60      1.55      1.53
R-Sq              56.92    61.09    64.22     67.39     69.80     70.61
R-Sq(adj)         56.64    60.58    63.51     66.52     68.78     69.42
PRESS           524.771  481.746  446.827   410.292   388.759   381.801
R-Sq(pred)        55.59    59.23    62.19     65.28     67.10     67.69
```

Regression Analysis: PTVT versus 6 independent variables

```
The regression equation is
      PTVT = 10.2 + 0.00208 SH^2 - 0.0555 SH.Y07 - 0.236 SH
            - 0.00211 USD_CTN.Y06 + 6.73 RL.Y07 + 0.00137 SM


Predictor              Coef    SE Coef      T      P     VIF
Constant             10.212      2.058   4.96  0.000
SH^2              0.0020767  0.0003117   6.66  0.000  75.048
SH.Y07             -0.05547    0.01145  -4.84  0.000   9.310
SH                 -0.23597    0.05212  -4.53  0.000  73.883
USD_CTN.Y06      -0.0021067  0.0005697  -3.70  0.000   1.049
RL.Y07                6.733      1.948   3.46  0.001   9.122
SM                0.0013712  0.0006767   2.03  0.045   1.179


S = 1.53183         R-Sq = 70.6%        R-Sq(adj) = 69.4%

PRESS = 381.801     R-Sq(pred) = 67.69%


Analysis of Variance

Source           DF       SS      MS      F      P
Regression        6   834.47  139.08  59.27  0.000
Residual Error  148   347.28    2.35
Total           154  1181.75


Source           DF  Seq SS
SH^2              1  672.70
SH.Y07            1   49.21
SH                1   36.97
USD_CTN.Y06       1   37.44
RL.Y07            1   28.51
SM                1    9.63
```

```
Unusual Observations

Obs      SH^2       PTVT     Fit     SE Fit   Residual   St Resid
2         9216      1.450    4.718   0.386    -3.268     -2.20R
3        11664      2.800    7.174   0.399    -4.374     -2.96R
4        12996      7.790    6.752   0.595     1.038      0.74 X
5        11664      6.870    5.501   0.607     1.369      0.97 X
14        5184      6.900    3.193   0.327     3.707      2.48R
18        5184      7.300    5.743   0.756     1.557      1.17 X
21        4356      7.630    4.972   0.568     2.658      1.87 X
33        3844      1.890    1.613   0.574     0.277      0.20 X
45        5184      6.790    3.698   0.225     3.092      2.04R
49        5184      7.040    3.959   0.191     3.081      2.03R
56        5776      1.540    2.689   0.586    -1.149     -0.81 X
59        3364      1.520    1.334   0.653     0.186      0.13 X
66       12996      6.970   10.311   0.329    -3.341     -2.23R
85       11664      5.770    9.232   0.254    -3.462     -2.29R
116      11664      5.180    9.096   0.260    -3.916     -2.59R
152       3364      7.730    4.339   0.273     3.391      2.25R
R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.
```

Residual Plots for PTVT



Correlations Matrix for Very Large/Large Mines

	PTVT	SM	SH	SH^2	RL.Y07	SH.Y07
SM	-0.164					
SH	0.727	-0.317				
SH^2	0.754	-0.337	0.993			
RL.Y07	-0.101	-0.111	0.061	0.053		
SH.Y07	-0.087	-0.157	0.158	0.152	0.939	
USD_CTN.Y06	-0.185	-0.016	-0.078	-0.073	-0.196	-0.196

Cell Contents: Pearson correlation

Appendix C-2: Minitab output for Medium Mines

Iteration 1: All Medium mines in

Stepwise Regression: PTVT versus 49 independent variables

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is PTVT on 49 predictors, with N = 155

Step	1	2	3	4	5	6
Constant	3.231	2.465	2.402	2.465	2.115	5.453
RL.LW	10.25	10.43	5.01	3.83	3.98	4.20
T-Value	11.07	12.31	3.20	2.54	2.72	2.98
P-Value	0.000	0.000	0.002	0.012	0.007	0.003
RL^2		2.76	2.94	2.76	2.56	2.36
T-Value		5.57	6.18	6.13	5.79	5.51
P-Value		0.000	0.000	0.000	0.000	0.000
NFDL_IR.LW			0.397	0.669	0.663	0.673
T-Value			4.04	5.99	6.12	6.44
P-Value			0.000	0.000	0.000	0.000
USD_CTN.LW				-0.00237	-0.00239	-0.00235
T-Value				-4.39	-4.55	-4.64
P-Value				0.000	0.000	0.000
SH^2					0.00009	0.00076
T-Value					3.16	3.99
P-Value					0.002	0.000
SH						-0.098
T-Value						-3.56
P-Value						0.001
S	1.25	1.14	1.09	1.03	0.996	0.959
R-Sq	44.46	53.86	58.36	63.09	65.41	68.14
R-Sq(adj)	44.10	53.26	57.53	62.11	64.25	66.84
Mallows Cp	182.6	128.1	103.1	76.7	64.7	50.4
PRESS	257.087	218.748	256.100	228.640	227.976	208.562
R-Sq(pred)	39.86	48.83	40.09	46.52	46.67	51.21

Step	7	8	9	10	11	12
Constant	5.534	5.513	5.452	5.220	5.062	5.080
RL.LW	3.9	3.8	4.1	8.7	15.8	16.0
T-Value	2.86	2.80	3.10	4.18	6.00	6.16
P-Value	0.005	0.006	0.002	0.000	0.000	0.000
RL^2	2.14	1.81	1.79	1.67	1.77	1.72
T-Value	5.00	4.21	4.27	4.06	4.52	4.46
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
NFDL_IR.LW	0.664	0.672	0.647	0.724	0.484	0.478
T-Value	6.50	6.76	6.63	7.31	4.36	4.37
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
USD_CTN.LW	-0.00208	-0.00209	-0.00203	-0.00124	-0.00403	-0.00400
T-Value	-4.15	-4.28	-4.25	-2.27	-4.70	-4.73
P-Value	0.000	0.000	0.000	0.025	0.000	0.000
SH^2	0.00077	0.00074	0.00072	0.00068	0.00067	0.00067
T-Value	4.12	4.07	4.09	3.94	4.05	4.11
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
SH	-0.098	-0.092	-0.088	-0.080	-0.077	-0.076
T-Value	-3.61	-3.49	-3.40	-3.18	-3.21	-3.22
P-Value	0.000	0.001	0.001	0.002	0.002	0.002
SH.Y07	-0.0081	-0.0102	-0.0126	-0.0132	-0.0133	-0.0143
T-Value	-2.81	-3.52	-4.27	-4.57	-4.82	-5.22
P-Value	0.006	0.001	0.000	0.000	0.000	0.000
USD_CTN.Y06		-0.00138	-0.00168	-0.00167	-0.00160	-0.00124
T-Value		-3.06	-3.71	-3.77	-3.80	-2.81
P-Value		0.003	0.000	0.000	0.000	0.006
SH.Y05			-0.0080	-0.0080	-0.0085	-0.0094
T-Value			-2.85	-2.92	-3.23	-3.61
P-Value			0.005	0.004	0.002	0.000
SH.LW				-0.050	-0.180	-0.183
T-Value				-2.83	-4.99	-5.14
P-Value				0.005	0.000	0.000
CTN_D.LW					10.3	10.4
T-Value					4.08	4.17
P-Value					0.000	0.000
SM.Y06						-0.00097
T-Value						-2.36
P-Value						0.020
S	0.938	0.912	0.891	0.870	0.826	0.813
R-Sq	69.76	71.58	73.09	74.50	77.16	78.02
R-Sq(adj)	68.32	70.03	71.42	72.73	75.40	76.16
Mallows Cp	42.6	33.7	26.6	20.1	6.2	3.0
PRESS	202.630	205.595	187.592	190.856	114.501	110.253
R-Sq(pred)	52.60	51.91	56.12	55.36	73.22	74.21

Regression Analysis: PTVT versus 12 independent variables

```
The regression equation is
      PTVT = 5.08 + 16.0 RL.LW + 1.72 RL^2 + 0.478 NFDL_IR.LW
            - 0.00400 USD_CTN.LW + 0.000666 SH^2 - 0.0763 SH - 0.0143 SH.Y07
            - 0.00124 USD_CTN.Y06 - 0.00941 SH.Y05 - 0.183 SH.LW
            + 10.4 CTN_D.LW - 0.000974 SM.Y06
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	5.0802	0.8149	6.23	0.000	
RL.LW	16.022	2.600	6.16	0.000	18.489
RL^2	1.7236	0.3863	4.46	0.000	1.189
NFDL_IR.LW	0.4777	0.1094	4.37	0.000	8.342
USD_CTN.LW	0.0040034	0.0008456	-4.73	0.000	7.972
SH^2	0.0006665	0.0001622	4.11	0.000	51.381
SH	-0.07635	0.02368	-3.22	0.002	51.511
SH.Y07	-0.014333	0.002745	-5.22	0.000	1.334
USD_CTN.Y06	-0.0012419	0.0004415	-2.81	0.006	1.338
SH.Y05	-0.009405	0.002604	-3.61	0.000	1.206
SH.LW	-0.18300	0.03558	-5.14	0.000	70.008
CTN_D.LW	10.395	2.491	4.17	0.000	66.657
SM.Y06	-0.0009743	0.0004131	-2.36	0.020	1.273

S = 0.813474   R-Sq = 78.0%   R-Sq(adj) = 76.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	333.537	27.795	42.00	0.000
Residual Error	142	93.967	0.662		
Total	154	427.504			

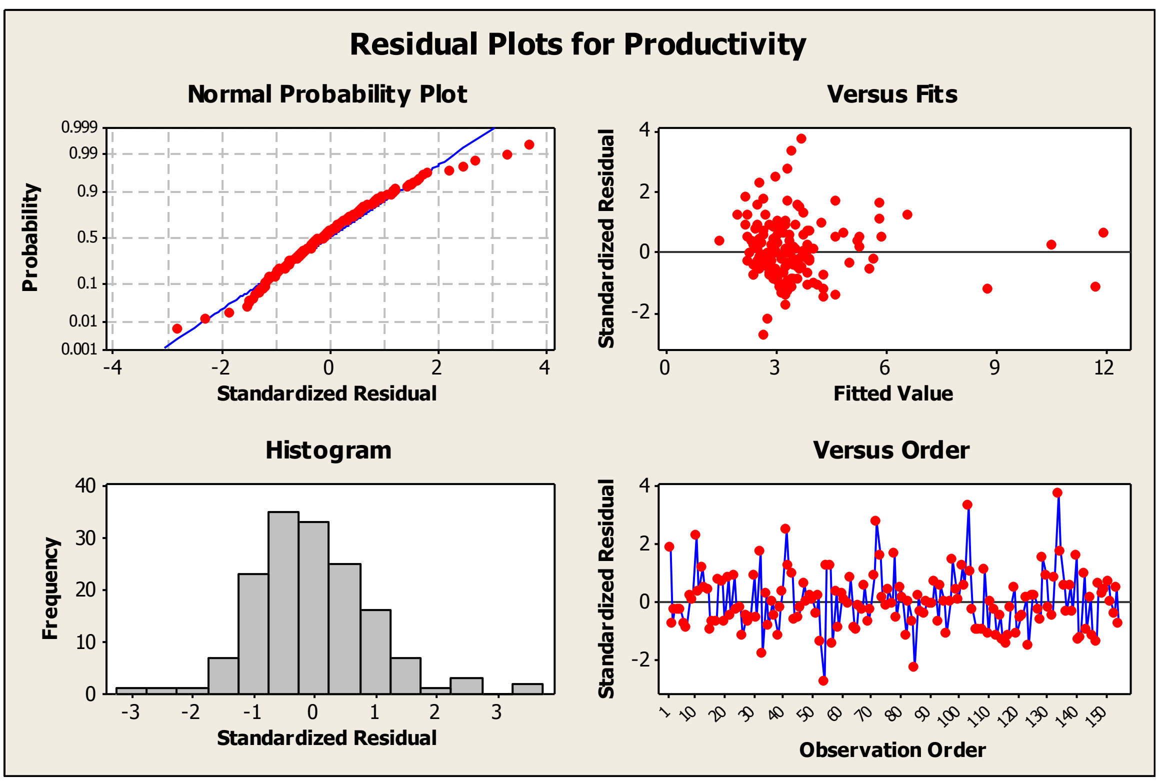
Source	DF	Seq SS
RL.LW	1	190.061
RL^2	1	40.211
NFDL_IR.LW	1	19.205
USD_CTN.LW	1	20.236
SH^2	1	9.921
SH	1	11.651
SH.Y07	1	6.937
USD_CTN.Y06	1	7.799
SH.Y05	1	6.427
SH.LW	1	6.053
CTN_D.LW	1	11.353
SM.Y06	1	3.682

Unusual Observations

Obs	RL.LW	PTVT	Fit	SE Fit	Residual	St Resid
10	0.000	4.3700	2.5598	0.1321	1.8102	2.26R
24	0.212	5.5700	5.6455	0.7708	-0.0755	-0.29 X
25	0.365	3.6100	3.6749	0.7570	-0.0649	-0.22 X
41	0.000	5.0000	2.9940	0.1050	2.0060	2.49R
54	0.000	0.5000	2.6614	0.2519	-2.1614	-2.79R
55	0.401	7.4300	6.5792	0.4117	0.8508	1.21 X
58	0.000	1.6500	1.4553	0.5289	0.1947	0.32 X
62	0.000	2.2000	2.2333	0.5981	-0.0333	-0.06 X
72	0.000	5.4900	3.3012	0.1458	2.1888	2.73R
80	0.530	4.9000	4.6286	0.5628	0.2714	0.46 X
85	0.000	0.9300	2.7464	0.1462	-1.8164	-2.27R
86	0.480	10.6800	10.5508	0.5249	0.1292	0.21 X
103	0.000	6.0800	3.4081	0.1046	2.6719	3.31R
117	0.687	11.2900	11.7551	0.7074	-0.4651	-1.16 X
128	0.457	5.2200	5.5685	0.5805	-0.3485	-0.61 X
134	0.000	6.6900	3.7018	0.1165	2.9882	3.71R
142	0.353	8.1100	8.7727	0.6150	-0.6627	-1.24 X
148	0.507	12.4100	12.0049	0.4856	0.4051	0.62 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for PTVT



Iteration 2: Medium mine no. 134 out

Stepwise Regression: PTVT versus 49 independent variables

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is PTVT on 49 predictors, with N = 154

Step	1	2	3	4	5	6
Constant	3.207	2.470	2.407	2.469	2.093	5.362
RL.LW	10.30	10.47	5.06	3.87	4.04	4.25
T-Value	11.37	12.61	3.31	2.64	2.86	3.13
P-Value	0.000	0.000	0.001	0.009	0.005	0.002
RL^2		2.67	2.84	2.67	2.44	2.26
T-Value		5.48	6.11	6.07	5.70	5.43
P-Value		0.000	0.000	0.000	0.000	0.000
NFDL_IR.LW			0.396	0.668	0.663	0.672
T-Value			4.12	6.15	6.32	6.67
P-Value			0.000	0.000	0.000	0.000
USD_CTN.LW				-0.00237	-0.00239	-0.00235
T-Value				-4.51	-4.73	-4.83
P-Value				0.000	0.000	0.000
SH^2					0.00010	0.00075
T-Value					3.52	4.09
P-Value					0.001	0.000
SH						-0.096
T-Value						-3.61
P-Value						0.000
S	1.22	1.12	1.06	0.998	0.962	0.926
R-Sq	45.98	54.94	59.53	64.40	67.14	69.82
R-Sq(adj)	45.62	54.34	58.72	63.44	66.03	68.59
Mallows Cp	200.5	144.4	116.6	87.0	71.2	55.8
PRESS	244.872	209.406	246.932	219.068	217.223	198.414
R-Sq(pred)	41.31	49.81	40.82	47.49	47.94	52.44

Step	7	8	9	10	11	12
Constant	5.442	5.422	5.185	5.035	4.973	4.992
RL.LW	4.0	3.8	8.5	15.4	16.0	16.1
T-Value	3.01	2.96	4.17	5.95	6.34	6.51
P-Value	0.003	0.004	0.000	0.000	0.000	0.000
RL^2	2.03	1.71	1.59	1.68	1.67	1.63
T-Value	4.92	4.13	3.91	4.35	4.47	4.41
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
NFDL_IR.LW	0.664	0.672	0.750	0.519	0.486	0.480
T-Value	6.74	7.02	7.73	4.77	4.58	4.60
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
USD_CTN.LW	-0.00209	-0.00210	-0.00129	-0.00401	-0.00404	-0.00401
T-Value	-4.33	-4.47	-2.42	-4.74	-4.92	-4.97
P-Value	0.000	0.000	0.017	0.000	0.000	0.000
SH^2	0.00076	0.00073	0.00069	0.00067	0.00066	0.00066
T-Value	4.24	4.19	4.04	4.15	4.20	4.27
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
SH	-0.096	-0.090	-0.083	-0.080	-0.076	-0.075
T-Value	-3.67	-3.55	-3.33	-3.38	-3.29	-3.30
P-Value	0.000	0.001	0.001	0.001	0.001	0.001
SH.Y07	-0.0080	-0.0100	-0.0107	-0.0106	-0.0130	-0.0140
T-Value	-2.86	-3.60	-3.90	-4.07	-4.95	-5.35
P-Value	0.005	0.000	0.000	0.000	0.000	0.000
USD_CTN.Y06		-0.00136	-0.00135	-0.00127	-0.00156	-0.00122
T-Value		-3.14	-3.18	-3.14	-3.89	-2.90
P-Value		0.002	0.002	0.002	0.000	0.004
SH.LW			-0.050	-0.177	-0.181	-0.184
T-Value			-2.92	-4.98	-5.26	-5.42
P-Value			0.004	0.000	0.000	0.000
CTN_D.LW				10.0	10.3	10.4
T-Value				4.02	4.28	4.38
P-Value				0.000	0.000	0.000
SH.Y05					-0.0080	-0.0090
T-Value					-3.22	-3.60
P-Value					0.002	0.000
SM.Y06						-0.00094
T-Value						-2.38
P-Value						0.018
S	0.904	0.878	0.856	0.814	0.788	0.776
R-Sq	71.42	73.24	74.74	77.30	78.84	79.66
R-Sq(adj)	70.05	71.76	73.16	75.71	77.20	77.93
Mallows Cp	47.4	37.6	29.9	15.3	7.3	4.0
PRESS	192.721	195.731	195.383	113.193	104.568	100.565
R-Sq(pred)	53.81	53.09	53.17	72.87	74.94	75.90

Regression Analysis: PTVT versus 12 independent variables

```
The regression equation is
      PTVT = 4.99 + 16.1 RL.LW + 1.63 RL^2 + 0.480 NFDL_IR.LW
            - 0.00401 USD_CTN.LW+ 0.000660 SH^2 - 0.0746 SH - 0.0140 SH.Y07
            - 0.00122 USD_CTN.Y06 - 0.184 SH.LW + 10.4 CTN_D.LW
            - 0.00896 SH.Y05 - 0.000939 SM.Y06
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	4.9916	0.7774	6.42	0.000	
RL.LW	16.136	2.480	6.51	0.000	18.485
RL^2	1.6294	0.3692	4.41	0.000	1.189
NFDL_IR.LW	0.4796	0.1043	4.60	0.000	8.339
USD_CTN.LW	-0.0040075	0.0008064	-4.97	0.000	7.970
SH^2	0.0006599	0.0001547	4.27	0.000	51.183
SH	-0.07464	0.02259	-3.30	0.001	51.297
SH.Y07	-0.014016	0.002619	-5.35	0.000	1.333
USD_CTN.Y06	-0.0012196	0.0004210	-2.90	0.004	1.337
SH.LW	-0.18377	0.03393	-5.42	0.000	69.983
CTN_D.LW	10.398	2.375	4.38	0.000	66.634
SH.Y05	-0.008957	0.002486	-3.60	0.000	1.206
SM.Y06	-0.0009391	0.0003940	-2.38	0.018	1.273

S = 0.775743   R-Sq = 79.7%   R-Sq(adj) = 77.9%

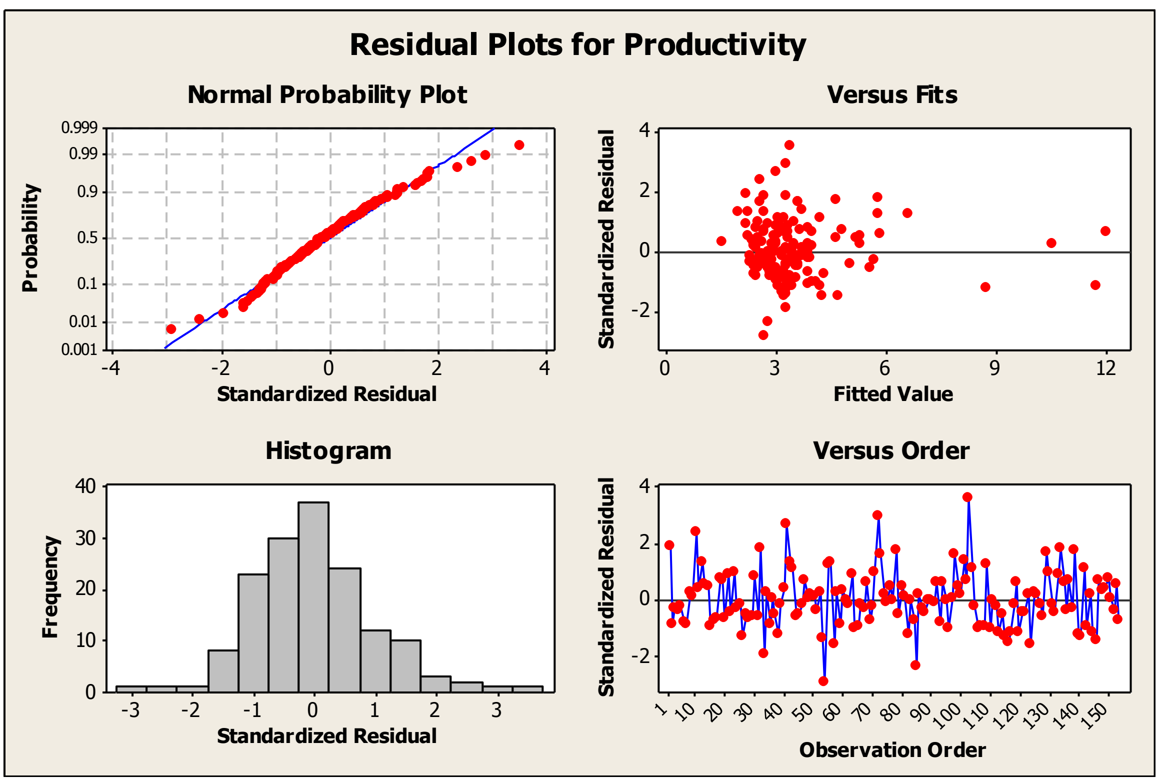
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	332.377	27.698	46.03	0.000
Residual Error	141	84.851	0.602		
Total	153	417.227			

Source	DF	Seq SS
RL.LW	1	191.833
RL^2	1	37.395
NFDL_IR.LW	1	19.132
USD_CTN.LW	1	20.318
SH^2	1	11.466
SH	1	11.162
SH.Y07	1	6.698
USD_CTN.Y06	1	7.571
SH.LW	1	6.248
CTN_D.LW	1	10.695
SH.Y05	1	6.442
SM.Y06	1	3.419

```
Unusual Observations

Obs  RL.LW         PTVT      Fit   SE Fit   Residual    St Resid
 10  0.000       4.3700   2.5373   0.1261     1.8327      2.39R
 24  0.212       5.5700   5.6518   0.7350    -0.0818     -0.33 X
 25  0.365       3.6100   3.6649   0.7219    -0.0549     -0.19 X
 41  0.000       5.0000   2.9610   0.1005     2.0390      2.65R
 54  0.000       0.5000   2.6361   0.2403    -2.1361     -2.90R
 55  0.401       7.4300   6.5852   0.3926     0.8448      1.26 X
 58  0.000       1.6500   1.4972   0.5044     0.1528      0.26 X
 62  0.000       2.2000   2.2864   0.5705    -0.0864     -0.16 X
 72  0.000       5.4900   3.2574   0.1395     2.2326      2.93R
 80  0.530       4.9000   4.6452   0.5367     0.2548      0.45 X
 85  0.000       0.9300   2.7412   0.1394    -1.8112     -2.37R
 86  0.480      10.6800  10.5654   0.5005     0.1146      0.19 X
103  0.000       6.0800   3.3616   0.1005     2.7184      3.53R
117  0.687      11.2900  11.7491   0.6745    -0.4591     -1.20 X
128  0.457       5.2200   5.5570   0.5535    -0.3370     -0.62 X
141  0.353       8.1100   8.7613   0.5865    -0.6513     -1.28 X
147  0.507      12.4100  12.0000   0.4631     0.4100      0.66 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.
```

Residual Plots for PTVT



Iteration 3: Medium mine no. 103 out

Stepwise Regression: PTVT versus 49 independent variables

```
Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05

Response is PTVT on 49 predictors, with N = 153


Step               1        2        3         4         5         6
Constant       3.187    2.448    2.385     2.447     2.047     5.222

RL.LW          10.34    10.51     5.09      3.90      4.09      4.29
T-Value        11.60    12.92     3.40      2.73      2.99      3.26
P-Value        0.000    0.000    0.001     0.007     0.003     0.001

RL^2                     2.68     2.85      2.68      2.44      2.25
T-Value                  5.60     6.27      6.25      5.88      5.62
P-Value                 0.000    0.000     0.000     0.000     0.000

NFDL_IR.LW                       0.397     0.669     0.663     0.672
T-Value                           4.23      6.31      6.54      6.90
P-Value                          0.000     0.000     0.000     0.000

USD_CTN.LW                              -0.00237  -0.00239  -0.00235
T-Value                                    -4.62     -4.88     -4.99
P-Value                                    0.000     0.000     0.000

SH^2                                               0.00010   0.00074
T-Value                                               3.85      4.16
P-Value                                              0.000     0.000

SH                                                            -0.093
T-Value                                                        -3.62
P-Value                                                        0.000

S               1.20     1.09     1.04     0.973     0.930     0.894
R-Sq           47.10    56.26    60.94     65.87     68.99     71.55
R-Sq(adj)      46.75    55.67    60.16     64.95     67.94     70.38
Mallows Cp     231.7    167.8    136.1     102.6      82.2      65.7
PRESS        236.447  200.835  238.076   210.612   207.777   189.778
R-Sq(pred)     42.39    51.06    41.99     48.68     49.37     53.76
```

Step	7	8	9	10	11	12
Constant	5.302	5.287	5.049	4.896	4.842	4.862
RL.LW	4.0	3.9	8.6	15.6	16.1	16.2
T-Value	3.15	3.10	4.37	6.26	6.65	6.83
P-Value	0.002	0.002	0.000	0.000	0.000	0.000
RL^2	2.04	1.73	1.60	1.70	1.69	1.65
T-Value	5.11	4.32	4.11	4.59	4.71	4.66
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
NFDL_IR.LW	0.664	0.672	0.751	0.518	0.486	0.480
T-Value	6.98	7.27	8.03	4.96	4.79	4.81
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
USD_CTN.LW	-0.00210	-0.00211	-0.00129	-0.00403	-0.00405	-0.00402
T-Value	-4.49	-4.64	-2.51	-4.97	-5.16	-5.21
P-Value	0.000	0.000	0.013	0.000	0.000	0.000
SH^2	0.00075	0.00072	0.00068	0.00066	0.00065	0.00065
T-Value	4.31	4.26	4.11	4.25	4.30	4.37
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
SH	-0.093	-0.088	-0.080	-0.077	-0.073	-0.072
T-Value	-3.69	-3.58	-3.35	-3.41	-3.33	-3.35
P-Value	0.000	0.000	0.001	0.001	0.001	0.001
SH.Y07	-0.0077	-0.0097	-0.0103	-0.0102	-0.0125	-0.0135
T-Value	-2.86	-3.60	-3.92	-4.11	-4.98	-5.39
P-Value	0.005	0.000	0.000	0.000	0.000	0.000
USD_CTN.Y06		-0.00132	-0.00130	-0.00122	-0.00150	-0.00117
T-Value		-3.13	-3.19	-3.15	-3.90	-2.91
P-Value		0.002	0.002	0.002	0.000	0.004
SH.LW			-0.051	-0.178	-0.182	-0.185
T-Value			-3.05	-5.23	-5.51	-5.68
P-Value			0.003	0.000	0.000	0.000
CTN_D.LW				10.1	10.4	10.4
T-Value				4.22	4.48	4.59
P-Value				0.000	0.000	0.000
SH.Y05					-0.0076	-0.0085
T-Value					-3.19	-3.58
P-Value					0.002	0.000
SM.Y06						-0.00090
T-Value						-2.39
P-Value						0.018
S	0.873	0.848	0.824	0.780	0.755	0.743
R-Sq	73.08	74.79	76.33	78.97	80.39	81.16
R-Sq(adj)	71.78	73.39	74.84	77.49	78.86	79.54
Mallows Cp	56.8	46.4	37.3	20.3	12.1	8.6
PRESS	184.517	187.704	187.350	104.331	96.3273	92.6303
R-Sq(pred)	55.04	54.26	54.35	74.58	76.53	77.43

Regression Analysis: PTVT versus 12 independent variables

```
The regression equation is
      PTVT = 4.86 + 16.2 RL.LW + 1.65 RL^2 + 0.480 NFDL_IR.LW
            - 0.00402 USD_CTN.LW
            + 0.000648 SH^2 - 0.0725 SH - 0.0135 SH.Y07 - 0.00117 USD_CTN.Y06
            - 0.185 SH.LW + 10.4 CTN_D.LW - 0.00853 SH.Y05 - 0.000901 SM.Y06
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	4.8625	0.7457	6.52	0.000	
RL.LW	16.218	2.376	6.83	0.000	18.480
RL^2	1.6489	0.3537	4.66	0.000	1.189
NFDL_IR.LW	0.48028	0.09992	4.81	0.000	8.337
USD_CTN.LW	-0.0040248	0.0007726	-5.21	0.000	7.969
SH^2	0.0006485	0.0001483	4.37	0.000	50.999
SH	-0.07247	0.02165	-3.35	0.001	51.097
SH.Y07	-0.013543	0.002512	-5.39	0.000	1.334
USD_CTN.Y06	-0.0011733	0.0004036	-2.91	0.004	1.338
SH.LW	-0.18453	0.03251	-5.68	0.000	69.958
CTN_D.LW	10.442	2.276	4.59	0.000	66.612
SH.Y05	-0.008531	0.002384	-3.58	0.000	1.207
SM.Y06	-0.0009008	0.0003776	-2.39	0.018	1.273

S = 0.743230   R-Sq = 81.2%   R-Sq(adj) = 79.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	333.056	27.755	50.24	0.000
Residual Error	140	77.335	0.552		
Total	152	410.391			

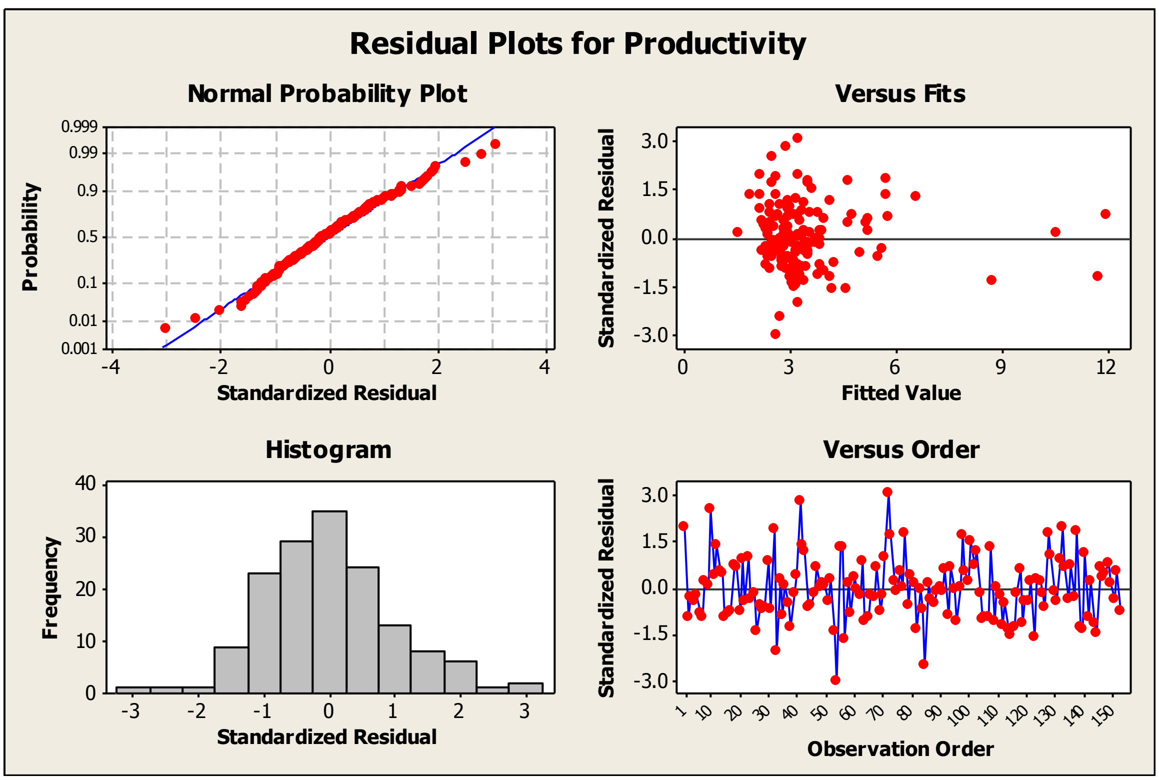
Source	DF	Seq SS
RL.LW	1	193.308
RL^2	1	37.561
NFDL_IR.LW	1	19.234
USD_CTN.LW	1	20.224
SH^2	1	12.818
SH	1	10.507
SH.Y07	1	6.250
USD_CTN.Y06	1	7.046
SH.LW	1	6.318
CTN_D.LW	1	10.836
SH.Y05	1	5.812
SM.Y06	1	3.143

Unusual Observations

Obs	RL.LW	PTVT	Fit	SE Fit	Residual	St Resid
10	0.000	4.3700	2.5118	0.1211	1.8582	2.53R
24	0.212	5.5700	5.6566	0.7042	-0.0866	-0.36 X
25	0.365	3.6100	3.6555	0.6917	-0.0455	-0.17 X
33	0.000	1.7900	3.2460	0.1580	-1.4560	-2.00R
41	0.000	5.0000	2.9267	0.0968	2.0733	2.81R
54	0.000	0.5000	2.6328	0.2302	-2.1328	-3.02R
55	0.401	7.4300	6.5988	0.3761	0.8312	1.30 X
58	0.000	1.6500	1.5541	0.4836	0.0959	0.17 X
62	0.000	2.2000	2.3289	0.5467	-0.1289	-0.26 X
72	0.000	5.4900	3.2353	0.1338	2.2547	3.08R
80	0.530	4.9000	4.6615	0.5142	0.2385	0.44 X
85	0.000	0.9300	2.7311	0.1336	-1.8011	-2.46R
86	0.480	10.6800	10.5785	0.4796	0.1015	0.18 X
116	0.687	11.2900	11.7452	0.6463	-0.4552	-1.24 X
127	0.457	5.2200	5.5404	0.5304	-0.3204	-0.62 X
140	0.353	8.1100	8.7514	0.5619	-0.6414	-1.32 X
146	0.507	12.4100	11.9917	0.4437	0.4183	0.70 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for PTVT



Iteration 4: Medium mine no. 72 out

Stepwise Regression: PTVT versus 49 independent variables

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is PTVT on 49 predictors, with N = 152

Step	1	2	3	4	5	6
Constant	3.171	2.454	2.391	2.453	2.035	5.168
RL.LW	10.37	10.54	5.12	3.93	4.13	4.33
T-Value	11.74	13.02	3.44	2.78	3.07	3.34
P-Value	0.000	0.000	0.001	0.006	0.003	0.001
RL^2		2.61	2.78	2.61	2.35	2.17
T-Value		5.48	6.14	6.12	5.72	5.47
P-Value		0.000	0.000	0.000	0.000	0.000
NFDL_IR.LW			0.397	0.669	0.662	0.671
T-Value			4.25	6.37	6.63	6.99
P-Value			0.000	0.000	0.000	0.000
USD_CTN.LW				-0.00237	-0.00240	-0.00236
T-Value				-4.67	-4.96	-5.08
P-Value				0.000	0.000	0.000
SH^2					0.00011	0.00073
T-Value					4.07	4.20
P-Value					0.000	0.000
SH						-0.092
T-Value						-3.63
P-Value						0.000
S	1.19	1.09	1.03	0.965	0.917	0.881
R-Sq	47.88	56.61	61.33	66.33	69.76	72.28
R-Sq(adj)	47.53	56.03	60.55	65.41	68.72	71.13
Mallows Cp	255.1	189.6	155.1	118.4	93.9	76.4
PRESS	231.036	197.574	234.966	207.178	203.468	185.833
R-Sq(pred)	43.13	51.36	42.16	49.00	49.91	54.25

Step	7	8	9	10	11	12
Constant	5.248	5.234	4.992	4.839	4.771	4.791
RL.LW	4.1	3.9	8.7	15.7	16.2	16.4
T-Value	3.23	3.19	4.50	6.43	6.93	7.12
P-Value	0.002	0.002	0.000	0.000	0.000	0.000
RL^2	1.96	1.65	1.53	1.62	1.60	1.56
T-Value	4.97	4.18	3.96	4.45	4.58	4.53
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
NFDL_IR.LW	0.663	0.671	0.751	0.518	0.483	0.477
T-Value	7.08	7.38	8.18	5.07	4.90	4.93
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
USD_CTN.LW	-0.00211	-0.00211	-0.00129	-0.00403	-0.00406	-0.00403
T-Value	-4.58	-4.73	-2.55	-5.07	-5.32	-5.38
P-Value	0.000	0.000	0.012	0.000	0.000	0.000
SH^2	0.00074	0.00071	0.00067	0.00066	0.00064	0.00064
T-Value	4.35	4.31	4.16	4.31	4.39	4.47
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
SH	-0.092	-0.086	-0.079	-0.076	-0.071	-0.071
T-Value	-3.69	-3.58	-3.35	-3.42	-3.34	-3.36
P-Value	0.000	0.000	0.001	0.001	0.001	0.001
SH.Y07	-0.0077	-0.0096	-0.0102	-0.0102	-0.0127	-0.0137
T-Value	-2.88	-3.62	-3.96	-4.16	-5.19	-5.62
P-Value	0.005	0.000	0.000	0.000	0.000	0.000
USD_CTN.Y06		-0.00131	-0.00129	-0.00121	-0.00152	-0.00119
T-Value		-3.16	-3.22	-3.19	-4.06	-3.04
P-Value		0.002	0.002	0.002	0.000	0.003
SH.LW			-0.051	-0.179	-0.183	-0.186
T-Value			-3.14	-5.36	-5.72	-5.90
P-Value			0.002	0.000	0.000	0.000
CTN_D.LW				10.1	10.4	10.5
T-Value				4.31	4.64	4.75
P-Value				0.000	0.000	0.000
SH.Y05					-0.0084	-0.0093
T-Value					-3.61	-4.01
P-Value					0.000	0.000
SM.Y06						-0.00090
T-Value						-2.47
P-Value						0.015
S	0.860	0.834	0.810	0.764	0.733	0.720
R-Sq	73.78	75.49	77.08	79.76	81.48	82.26
R-Sq(adj)	72.51	74.12	75.63	78.32	80.02	80.72
Mallows Cp	66.8	55.5	45.2	26.6	15.2	11.2
PRESS	180.697	183.897	183.172	100.054	90.7991	87.2831
R-Sq(pred)	55.52	54.73	54.91	75.37	77.65	78.51

Regression Analysis: PTVT versus 12 independent variables

```
The regression equation is
      PTVT = 4.79 + 16.4 RL.LW + 1.56 RL^2 + 0.477 NFDL_IR.LW
            - 0.00403 USD_CTN.LW + 0.000642 SH^2 - 0.0706 SH - 0.0137 SH.Y07
            - 0.00119 USD_CTN.Y06 - 0.186 SH.LW + 10.5 CTN_D.LW
            - 0.00933 SH.Y05 - 0.000902 SM.Y06
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	4.7913	0.7228	6.63	0.000	
RL.LW	16.390	2.303	7.12	0.000	18.483
RL^2	1.5573	0.3439	4.53	0.000	1.189
NFDL_IR.LW	0.47715	0.09682	4.93	0.000	8.335
USD_CTN.LW	-0.0040292	0.0007486	-5.38	0.000	7.967
SH^2	0.0006422	0.0001437	4.47	0.000	50.799
SH	-0.07058	0.02098	-3.36	0.001	50.897
SH.Y07	-0.013687	0.002435	-5.62	0.000	1.333
USD_CTN.Y06	-0.0011904	0.0003911	-3.04	0.003	1.337
SH.LW	-0.18577	0.03150	-5.90	0.000	69.941
CTN_D.LW	10.485	2.205	4.75	0.000	66.591
SH.Y05	-0.009329	0.002324	-4.01	0.000	1.209
SM.Y06	-0.0009023	0.0003659	-2.47	0.015	1.272

S = 0.720118   R-Sq = 82.3%   R-Sq(adj) = 80.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	334.149	27.846	53.70	0.000
Residual Error	139	72.081	0.519		
Total	151	406.230			

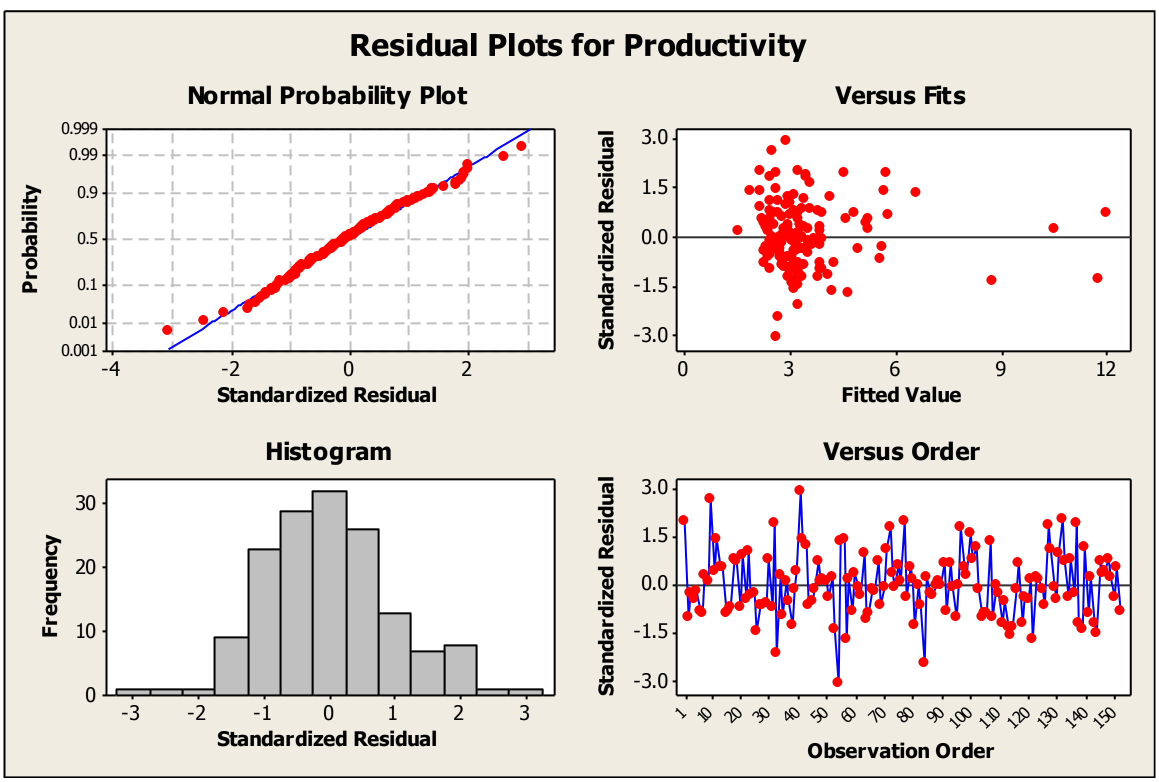
Source	DF	Seq SS
RL.LW	1	194.488
RL^2	1	35.492
NFDL_IR.LW	1	19.172
USD_CTN.LW	1	20.291
SH^2	1	13.940
SH	1	10.222
SH.Y07	1	6.125
USD_CTN.Y06	1	6.939
SH.LW	1	6.468
CTN_D.LW	1	10.858
SH.Y05	1	7.001
SM.Y06	1	3.154

Unusual Observations

Obs	RL.LW	PTVT	Fit	SE Fit	Residual	St Resid
1	0.000	3.5800	2.1722	0.1597	1.4078	2.00R
10	0.000	4.3700	2.4945	0.1174	1.8755	2.64R
24	0.212	5.5700	5.6429	0.6823	-0.0729	-0.32 X
25	0.365	3.6100	3.6743	0.6702	-0.0643	-0.24 X
33	0.000	1.7900	3.2821	0.1535	-1.4921	-2.12R
41	0.000	5.0000	2.9094	0.0939	2.0906	2.93R
54	0.000	0.5000	2.6117	0.2231	-2.1117	-3.08R
58	0.000	1.6500	1.5600	0.4685	0.0900	0.16 X
62	0.000	2.2000	2.3567	0.5298	-0.1567	-0.32 X
79	0.530	4.9000	4.6315	0.4983	0.2685	0.52 X
84	0.000	0.9300	2.6814	0.1304	-1.7514	-2.47R
85	0.480	10.6800	10.5524	0.4647	0.1276	0.23 X
115	0.687	11.2900	11.7537	0.6262	-0.4637	-1.30 X
126	0.457	5.2200	5.5563	0.5139	-0.3363	-0.67 X
132	0.000	4.6700	3.2335	0.1013	1.4365	2.01R
139	0.353	8.1100	8.7670	0.5445	-0.6570	-1.39 X
145	0.507	12.4100	12.0016	0.4299	0.4084	0.71 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for PTVT



Iteration 5: Medium mine no. 54 out

Stepwise Regression: PTVT versus 49 independent variables

```
Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05

Response is PTVT on 49 predictors, with N = 151


Step               1        2        3         4         5         6
Constant       3.190    2.465    2.401     2.464     2.055     5.030

RL.LW          10.33    10.50     5.08      3.89      4.09      4.28
T-Value        11.86    13.23     3.50      2.83      3.12      3.39
P-Value        0.000    0.000    0.001     0.005     0.002     0.001

RL^2                     2.65     2.82      2.64      2.39      2.22
T-Value                  5.66     6.36      6.37      5.98      5.72
P-Value                 0.000    0.000     0.000     0.000     0.000

NFDL_IR.LW                       0.396     0.668     0.662     0.671
T-Value                           4.35      6.54      6.82      7.17
P-Value                          0.000     0.000     0.000     0.000

USD_CTN.LW                              -0.00237  -0.00240  -0.00236
T-Value                                    -4.80     -5.10     -5.22
P-Value                                    0.000     0.000     0.000

SH^2                                               0.00010   0.00070
T-Value                                               4.08      4.10
P-Value                                              0.000     0.000

SH                                                            -0.088
T-Value                                                        -3.53
P-Value                                                        0.001

S               1.17     1.07     1.01     0.939     0.892     0.859
R-Sq           48.54    57.70    62.51     67.62     70.95     73.27
R-Sq(adj)      48.19    57.12    61.75     66.73     69.95     72.15
Mallows Cp     261.5    190.8    154.6     116.0      91.5      75.2
PRESS        223.778  189.342  226.827   198.972   195.732   179.532
R-Sq(pred)     43.70    52.37    42.94     49.95     50.76     54.84
```

Step	7	8	9	10	11
Constant	5.109	4.861	4.702	4.657	4.682
RL.LW	4.0	8.9	16.1	16.6	16.6
T-Value	3.28	4.60	6.66	7.05	7.39
P-Value	0.001	0.000	0.000	0.000	0.000
RL^2	2.00	1.87	1.95	2.17	2.05
T-Value	5.21	4.99	5.54	6.23	6.11
P-Value	0.000	0.000	0.000	0.000	0.000
NFDL_IR.LW	0.663	0.744	0.502	0.470	0.466
T-Value	7.28	8.11	4.95	4.74	4.92
P-Value	0.000	0.000	0.000	0.000	0.000
USD_CTN.LW	-0.00210	-0.00127	-0.00411	-0.00420	-0.00414
T-Value	-4.70	-2.50	-5.21	-5.47	-5.64
P-Value	0.000	0.013	0.000	0.000	0.000
SH^2	0.00071	0.00066	0.00065	0.00063	0.00062
T-Value	4.26	4.11	4.26	4.31	4.43
P-Value	0.000	0.000	0.000	0.000	0.000
SH	-0.087	-0.079	-0.076	-0.074	-0.071
T-Value	-3.60	-3.36	-3.43	-3.43	-3.44
P-Value	0.000	0.001	0.001	0.001	0.001
SH.Y07	-0.0079	-0.0086	-0.0086	-0.0100	-0.0120
T-Value	-3.06	-3.41	-3.65	-4.27	-5.23
P-Value	0.003	0.001	0.000	0.000	0.000
SH.LW		-0.052	-0.185	-0.190	-0.191
T-Value		-3.20	-5.58	-5.88	-6.20
P-Value		0.002	0.000	0.000	0.000
CTN_D.LW			10.5	10.9	10.9
T-Value			4.51	4.84	5.06
P-Value			0.000	0.000	0.000
RL.Y05				-0.88	-1.08
T-Value				-3.07	-3.86
P-Value				0.003	0.000
SM.Y06					-0.00125
T-Value					-3.75
P-Value					0.000
S	0.835	0.809	0.759	0.738	0.706
R-Sq	74.91	76.59	79.54	80.83	82.59
R-Sq(adj)	73.68	75.27	78.24	79.46	81.21
Mallows Cp	64.2	52.8	31.4	23.2	11.2
PRESS	173.974	173.195	92.2026	87.5275	81.6263
R-Sq(pred)	56.23	56.43	76.80	77.98	79.47

Regression Analysis: PTVT versus 11 independent variables

```
The regression equation is
      PTVT = 4.68 + 16.6 RL.LW + 2.05 RL^2 + 0.466 NFDL_IR.LW
            - 0.00414 USD_CTN.LW + 0.000624 SH^2 - 0.0707 SH - 0.0120 SH.Y07
            - 0.191 SH.LW + 10.9 CTN_D.LW - 1.08 RL.Y05 - 0.00125 SM.Y06
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	4.6820	0.7092	6.60	0.000	
RL.LW	16.641	2.251	7.39	0.000	18.393
RL^2	2.0497	0.3355	6.11	0.000	1.178
NFDL_IR.LW	0.46619	0.09482	4.92	0.000	8.323
USD_CTN.LW	-0.0041388	0.0007336	-5.64	0.000	7.968
SH^2	0.0006243	0.0001409	4.43	0.000	50.870
SH	-0.07072	0.02054	-3.44	0.001	50.801
SH.Y07	-0.012036	0.002300	-5.23	0.000	1.238
SH.LW	-0.19116	0.03083	-6.20	0.000	69.744
CTN_D.LW	10.922	2.160	5.06	0.000	66.561
RL.Y05	-1.0799	0.2797	-3.86	0.000	1.167
SM.Y06	-0.0012523	0.0003341	-3.75	0.000	1.105

S = 0.705597   R-Sq = 82.6%   R-Sq(adj) = 81.2%

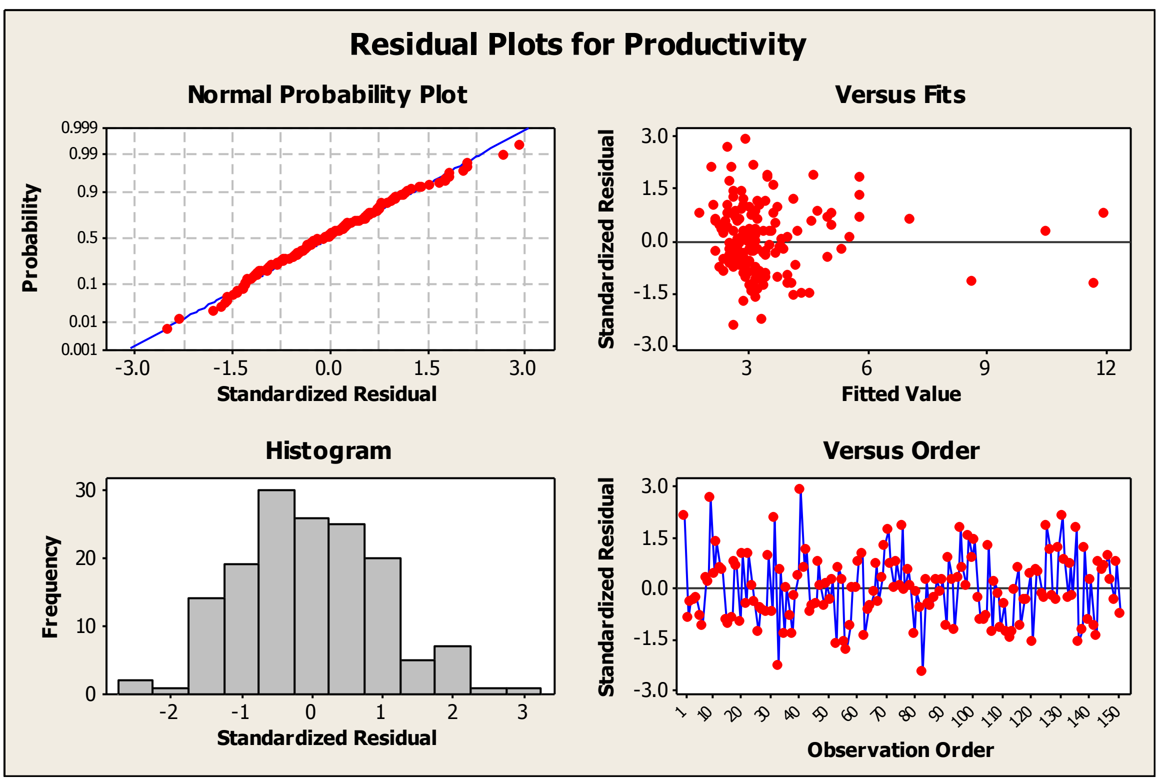
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	11	328.305	29.846	59.95	0.000
Residual Error	139	69.204	0.498		
Total	150	397.509			

Source	DF	Seq SS
RL.LW	1	192.951
RL^2	1	36.393
NFDL_IR.LW	1	19.143
USD_CTN.LW	1	20.312
SH^2	1	13.252
SH	1	9.189
SH.Y07	1	6.527
SH.LW	1	6.694
CTN_D.LW	1	11.737
RL.Y05	1	5.114
SM.Y06	1	6.994

```
Unusual Observations

Obs  RL.LW        PTVT      Fit   SE Fit   Residual    St Resid
  1  0.000      3.5800   2.1160   0.1562     1.4640      2.13R
 10  0.000      4.3700   2.5010   0.1150     1.8690      2.68R
 24  0.212      5.5700   5.5539   0.6679     0.0161      0.07 X
 25  0.365      3.6100   3.7078   0.6565    -0.0978     -0.38 X
 32  0.000      4.0300   2.5847   0.1331     1.4453      2.09R
 33  0.000      1.7900   3.3757   0.1344    -1.5857     -2.29R
 41  0.000      5.0000   2.9481   0.0928     2.0519      2.93R
 61  0.000      2.2000   1.7957   0.4863     0.4043      0.79 X
 78  0.530      4.9000   4.6154   0.4869     0.2846      0.56 X
 83  0.000      0.9300   2.6482   0.1313    -1.7182     -2.48R
 84  0.480     10.6800  10.5323   0.4566     0.1477      0.27 X
114  0.687     11.2900  11.7320   0.6135    -0.4420     -1.27 X
125  0.457      5.2200   5.3491   0.4996    -0.1291     -0.26 X
131  0.000      4.6700   3.1664   0.0935     1.5036      2.15R
138  0.353      8.1100   8.6691   0.5324    -0.5591     -1.21 X
144  0.507     12.4100  11.9664   0.4211     0.4436      0.78 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.
```

Residual Plots for PTVT



Correlation matrix

	PTVT	SH	RL^2	SH^2	SH.Y07	SM.Y06	RL.Y05
USD_CTN.LW							
SH	0.192						
RL^2	0.277	0.135					
SH^2	0.204	0.990	0.156				
SH.Y07	-0.196	0.148	-0.181	0.144			
SM.Y06	-0.144	0.055	-0.087	0.052	-0.166		
RL.Y05	0.002	0.062	0.252	0.063	-0.223	-0.160	
USD_CTN.LW	0.282	0.005	-0.124	-0.015	0.168	-0.052	-0.044
CTN_D.LW	0.536	0.016	-0.125	-0.010	0.069	-0.067	-0.014
RL.LW	0.697	-0.015	-0.036	-0.037	-0.040	-0.071	0.028
NFDL_IR.LW	0.695	-0.002	-0.078	-0.025	0.021	-0.065	-0.029
SH.LW	0.609	0.019	-0.090	-0.009	0.004	-0.070	0.008

	USD_CTN.LW	CTN_D.LW	RL.LW	NFDL_IR.LW
CTN_D.LW	0.859			
RL.LW	0.531	0.803		
NFDL_IR.LW	0.699	0.899	0.856	
SH.LW	0.700	0.946	0.930	0.900

Cell Contents: Pearson correlation

Appendix C-3: Minitab output for Small/Very Small Mines

Iteration 1: All Small/Very Small mines in

Stepwise Regression: PTVT versus 42 independent variables

```
  Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05


Response is PTVT on 42 predictors, with N = 155


Step              1        2        3        4
Constant      3.194    3.054    3.033    2.981

NFDL_IR^2   0.00430  0.00394  0.00356  0.00349
T-Value        5.49     5.12     4.59     4.55
P-Value       0.000    0.000    0.000    0.000

SM.Y04               0.00268  0.00279  0.00287
T-Value                 3.23     3.41     3.54
P-Value                0.002    0.001    0.001

SM^2                          0.00000  0.00000
T-Value                          2.35     2.42
P-Value                         0.020    0.017

USD_CTN.Y05                            0.00081
T-Value                                   2.10
P-Value                                  0.038

S              2.39     2.32     2.28     2.26
R-Sq          16.44    21.80    24.57    26.72
R-Sq(adj)     15.90    20.77    23.07    24.76
Mallows Cp      9.3      1.1     -2.2     -4.4
PRESS       938.462  910.283  1495.98  1389.49
R-Sq(pred)    10.18    12.88     0.00     0.00
```

Regression Analysis: PTVT versus 4 independent variables

```
The regression equation is
      PTVT = 3.01 + 0.00388 NFDL_IR^2 + 0.00275 SM.Y04 - 0.00004 SM.Y07
             + 0.000787 USD_CTN.Y05
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	3.0061	0.2096	14.34	0.000	
NFDL_IR^2	0.0038830	0.0007652	5.07	0.000	1.023
SM.Y04	0.0027478	0.0008271	3.32	0.001	1.029
SM.Y07	-0.000044	0.001004	-0.04	0.965	1.007
USD_CTN.Y05	0.0007865	0.0003927	2.00	0.047	1.004

S = 2.30313   R-Sq = 23.8%   R-Sq(adj) = 21.8%

PRESS = 893.824   R-Sq(pred) = 14.45%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	249.144	62.286	11.74	0.000
Residual Error	150	795.663	5.304		
Total	154	1044.807			

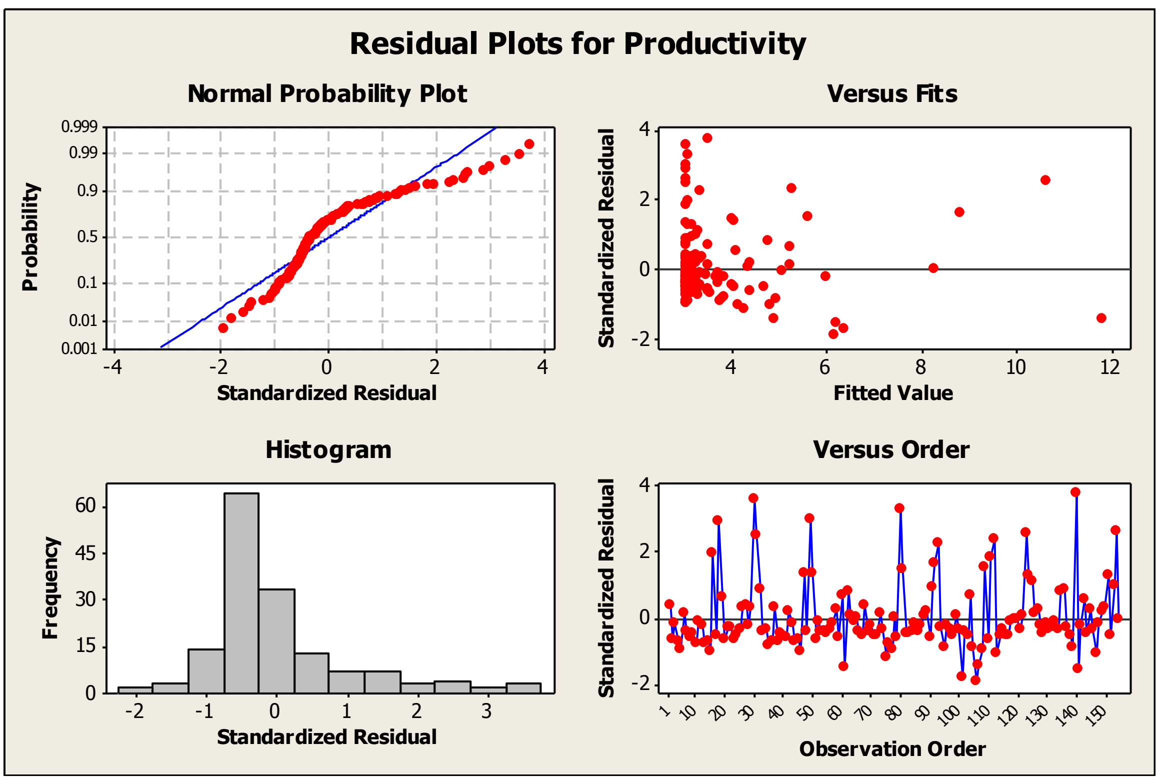
Source	DF	Seq SS
NFDL_IR^2	1	171.790
SM.Y04	1	55.996
SM.Y07	1	0.076
USD_CTN.Y05	1	21.282

Unusual Observations

Obs	NFDL_IR^2	PTVT	Fit	SE Fit	Residual	St Resid
16	23	7.230	3.057	0.881	4.173	1.96 X
18	0	9.650	3.006	0.210	6.644	2.90R
28	133	3.090	3.456	1.453	-0.366	-0.20 X
30	0	11.210	3.006	0.210	8.204	3.58R
31	0	8.770	3.006	0.210	5.764	2.51R
49	0	9.880	3.006	0.210	6.874	3.00R
61	2273	9.530	11.832	1.692	-2.302	-1.47 X
78	169	8.260	8.259	2.272	0.001	0.00 X
80	0	10.610	3.069	0.208	7.541	3.29R
92	1477	12.100	8.825	1.087	3.275	1.61 X
93	60	8.480	3.333	0.200	5.147	2.24R
101	149	2.540	6.376	0.801	-3.836	-1.78 X
106	130	1.940	6.148	0.757	-4.208	-1.93 X
112	130	10.520	5.259	0.503	5.261	2.34R
123	695	15.200	10.634	1.445	4.566	2.55RX
140	124	12.080	3.488	0.204	8.592	3.75R
154	0	9.000	3.006	0.210	5.994	2.61R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for PTVT



Iteration 2: Small/Very Small mines no. 140, 80, 49, 30 & 18 out

Stepwise Regression: PTVT versus 42 independent variables

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is PTVT on 42 predictors, with N = 150

Step	1	2	3	4	5	6
Constant	2.419	2.427	2.373	1.618	1.601	1.607
NFDL_IR	0.154	0.130	0.108	0.103	0.098	0.080
T-Value	7.14	5.99	5.06	4.90	4.70	3.58
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
SM		0.00021	0.00022	0.00024	0.00024	0.00023
T-Value		3.61	4.01	4.28	4.40	4.24
P-Value		0.000	0.000	0.000	0.000	0.000
SM.Y04			0.00266	0.00258	0.00269	0.00289
T-Value			4.00	3.94	4.17	4.49
P-Value			0.000	0.000	0.000	0.000
SH				0.0150	0.0149	0.0158
T-Value				2.50	2.52	2.69
P-Value				0.013	0.013	0.008
USD_CTN.Y05					0.00074	0.00067
T-Value					2.47	2.23
P-Value					0.015	0.027
SM.Y05						0.00050
T-Value						2.14
P-Value						0.034
S	1.98	1.90	1.81	1.78	1.75	1.73
R-Sq	25.62	31.69	38.44	40.99	43.39	45.15
R-Sq(adj)	25.12	30.76	37.18	39.36	41.42	42.85
Mallows Cp	35.1	22.3	7.9	3.7	-0.1	-2.4
PRESS	607.542	569.284	553.407	548.888	545.713	634.273
R-Sq(pred)	22.13	27.03	29.07	29.65	30.05	18.70

Regression Analysis: PTVT versus 6 independent variables

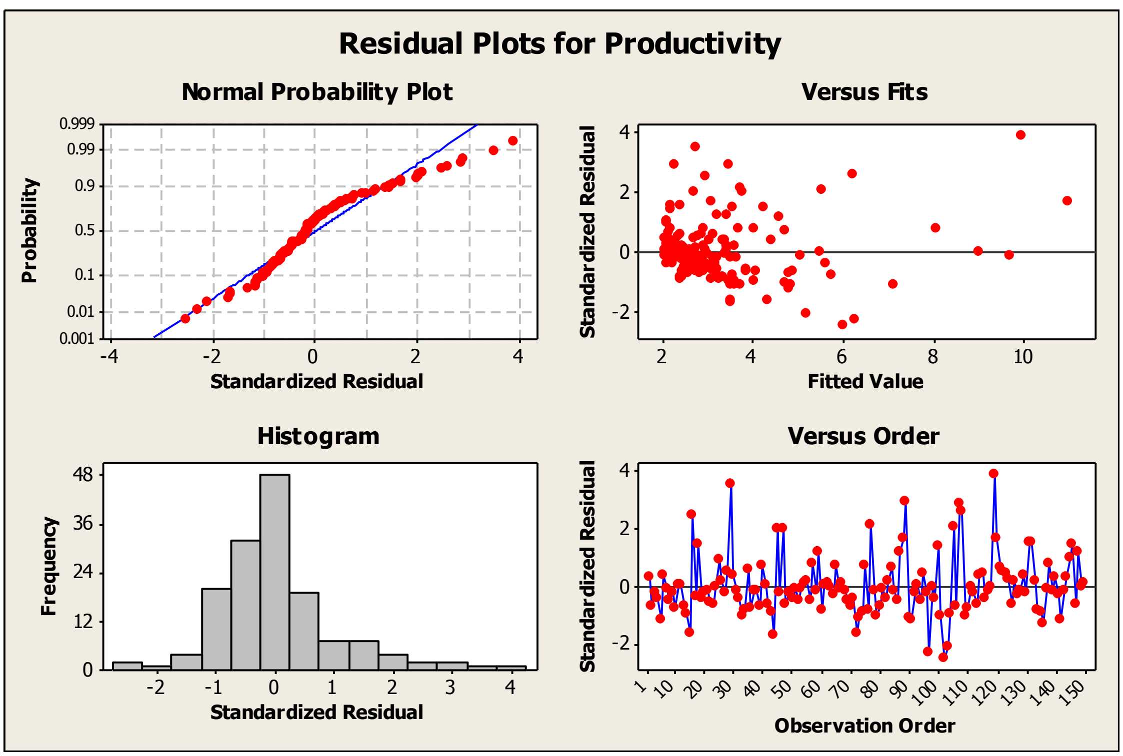
```
The regression equation is
      PTVT = 1.61 + 0.0798 NFDL_IR + 0.000228 SM + 0.00289 SM.Y04 + 0.0158 SH
            + 0.000665 USD_CTN.Y05 + 0.000505 SM.Y05


Predictor              Coef     SE Coef     T      P    VIF
Constant             1.6067      0.3449  4.66  0.000
NFDL_IR             0.07981     0.02231  3.58  0.000  1.405
SM               0.00022758  0.00005364  4.24  0.000  1.122
SM.Y04            0.0028930   0.0006436  4.49  0.000  1.102
SH                 0.015757    0.005847  2.69  0.008  1.023
USD_CTN.Y05       0.0006651   0.0002984  2.23  0.027  1.027
SM.Y05            0.0005046   0.0002353  2.14  0.034  1.258


S = 1.72988   R-Sq = 45.2%   R-Sq(adj) = 42.8%

PRESS = 634.273   R-Sq(pred) = 18.70%


Analysis of Variance

Source           DF       SS      MS      F      P
Regression        6  352.251  58.709  19.62  0.000
Residual Error  143  427.926   2.992
Total           149  780.177


Source           DF   Seq SS
NFDL_IR           1  199.917
SM                1   47.351
SM.Y04            1   52.636
SH                1   19.874
USD_CTN.Y05       1   18.716
SM.Y05            1   13.758
```

Unusual Observations

Obs	NFDL_IR	PTVT	Fit	SE Fit	Residual	St Resid
16	4.8	7.230	2.939	0.154	4.291	2.49R
29	0.0	8.770	2.741	0.224	6.029	3.51R
47	16.5	7.190	3.754	0.307	3.436	2.02R
58	47.7	9.530	9.697	1.108	-0.167	-0.13 X
60	13.7	4.460	5.755	0.661	-1.295	-0.81 X
75	13.0	8.260	8.058	1.708	0.202	0.74 X
77	15.6	7.340	3.732	0.278	3.608	2.11R
88	38.4	12.100	11.006	1.604	1.094	1.69 X
89	7.7	8.480	3.478	0.197	5.002	2.91R
90	21.4	5.370	7.133	0.681	-1.763	-1.11 X
97	12.2	2.540	6.256	0.621	-3.716	-2.30R
102	11.4	1.940	6.016	0.588	-4.076	-2.51R
103	5.9	1.630	5.170	0.398	-3.540	-2.10R
105	14.2	8.950	5.519	0.400	3.431	2.04R
107	0.0	7.230	2.268	0.189	4.962	2.89R
108	11.4	10.520	6.203	0.481	4.317	2.60R
119	26.4	15.200	9.969	1.086	5.231	3.88RX
149	0.0	9.000	9.010	1.595	-0.010	-0.01 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for PTVT



Iteration 3: Small/Very Small mines no. 119, 89 & 29 out

Stepwise Regression: PTVT versus 42 independent variables

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is PTVT on 42 predictors, with N = 147

Step	1	2	3	4	5	6
Constant	2.394	2.403	1.630	1.612	1.459	1.454
NFDL_IR	0.135	0.109	0.103	0.098	0.098	0.079
T-Value	6.98	5.72	5.47	5.31	5.38	4.12
P-Value	0.000	0.000	0.000	0.000	0.000	0.000
SM		0.00023	0.00024	0.00024	0.00025	0.00024
T-Value		4.50	4.85	5.00	5.23	5.08
P-Value		0.000	0.000	0.000	0.000	0.000
SH			0.0155	0.0154	0.0154	0.0166
T-Value			2.89	2.93	2.99	3.26
P-Value			0.005	0.004	0.003	0.001
USD_CTN.Y05				0.00071	0.00075	0.00067
T-Value				2.68	2.89	2.61
P-Value				0.008	0.005	0.010
RL.Y04					1.36	1.45
T-Value					2.50	2.71
P-Value					0.014	0.008
SM.Y05						0.00052
T-Value						2.58
P-Value						0.011
S	1.73	1.63	1.59	1.55	1.53	1.50
R-Sq	25.13	34.37	37.98	40.97	43.47	46.04
R-Sq(adj)	24.62	33.46	36.68	39.31	41.47	43.72
Mallows Cp	45.3	24.1	17.0	11.5	7.2	2.8
PRESS	453.808	400.789	389.985	453.256	400.494	498.950
R-Sq(pred)	21.82	30.95	32.82	21.92	31.01	14.04

Regression Analysis: PTVT versus 6 independent variables

```
The regression equation is
      PTVT = 1.45 + 0.0790 NFDL_IR + 0.000236 SM + 0.0166 SH
            + 0.000674 USD_CTN.Y05 + 1.45 RL.Y04 + 0.000523 SM.Y05
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	1.4542	0.3044	4.78	0.000	
NFDL_IR	0.07899	0.01916	4.12	0.000	1.311
SM	0.00023582	0.00004644	5.08	0.000	1.122
SH	0.016569	0.005088	3.26	0.001	1.026
USD_CTN.Y05	0.0006743	0.0002582	2.61	0.010	1.029
RL.Y04	1.4522	0.5358	2.71	0.008	1.012
SM.Y05	0.0005233	0.0002029	2.58	0.011	1.251

S = 1.49582   R-Sq = 46.0%   R-Sq(adj) = 43.7%

PRESS = 498.950   R-Sq(pred) = 14.04%

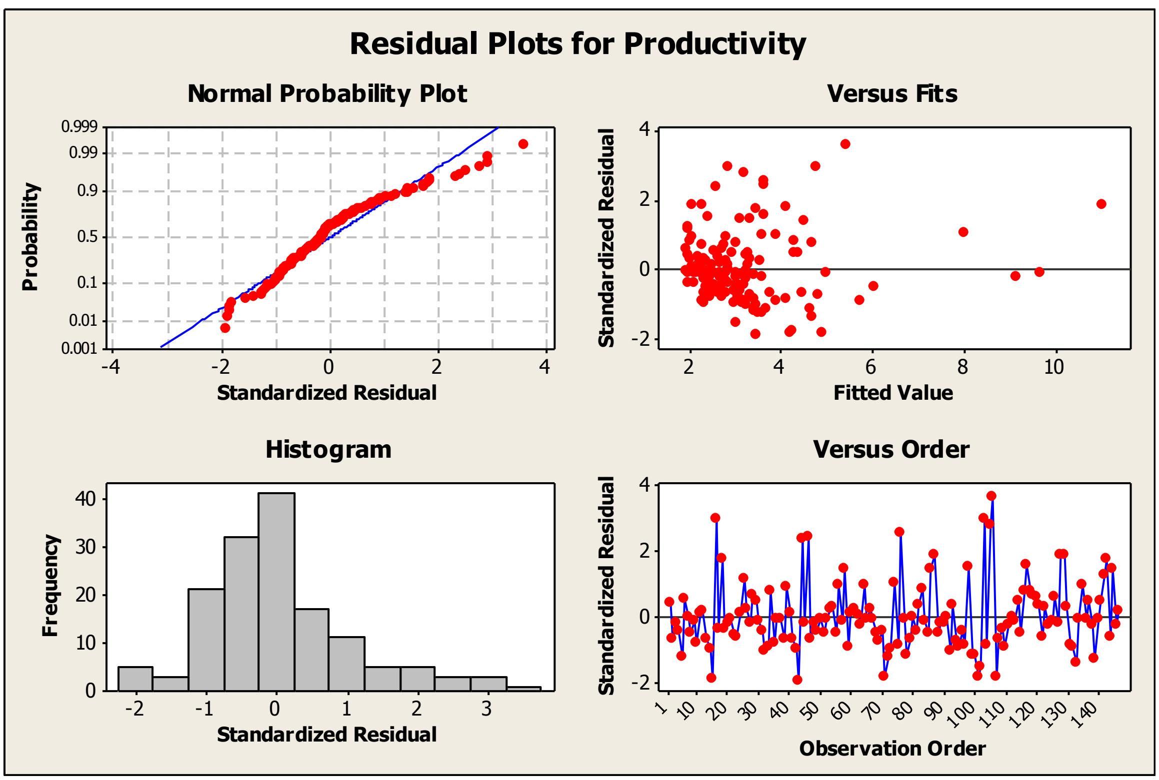
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	267.220	44.537	19.90	0.000
Residual Error	140	313.248	2.237		
Total	146	580.468			

Source	DF	Seq SS
NFDL_IR	1	145.897
SM	1	53.608
SH	1	20.964
USD_CTN.Y05	1	17.352
RL.Y04	1	14.525
SM.Y05	1	14.874

```
Unusual Observations

     Obs    NFDL_IR          PTVT      Fit   SE Fit   Residual   St Resid
      16        4.8         7.230    2.828    0.143      4.402      2.96R
      44        2.9         6.090    2.568    0.148      3.522      2.37R
      46       16.5         7.190    3.628    0.265      3.562      2.42R
      57       47.7         9.530    9.671    0.950     -0.141     -0.12 X
      59       13.7         4.460    5.729    0.570     -1.269     -0.92 X
      74       13.0         8.260    8.016    1.478      0.244      1.05 X
      76       15.6         7.340    3.608    0.241      3.732      2.53R
      87       38.4        12.100   11.063    1.388      1.037      1.86 X
      88       21.4         5.370    6.044    0.586     -0.674     -0.49 X
     103       14.2         8.950    4.769    0.453      4.181      2.93R
     105        0.0         7.230    3.172    0.358      4.058      2.79R
     106       11.4        10.520    5.408    0.471      5.112      3.60R
     146        0.0         9.000    9.135    1.380     -0.135     -0.23 X

     R denotes an observation with a large standardized residual.
     X denotes an observation whose X value gives it large leverage.
```

Residual Plots for PTVT



Iteration 4: Small/Very Small mines no. 106 & 16 out

Stepwise Regression: PTVT versus 42 independent variables

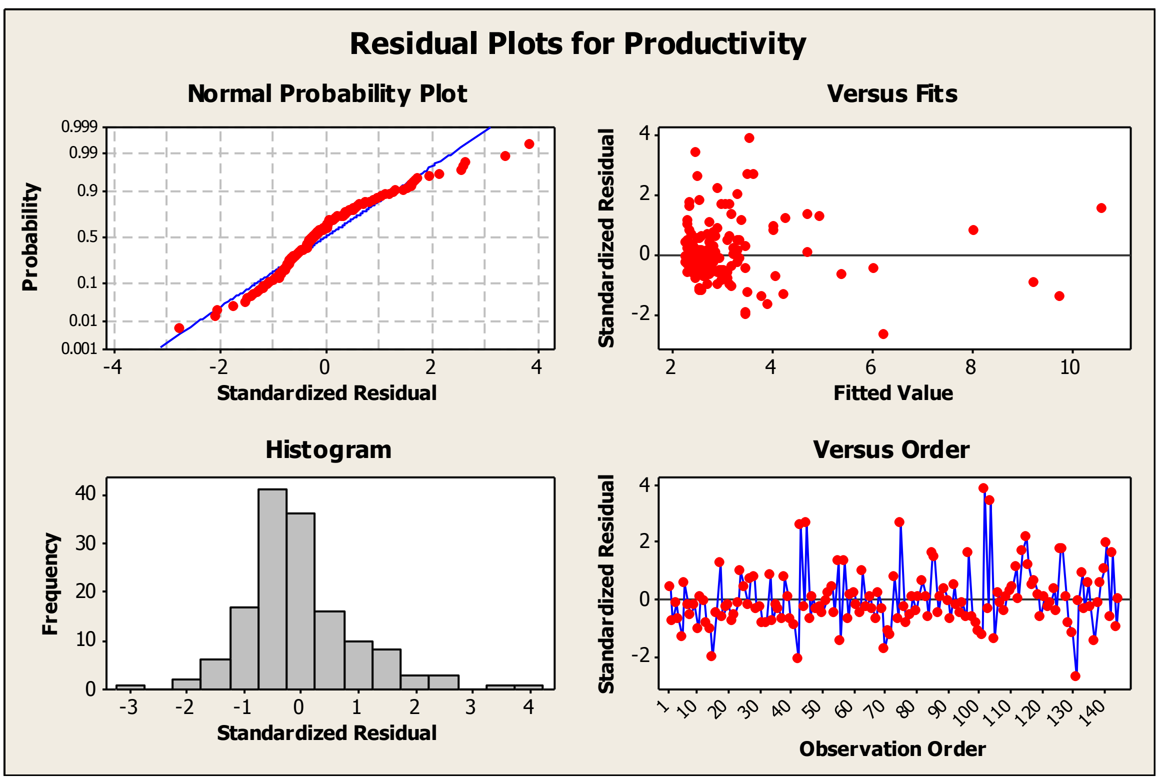
```
  Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05


Response is PTVT on 42 predictors, with N = 145


Step                    1        2        3          4          5
Constant            2.756    2.720    2.668      2.587      1.910

IR^2              0.00402  0.00303  0.00297    0.00442    0.00430
T-Value              7.46     5.30     5.37       6.03       6.00
P-Value             0.000    0.000    0.000      0.000      0.000

SM                         0.00019  0.00019    0.00022    0.00023
T-Value                       3.87     4.03       4.63       4.88
P-Value                      0.000    0.000      0.000      0.000

USD_CTN.Y05                         0.00084    0.00079    0.00078
T-Value                                3.34       3.23       3.26
P-Value                               0.001      0.002      0.001

SM.Y06                                        -0.00040   -0.00038
T-Value                                          -2.92      -2.86
P-Value                                          0.004      0.005

SH                                                         0.0132
T-Value                                                      2.74
P-Value                                                     0.007

S                    1.60     1.53     1.48       1.44       1.41
R-Sq                28.03    34.90    39.67      43.13      46.04
R-Sq(adj)           27.52    33.99    38.38      41.50      44.10
Mallows Cp           39.1     23.9     13.9        7.3        2.0
PRESS             403.267  456.802  434.872    574.093    631.311
R-Sq(pred)          20.87    10.36    14.67       0.00       0.00
```

Regression Analysis: PTVT versus 5 independent variables

```
The regression equation is
      PTVT = 1.91 + 0.00430 NFDL_IR^2 + 0.000229 SM + 0.000783 USD_CTN.Y05
            - 0.000383 SM.Y06 + 0.0132 SH


Predictor              Coef     SE Coef      T      P    VIF
Constant             1.9097      0.2787   6.85  0.000
NFDL_IR^2         0.0043034   0.0007169   6.00  0.000  2.301
SM               0.00022929  0.00004698   4.88  0.000  1.299
USD_CTN.Y05       0.0007835   0.0002401   3.26  0.001  1.005
SM.Y06           -0.0003828   0.0001339  -2.86  0.005  2.310
SH                 0.013216    0.004828   2.74  0.007  1.007


S = 1.40657   R-Sq = 46.0%   R-Sq(adj) = 44.1%

PRESS = 631.311   R-Sq(pred) = 0.00%


Analysis of Variance

Source           DF       SS      MS      F      P
Regression        5  234.608  46.922  23.72  0.000
Residual Error  139  275.001   1.978
Total           144  509.609


Source            DF   Seq SS
NFDL_IR^2          1  142.818
SM                 1   35.048
USD_CTN.Y05        1   24.277
SM.Y06             1   17.642
SH                 1   14.823
```

```
Unusual Observations

Obs  NFDL_IR^2       PTVT     Fit  SE Fit    Residual  St Resid
 15          0      0.740   3.496   0.356      -2.756    -2.02R
 28         15      5.020   4.024   0.521       0.996     0.76 X
 42          0      0.680   3.496   0.356      -2.816    -2.07R
 43          9      6.090   2.499   0.148       3.591     2.57R
 45        271      7.190   3.517   0.159       3.673     2.63R
 56       2273      9.530   9.802   1.394      -0.272    -1.44 X
 58        187      4.460   5.407   0.542      -0.947    -0.73 X
 73        169      8.260   8.084   1.389       0.176     0.79 X
 75        243      7.340   3.624   0.182       3.716     2.66R
 86       1477     12.100  10.658   1.013       1.442     1.48 X
 87        459      5.370   6.045   0.576      -0.675    -0.53 X
102        202      8.950   3.548   0.155       5.402     3.86R
104          0      7.230   2.465   0.136       4.765     3.40R
115         17      5.960   2.935   0.159       3.025     2.16R
116         65      5.790   4.268   0.517       1.522     1.16 X
131        822      2.700   6.254   0.555      -3.554    -2.75RX
144          0      9.000   9.257   1.382      -0.257    -0.98 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.
```

Residual Plots for PTVT



Iteration 5: Small/Very Small mines no. 104 & 102 out

Stepwise Regression: PTVT versus 42 independent variables

```
  Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05


Response is PTVT on 42 predictors, with N = 143


Step                 1        2        3        4         5
Constant         2.689    2.651    2.597    1.838     1.796

NFDL_IR^2      0.00399  0.00299  0.00292  0.00286   0.00418
T-Value           7.91     5.63     5.75     5.82      6.42
P-Value          0.000    0.000    0.000    0.000     0.000

SM                      0.00020  0.00020  0.00021   0.00023
T-Value                    4.27     4.49     4.84      5.47
P-Value                   0.000    0.000    0.000     0.000

USD_CTN.Y05                      0.00086  0.00085   0.00081
T-Value                             3.73     3.79      3.71
P-Value                            0.000    0.000     0.000

SH                                         0.0147    0.0141
T-Value                                      3.27      3.22
P-Value                                     0.001     0.002

SM.Y06                                             -0.00036
T-Value                                               -2.98
P-Value                                               0.003

S                 1.50     1.42     1.35     1.31      1.27
R-Sq             30.76    38.74    44.31    48.32     51.47
R-Sq(adj)        30.27    37.87    43.11    46.82     49.70
Mallows Cp        53.3     33.1     19.7     10.5       3.8
PRESS          351.311  389.991  372.342  332.723   623.068
R-Sq(pred)       23.28    14.84    18.69    27.34      0.00
```

Regression Analysis: PTVT versus 5 independent variables

```
The regression equation is
      PTVT = 1.80 + 0.00418 NFDL_IR^2 + 0.000233 SM
            + 0.000806 USD_CTN.Y05 - 0.000362 SM.Y06 + 0.0141 SH
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	1.7964	0.2533	7.09	0.000	
NFDL_IR^2	0.0041814	0.0006510	6.42	0.000	2.309
SM	0.00023261	0.00004255	5.47	0.000	1.299
USD_CTN.Y05	0.0008058	0.0002174	3.71	0.000	1.006
SM.Y06	-0.0003621	0.0001214	-2.98	0.003	2.317
SH	0.014109	0.004375	3.22	0.002	1.007

S = 1.27368   R-Sq = 51.5%   R-Sq(adj) = 49.7%

PRESS = 623.068   R-Sq(pred) = 0.00%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	235.692	47.138	29.06	0.000
Residual Error	137	222.249	1.622		
Total	142	457.941			

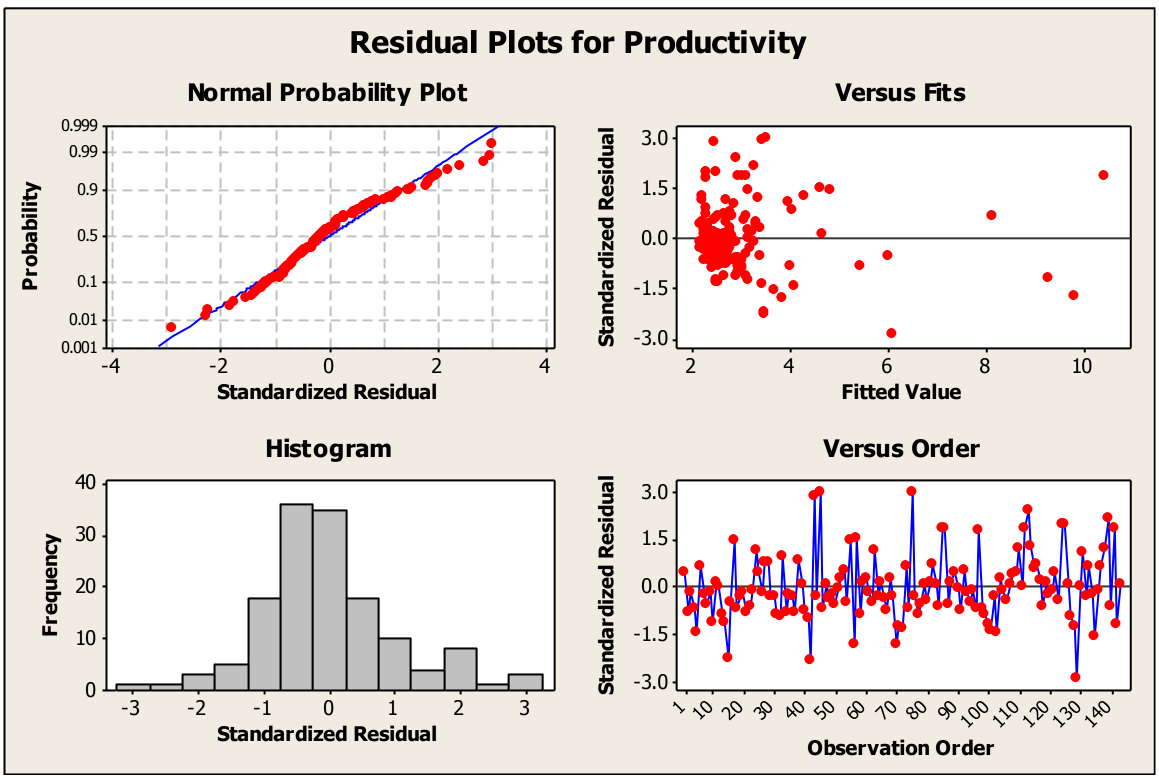
Source	DF	Seq SS
NFDL_IR^2	1	140.848
SM	1	36.571
USD_CTN.Y05	1	25.487
SM.Y06	1	15.915
SH	1	16.871

Unusual Observations

Obs	NFDL_IR^2	PTVT	Fit	SE Fit	Residual	St Resid
15	0	0.740	3.489	0.322	-2.749	-2.23R
28	15	5.020	4.050	0.472	0.970	0.82 X
42	0	0.680	3.489	0.322	-2.809	-2.28R
43	9	6.090	2.438	0.134	3.652	2.88R
45	271	7.190	3.426	0.145	3.764	2.97R
56	2273	9.530	9.831	1.262	-0.301	-1.77 X
58	187	4.460	5.427	0.491	-0.967	-0.82 X
73	169	8.260	8.126	1.257	0.134	0.66 X
75	243	7.340	3.521	0.166	3.819	3.02R
86	1477	12.100	10.430	0.920	1.670	1.90 X
87	459	5.370	6.016	0.522	-0.646	-0.56 X
113	17	5.960	2.884	0.145	3.076	2.43R
114	65	5.790	4.287	0.469	1.503	1.27 X
125	0	5.020	2.474	0.118	2.546	2.01R
129	822	2.700	6.081	0.505	-3.381	-2.89RX
139	33	6.060	3.282	0.181	2.778	2.20R
142	0	9.000	9.279	1.251	-0.279	-1.18 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for PTVT



Iteration 6: Small/Very Small mine no. 75 out

Stepwise Regression: PTVT versus 42 independent variables

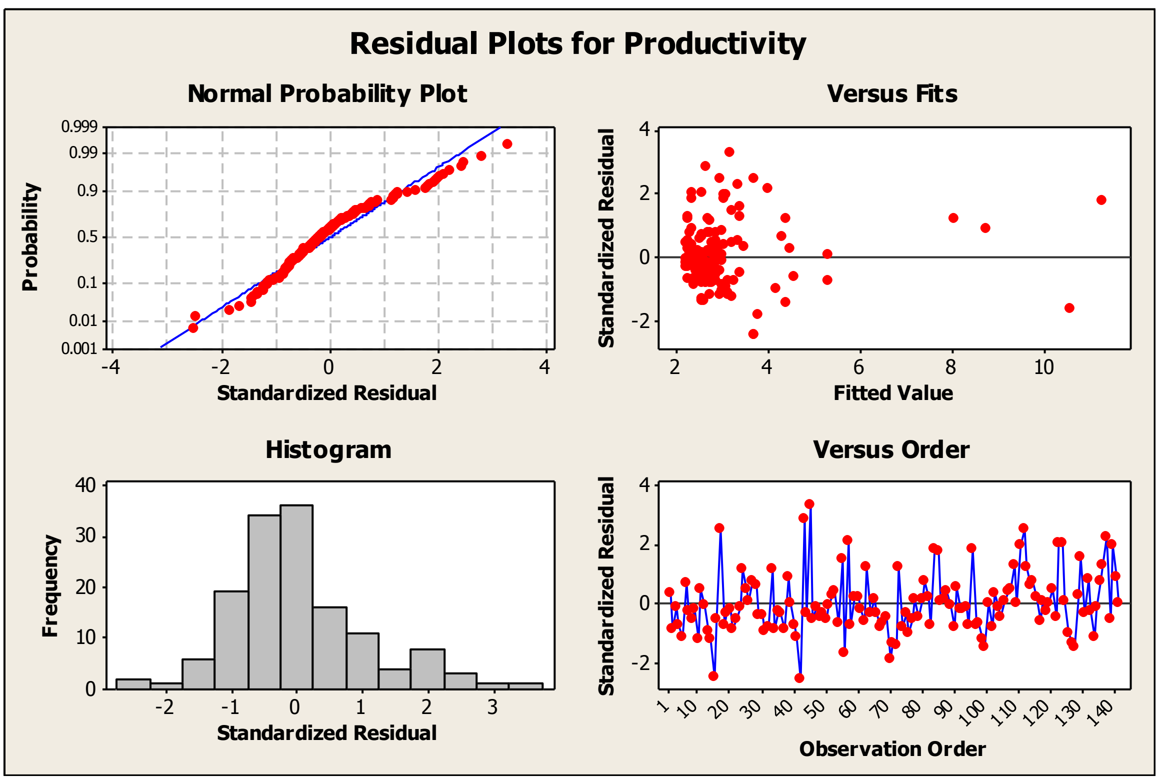
```
  Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05


Response is PTVT on 42 predictors, with N = 142


Step                  1        2        3        4        5
Constant          2.669    2.629    2.575    1.788    1.757

NFDL_IR^2       0.00393  0.00289  0.00283  0.00276  0.00209
T-Value            7.92     5.58     5.71     5.80     4.11
P-Value           0.000    0.000    0.000    0.000    0.000

SM                       0.00020  0.00020  0.00021  0.00021
T-Value                     4.49     4.73     5.14     5.28
P-Value                    0.000    0.000    0.000    0.000

USD_CTN.Y05                       0.00086  0.00085  0.00075
T-Value                              3.83     3.92     3.53
P-Value                             0.000    0.000    0.001

SH                                          0.0152   0.0159
T-Value                                       3.50     3.77
P-Value                                      0.001    0.000

SM.Y05                                              0.00053
T-Value                                                3.12
P-Value                                               0.002

S                  1.47     1.38     1.32     1.27     1.23
R-Sq              30.94    39.68    45.49    49.96    53.31
R-Sq(adj)         30.45    38.82    44.31    48.50    51.59
Mallows Cp         67.5     43.5     28.2     16.9      9.0
PRESS           335.366  347.476  324.051  303.581  372.767
R-Sq(pred)        23.67    20.91    26.25    30.91    15.16
```

Regression Analysis: PTVT versus 5 independent variables

```
The regression equation is
      PTVT = 1.76 + 0.00209 NFDL_IR^2 + 0.000213 SM
            + 0.000747 USD_CTN.Y05 + 0.000531 SM.Y05 + 0.0159 SH


Predictor              Coef     SE Coef     T      P    VIF
Constant             1.7571      0.2445  7.19  0.000
NFDL_IR^2         0.0020942   0.0005097  4.11  0.000  1.518
SM               0.00021266  0.00004027  5.28  0.000  1.251
USD_CTN.Y05       0.0007471   0.0002116  3.53  0.001  1.025
SM.Y05            0.0005312   0.0001701  3.12  0.002  1.303
SH                 0.015935    0.004223  3.77  0.000  1.008


S = 1.22818   R-Sq = 53.3%   R-Sq(adj) = 51.6%

PRESS = 372.767   R-Sq(pred) = 15.16%


Analysis of Variance

Source           DF       SS      MS      F      P
Regression        5  234.224  46.845  31.06  0.000
Residual Error  136  205.146   1.508
Total           141  439.370


Source            DF   Seq SS
NFDL_IR^2          1  135.959
SM                 1   38.404
USD_CTN.Y05        1   25.513
SM.Y05             1   12.875
SH                 1   21.472
```

```
Unusual Observations

       Obs  NFDL_IR^2       PTVT     Fit  SE Fit  Residual  St Resid
        15          0      0.740   3.669   0.309    -2.929    -2.46R
        17        574      6.650   3.673   0.288     2.977     2.49R
        28         15      5.020   4.288   0.454     0.732     0.64 X
        42          0      0.680   3.669   0.309    -2.989    -2.51R
        43          9      6.090   2.618   0.111     3.472     2.84R
        45        271      7.190   3.154   0.152     4.036     3.31R
        56       2273      9.530  10.573   1.049    -1.043    -1.63 X
        57        453      6.550   3.992   0.232     2.558     2.12R
        58        187      4.460   5.298   0.473    -0.838    -0.74 X
        73        169      8.260   8.025   1.213     0.235     1.22 X
        85       1477     12.100  11.288   1.141     0.812     1.78 X
        86        459      5.370   5.296   0.484     0.074     0.07 X
       112         17      5.960   2.940   0.138     3.020     2.47R
       113         65      5.790   4.395   0.452     1.395     1.22 X
       123          0      4.780   2.315   0.131     2.465     2.02R
       124          0      5.020   2.522   0.112     2.498     2.04R
       138         33      6.060   3.314   0.175     2.746     2.26R
       141          0      9.000   8.726   1.188     0.274     0.88 X

       R denotes an observation with a large standardized residual.
       X denotes an observation whose X value gives it large leverage.
```

Residual Plots for PTVT



Iteration 7: Small/Very Small mine no. 45 out

Stepwise Regression: PTVT versus 42 independent variables

```
  Alpha-to-Enter: 0.05  Alpha-to-Remove: 0.05


Response is PTVT on 42 predictors, with N = 141


Step                    1        2        3        4        5
Constant            2.650    2.609    2.554    1.742    1.707

NFDL_IR^2         0.00386  0.00280  0.00273  0.00266  0.00195
T-Value              7.89     5.50     5.65     5.75     3.97
P-Value             0.000    0.000    0.000    0.000    0.000

SM                         0.00021  0.00021  0.00022  0.00022
T-Value                       4.66     4.93     5.39     5.59
P-Value                      0.000    0.000    0.000    0.000

USD_CTN.Y05                         0.00087  0.00086  0.00075
T-Value                                3.98     4.09     3.69
P-Value                               0.000    0.000    0.000

SH                                            0.0157   0.0165
T-Value                                         3.72     4.05
P-Value                                        0.000    0.000

SM.Y05                                                0.00056
T-Value                                                  3.41
P-Value                                                 0.001

S                    1.45     1.35     1.28     1.23     1.18
R-Sq                30.93    40.32    46.50    51.44    55.29
R-Sq(adj)           30.44    39.46    45.33    50.02    53.63
Mallows Cp           68.1     42.2     25.9     13.2      3.8
PRESS             321.100  319.289  297.050  271.703  332.396
R-Sq(pred)          23.88    24.31    29.58    35.59    21.20
```

Regression Analysis: PTVT versus 5 independent variables

```
The regression equation is
      PTVT = 1.71 + 0.00195 NFDL_IR^2 + 0.000217 SM + 0.000751 USD_CTN.Y05
            + 0.000558 SM.Y05 + 0.0165 SH
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	1.7069	0.2358	7.24	0.000	
NFDL_IR^2	0.0019533	0.0004922	3.97	0.000	1.523
SM	0.00021659	0.00003877	5.59	0.000	1.252
USD_CTN.Y05	0.0007514	0.0002037	3.69	0.000	1.025
SM.Y05	0.0005584	0.0001639	3.41	0.001	1.305
SH	0.016472	0.004067	4.05	0.000	1.008

S = 1.18197   R-Sq = 55.3%   R-Sq(adj) = 53.6%

PRESS = 332.396   R-Sq(pred) = 21.20%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	233.216	46.643	33.39	0.000
Residual Error	135	188.603	1.397		
Total	140	421.819			

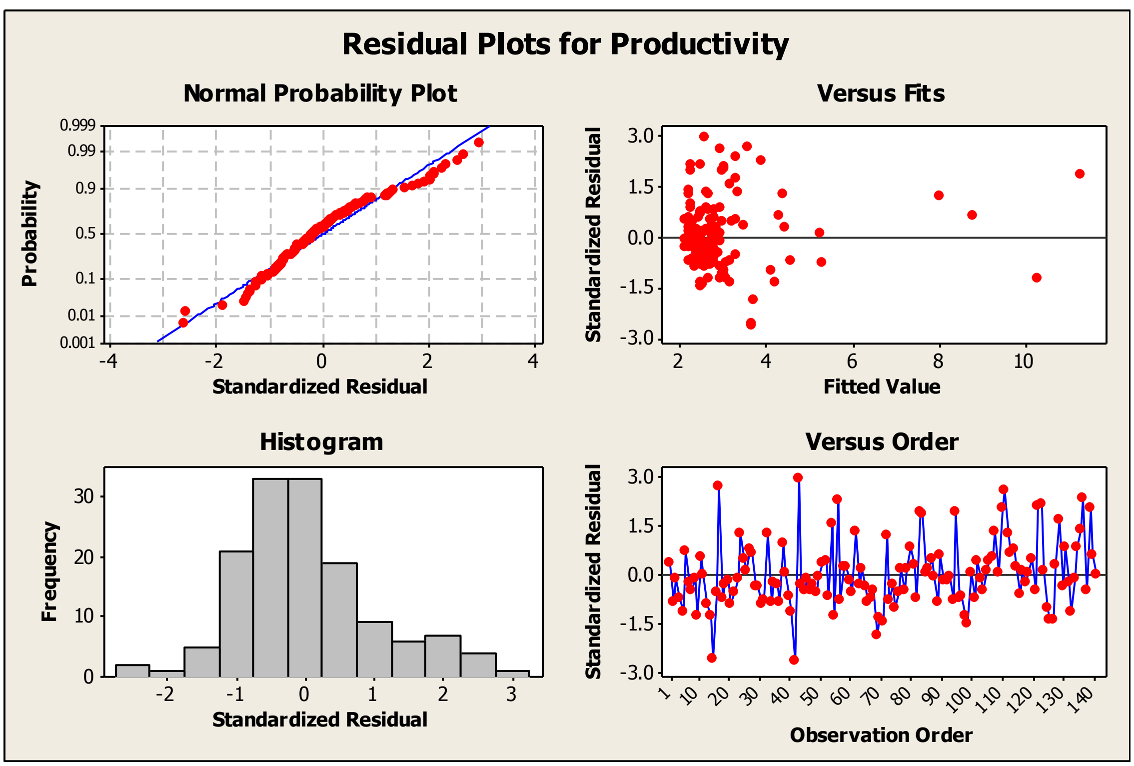
Source	DF	Seq SS
NFDL_IR^2	1	130.481
SM	1	39.618
USD_CTN.Y05	1	26.041
SM.Y05	1	14.165
SH	1	22.912

Unusual Observations

Obs	NFDL_IR^2	PTVT	Fit	SE Fit	Residual	St Resid
15	0	0.7400	3.6836	0.2975	-2.9436	-2.57R
17	574	6.6500	3.5659	0.2788	3.0841	2.69R
28	15	5.0200	4.3201	0.4367	0.6999	0.64 X
42	0	0.6800	3.6836	0.2975	-3.0036	-2.63R
43	9	6.0900	2.5938	0.1070	3.4962	2.97R
55	2273	9.5300	10.2876	1.0130	-0.7576	-1.24 X
56	453	6.5500	3.9191	0.2246	2.6309	2.27R
57	187	4.4600	5.3123	0.4550	-0.8523	-0.78 X
72	169	8.2600	8.0344	1.1674	0.2256	1.22 X
84	1477	12.1000	11.2910	1.0976	0.8090	1.84 X
85	459	5.3700	5.2667	0.4658	0.1033	0.10 X
110	28	5.4100	3.0153	0.1202	2.3947	2.04R
111	17	5.9600	2.9266	0.1331	3.0334	2.58R
112	65	5.7900	4.4195	0.4348	1.3705	1.25 X
122	0	4.7800	2.2834	0.1263	2.4966	2.12R
123	0	5.0200	2.4976	0.1075	2.5224	2.14R
137	33	6.0600	3.3076	0.1680	2.7524	2.35R
139	64	5.4500	3.0282	0.1167	2.4218	2.06R
140	0	9.0000	8.8146	1.1435	0.1854	0.62 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for PTVT



Correlation matrix

	PTVT	NFDL_IR^2	SM	USD_CTN.Y05	SM.Y05
NFDL_IR^2	0.556				
SM	0.522	0.445			
USD_CTN.Y05	0.266	0.035	0.010		
SM.Y05	0.453	0.463	0.211	0.148	
SH	0.213	0.014	-0.054	0.019	-0.041

Cell Contents: Pearson correlation