Predictive Analytics with Administrative Data from the Mine Safety and Health Administration

Yuwen Dai, Natalie Patten, Albert Lee, George Cave[1]

Summit Consulting, LLC
601 New Jersey Avenue, NW
Suite 400
Washington, DC 20001

Abstract

The U.S. Department of Labor (DOL), Mine Safety and Health Administration (MSHA) would like to predict at an early stage which mine operating firms with health and safety violations are at risk of failing to pay their fines on time. Summit Consulting, LLC (Summit), in collaboration with the DOL Chief Evaluation Office (CEO), used MSHA internal administrative data to develop predictive models that identify mine operators who are at high risk of failing to make timely payments to MSHA. Summit also created a dynamic, user-friendly Microsoft Excel tool (the Early Detection tool) based on the delinquency risk scores calculated by the predictive models.

The early identification of delinquent operators has three components: 1) identification of the probability to be delinquent, and 2) the severity of delinquency, given delinquency, in terms of duration, and 3) the compounding effect of delinquency probability and severity. Summit used logistic regression and survival analysis to predict the delinquency probability and duration, respectively. We developed a composite risk score to rank the delinquency risk of violations based on these two models. The early detection model has demonstrated predictive power to discriminate operators with high delinquency risk and has uncovered key relationships in the delinquency behavior in a statistical and systematic way. Summit implemented the early detection model in a Microsoft Excel application so that MSHA can periodically use in-house resources to conduct the analysis and use the results for early enforcement activities. The automated tool requires minimum inputs and operations from users and has a user-friendly, dynamic interface.

In this paper, we demonstrate that predictive analytics can increase MSHA's internal capacity for conducting and implementing program-enhancing data analytics. This serves as a proof of concept for using predictive analytics to improve agency performance.

Key Words: Predictive analytics, Administrative Data, Business Intelligence tool, Enforcement program-enhancing data analytics

Introduction

Under the Federal Mine Safety and Health Act of 1977 (Mine Act), the Mine Safety and Health Administration (MSHA) is entrusted to enforce compliance with mandatory safety and health standards at all mining and mineral processing operations in the United States. MSHA conducts regular inspections four times per year in each underground mine and twice per year in each surface mine, and may conduct spot inspections for locations

---

[1] Dai, Patten, Lee, and Cave: Summit Consulting, LLC.
Disclaimer: This article is the result of the authors' independent research and does not represent the views of Summit Consulting or the U.S. Department of Labor the Mine Safety and Health Administration.

warranting more intense scrutiny or in reaction to individual incidents and complaints [1-4]. MSHA assesses a civil penalty for every health and safety violation but is unable to collect about 10% of penalty fines.

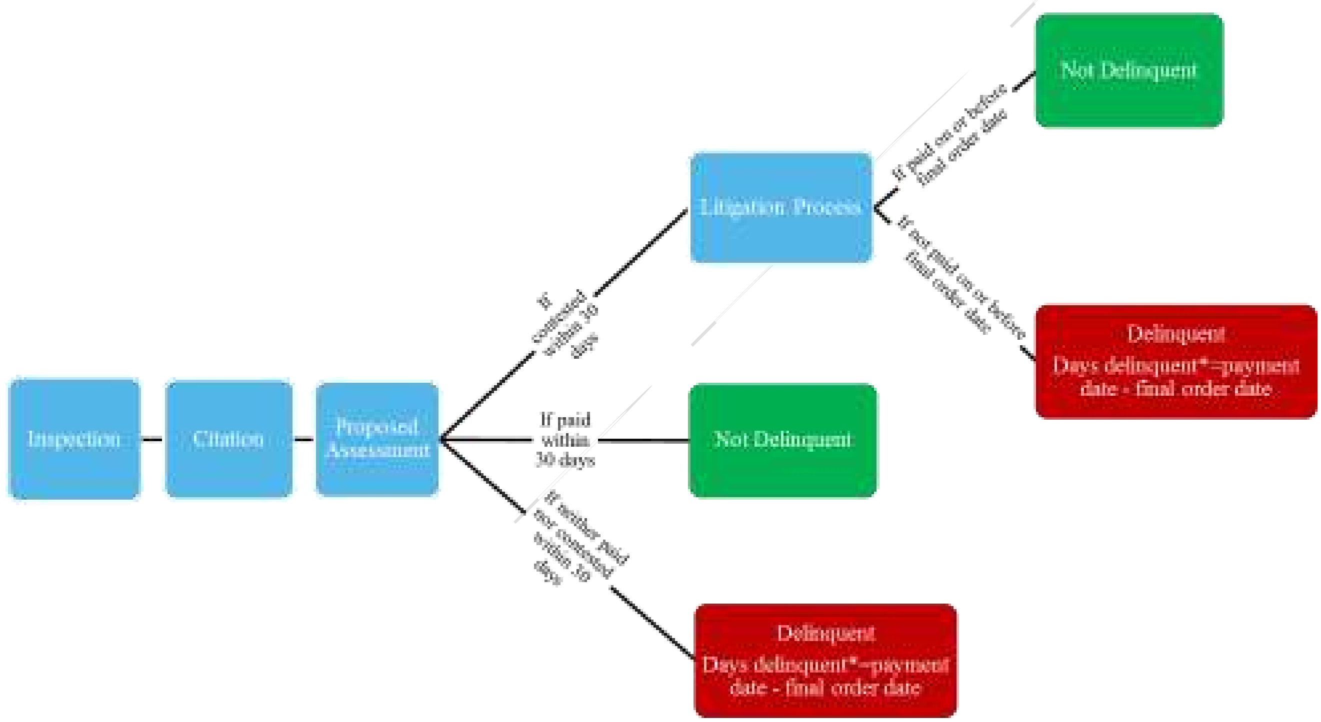
The Mine Improvement and New Emergency Response Act of 2006 (MINER Act) dictates the minimum and maximum penalty amounts for certain types of violations according to the following characteristics:

- Size of the business of the operator charged,
- History of previous violations,
- Repeat violations of the same standard,
- Negligence,
- Gravity of the safety and health hazard(s),
- Demonstrated good faith of the operator in abating the violation, and
- The effect of the penalty on the operator's ability to continue in business.

MSHA calculates penalty points for each characteristic and uses a conversion table that determines the penalty amount according to the total number of points associated with a violation. "Regular" assessments, which MSHA levies for most violations, range between $100 and $70,000. MSHA imposes "special" assessments for violations of extraordinary circumstances, which are not listed in the conversion table and can be valued at over $70,000.

The mine operator may request a safety and health conference to review the citations and orders. After the conference, or if a conference is not requested, MSHA provides a form that lists the assessments for the violations, the instructions for paying or contesting them, and the payment due date, which is called the final order date. The penalty is typically due 30 days after the assessment. However, a firm's successful contest of the violation or penalty postpones the final order date. MSHA considers fines that firms do not pay by the final order date delinquent. Figure 1 shows the process of fine delinquency or payment.

Figure 1: MSHA Assessment and Delinquency Process



* For this analysis, the number of days delinquent is calculated as the data extraction date (in this case September 11, 2014) minus the final order date, if the penalty remains unpaid at the time of extraction.

When a fine is delinquent for 45 days, MSHA sends the delinquent firm a first demand letter. MSHA follows up with a second demand letter if the firm remains delinquent. If the mine operator has not paid the fine within 120 days of the assessment (i.e. within 90 days of the final order date) and has not contested it, MSHA sends the debt to the Treasury Department (Treasury) for collection.[2]

If the fine remains unpaid, Treasury sends a demand letter 30 days after receiving the delinquency notice from MSHA. After 90 days, Treasury sends the debt to a private collection agency. If the debt remains unpaid, Treasury sends it to a second private collection agency, and notifies the credit bureaus. The debt then enters the Treasury Offset Program, where any payments owed to the company, such as tax refunds, will go to Treasury. Finally, Treasury may choose to send the debt to the Department of Justice.

Currently, MSHA uses the Scofflaw Program as an enforcement tool to pursue operators demonstrating egregious non-compliance in combination with delinquent civil penalties [5]. This program leverages legal action to prevent scofflaw operators from ignoring monetary penalties and to incentivize operators to improve mines' safety and health conditions. In the past three years, MSHA has brought nine scofflaw operators to federal district court and received judgments totaling over $3 million in penalties [6].

[2] After 120 days, MSHA's computer system automatically marks delinquencies. Those that are marked are sent to Treasury once a week. MSHA may send delinquencies before or after the 120 day mark in certain cases.

However, this enforcement takes place *after* the non-compliance and non-payment has already occurred. To implement early intervention programs and tools that increase timely payment of civil penalties, MSHA must predict mine owners' propensity for long term delinquencies at the time a violation occurs. Thus, for MSHA, the critical step to the success of the early intervention program is detecting at an early stage which operators with violations are at risk of failing to pay their fines in a timely manner *before* they appear on the list of operators and contractors with delinquent penalties. Predicting the propensity of long delinquency for a violation upon its issuance allows MSHA to implement enforcement tools that increase timely payment of civil penalties.

The early identification of delinquent operators has three components: 1) identification of the probability to be delinquent, 2) the severity of delinquency, given delinquency, in terms of duration, and 3) the compounding effect of delinquency probability and severity. In collaboration with DOL CEO and MSHA, Summit developed three research questions to address each of the three components: *1) What is the predicted probability that a mine operator will be at least 90 days delinquent? 2) Given that an operator is delinquent, what is the predicted duration of delinquency? 3) Which operators and contractors are at highest risk of long-term delinquency?* In the next three sections, we first discuss the administrative data used in this paper, then discuss the predictive model methodologies and model results, and conclude by briefly describing the applications of the predictive model to MSHA and the automated tool implemented in Microsoft Excel.

Data

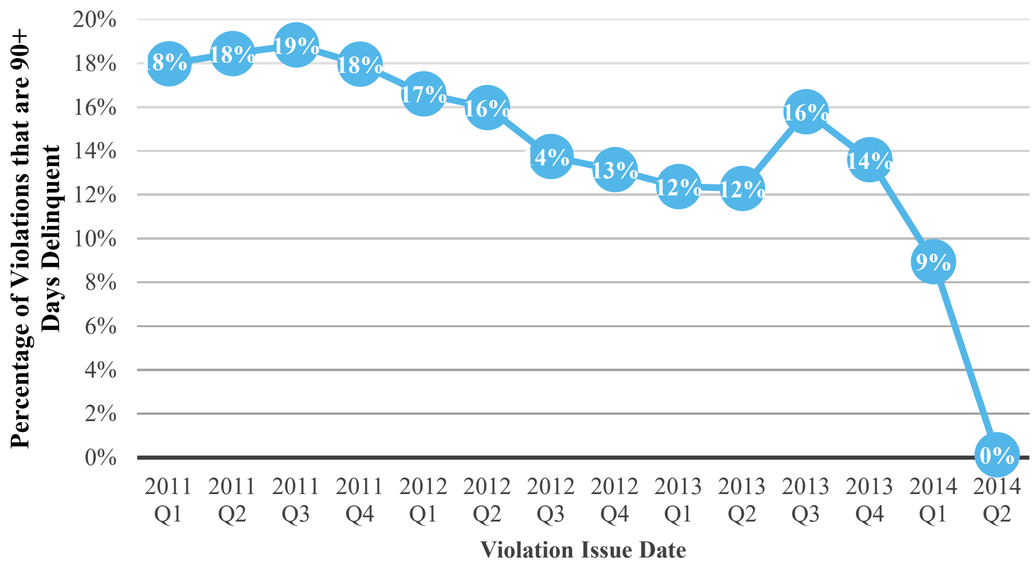
We used historical administrative data from the MSHA Standardized Information System (MSIS). These data contain information on mine characteristics, mine status history, mine operator history, mine employment and production information, accidents and injuries information, violation assessment and penalty information, payment amounts and dates, and delinquent case status. In this study, the unit of analysis was a unique violation and we developed the predictive models at violation level. We used the violation-level data issued from January 1, 2011 to September 11, 2014.

The outcome variables of interest were 1) a binary 90-day delinquency indicator and 2) the number of days in delinquency. Summit constructed both variables using the available data.

90-Day Delinquency Indicator: To address the first research question, we constructed the outcome variable of a delinquency indicator variable which takes a value of 1 if the penalty associated with a violation is at least 90 days delinquent and 0 otherwise. A delinquency occurs if the mine operator makes the penalty payment after the final due date. A 90-day delinquency occurs if the actual penalty payment date is greater than 90 days past the final due date, which includes cases whose penalty is still not paid as of our data cutoff date, September 11, 2014.

Figure 2 shows the unconditional delinquency rate for each year of the data from January 1, 2011 to June 30, 2014. The delinquency rate is defined as the number of violations that were 90+ days delinquent in the quarter divided by the number of total violations in the quarter.

Figure 2: Rate of 90-Day Delinquencies by Year and Quarter



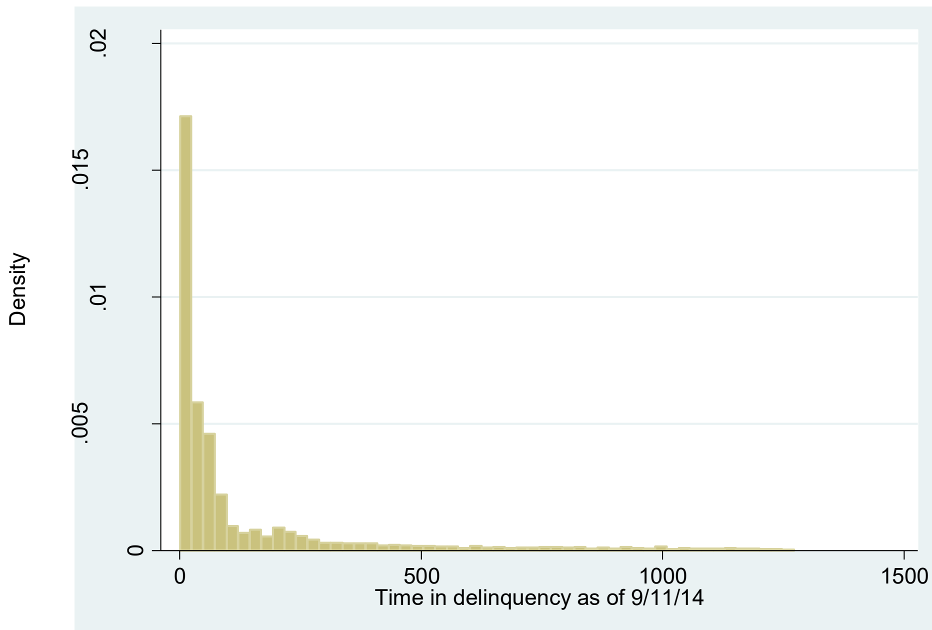
Source: MSIS Administrative Data Files, January 1, 2011-June 30, 2014

The figure shows that the percentage of violations that were delinquent for at least 90 days decreased each year from 2011 to 2014. We would expect an overall decreasing trend across years, especially a significant drop in 2014. The MSHA data extract provides a snapshot of data as of September 11, 2014. It includes the delinquency history performance observed on and before September 11, 2014, and leaves the payment behavior on all the open violations unobserved. The data is right-censored, which means that the event of interest (90-day delinquency) happens after the time of observation (September 11, 2014) and is not collected in the observed data. The longer the observation period, the more likely the 90-day delinquency is observed. The violations issued in early years (e.g., year 2011) have longer observation periods, and more delinquencies are observed in the data. The violations issued in more recent years (e.g., 2014) have a very short observation time period and have fewer delinquencies (close to 0%) observed. Some of those violations have not passed their payment due date by September 11, 2014 and most of them are still less than 90 days delinquent by September 11, 2014. Therefore, to avoid the right censoring issue in analyzing the 90 + days delinquency behavior, we used the violations issued between January 2011 and December 2013 in the first research question.

Time in Delinquency: To answer the second research question, we also created a new variable to indicate how long a payment is delinquent by taking the difference between the penalty due date and the actual payment date. We changed the time delinquent variable to zero if it had a negative value. A negative value would represent only those fines that firms paid before delinquency, so they have a delinquency duration of 0 days. This variable indicates how long a fine was delinquent before the violator paid it and how long a fine has been delinquent if it remains unpaid.

Figure 3 shows the distribution of time in delinquency for all the delinquent violations between January 2011 and December 2013, the same data as used in the research question 1. In the past, MSHA has collected about 90% of total penalties and about 10% have remained unpaid and in delinquency for more than six quarters.

Figure 3: Time in Delinquency Distribution, 2011 Q1 through 2013 Q4



Source: MSIS Administrative Data Files, January 1, 2011-December 31, 2013

Methodologies and Results

In this section, we describe the analytical approach used to address each of the three research questions. First, we discuss the justification for using the methodology. Second, we describe the process we used to specify the model and provide a brief description of the model specifications. Finally, we display the results and assess the model performance.

Research Question 1. What is the predicted probability that a penalty associated with a violation will be at least 90 days delinquent?

The purpose of this question is to predict whether a penalty for a given violation will be at least 90 days delinquent. After a penalty has been delinquent for 90 days, MSHA sends the debt to Treasury. If the operator has over $5,000 in delinquent penalties at that time, MSHA also places the operator on the scofflaw operators list. To predict which mines and violators will be delinquent at this point, we used 90 days in delinquency as the cut point for this model.

Binary logistic regression models the probability of an observation falling into one of the two outcomes. The probabilities can be used to predict the category in which the observation will fall. We used logistic regression to predict the binary outcome on 90-day delinquency.[3] We also determined which violation, mine, and violator characteristics are statistically significant in relation to 90-day delinquency.

This model assumes that violation i's probability of 90-day delinquency, p_i, is related to its observable characteristics through the following non-linear logistic function:

$$p_i = \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)} \tag{1}$$

[3] The variable equals one if a violation has been delinquent at least 90 days, and it equals zero otherwise.

X_i is a vector of violator, mine, and violation characteristics and β is a vector of regression coefficients that must be estimated. We estimated this model on all violations that were either paid or at least 90 days delinquent within the observed time period. Unpaid violations that have not yet reached 90 days after their final order date are not used to estimate the model because the dependent variable is unknown. As described in the data section, we used the violations issued between January 2011 and December 2013 for the logistic regression estimation, which includes 272,843 observations.

After compiling a comprehensive list of variables taken or derived from the MSIS database, we first used decision tree (CART) data exploratory analysis to screen a list of candidate variables. We used over 100 variables from the MSHA public use data, and from this exercise identified about 40 variables for potential use in the model. We began with the model containing all the candidate variables included in the MSIS data that, based on conversations with MSHA and our exploratory data analysis, could be associated with the dependent variable, the 90-day delinquency binary indicator. We used the stepwise model selection method, which is a hybrid of backward and forward selection, in which we remove variables with p-values greater than 0.20. We individually reintroduced the removed variables into the model and tested for significance. We kept significant variables and continued adding the other removed variables individually.

High correlation among predictor variables can cause unreliable and unstable coefficients estimation. We used the variance inflation factor (VIF) to test for multicollinearity on all the remaining predictor variables, except for the interaction term. The VIF shows how much multicollinearity inflates the variance of each coefficient. Although VIF information can only be obtained using OLS regression, we can compute it using an OLS regression with the same dependent (90-day delinquency) and predictor variables as the logistic regression. Typically, a VIF over 10 indicates that a variable should be dropped. **Error! Reference source not found.** shows the VIF for each predictor variable. We did not observe any predictor variable with VIF higher than 10, so none of the variables selected from the stepwise selection process are eliminated. In the end, we included 14 covariates and one interaction term in our final model. Table 2 displays the final list of covariates and their coefficients' sign and the p-value of Wald's test. We rank the covariates by the absolute value of their coefficients, while the upward arrow means positive coefficient and downward arrow means negative coefficient.

Table 1. Brief Covariate Specification for Logistic Regression

Rank	Variable	Sign	p-value
1	Percent of violations 90-days delinquent in prior year	+	0.00
2	Indicator of 90-day delinquency in prior year	+	0.00
3	Canvass Code	Mixed	N/A
4	Company Type	Mixed	N/A
5	Violation Type	Mixed	N/A
6	District	Mixed	N/A
7	Log of Penalty Amount	+	0.00
8	Interaction: (% of Violations 90-days Delinquent) x (Log Sum of Delinquent Penalties in Prior Year)	+	0.00
9	Rate of Injuries to Violations in Prior Year	+	0.05
10	Mine Size Points	-	0
11	Violator Type (Contractor)	-	0.25
12	Mine Type	Mixed	N/A
13	Log Number of employees	-	0.00
14	Log Number of Violations in Prior Year	-	0.00
15	Log Sum of Delinquent Penalties in Prior Year	-	0.00

To improve the enforcement efficiency and better deploy limited enforcement resources to more targeted areas, it is important for MSHA to understand the top risk drivers to the 90+ days delinquency for the entire portfolio of all MSHA violations as well as each individual violation in their portfolio. The coefficients in the logistic regression can provide some insight on how much each covariate impacts an increased risk of the 90+ days delinquency, but does not consider the difference in units between covariates and cannot provide relative importance between covariates. To quantify and rank order the contribution of a predictor variable to the increase in the risk of 90+ day delinquency, we have developed a method using the following formula:

$$Contribution\ of\ x_i = (x_i\beta_i \,/ \sum_{j=i}^{N}(x_i\beta_i)\big|x_j\beta_j > 0),$$

where x_i is the predictor variable i, β_i is the coefficient estimate of i, and $\sum_{j=i}^{N}(x_i\beta_i)\big|x_j\beta_j > 0$ is the sum of all positive $x\beta$ values. This metric allows us to focus only on the covariates that increase the 90+ day delinquency risk and see which explanatory variables make up the largest share of the positive portion of the likelihood of 90+ day delinquency.

Among the 15 covariates in the logistic regression, Table 2 shows the contribution for the top five contributors for delinquency risk increase. On average, the top contributor is the percentage of violations that were delinquent in the prior year, which accounted for 20% of the positive contributions in the model. The top three contributors accounted for 47% of the positive $\boldsymbol{x\beta}$ values, and the top five contributors accounted for 62%. This analysis provides MSHA insights on areas where they can focus their enforcement efforts given limited resources.

Table 2: Top Five Positive Contributors to 90-Day Delinquency

Characteristic	Percent of Positive Contribution
Percent of violations delinquent in prior year	20%
Indicator of 90+ day delinquency in prior year	15%
Primary Canvass - Coal Anthracite	11%
District - Coal 7	8%
District - Coal 12	8%
Top Three Contributors	47%
Top Five Contributors	62%

This method can also be used to identify the top risk contributors for each individual violation delinquency. This is particularly useful for extreme cases whose true top risk contributors are not well represented in the sample data population (results are not discussed in this paper). Often time, the extreme cases are the focus area MSHA can better deploy their enforcement resources.

We used k-fold cross validation, where k=5, and calculated the average area under the Receiver Operating Characteristic (ROC) curve to test model performance. K-fold cross validation is a model validation technique for 1) assessing the predictive performance of a statistical model to a data set outside the model estimation (or training) dataset and 2) measuring the stability of the model prediction to prevent over-fitting (adding extra parameters that damage the model's prediction when used on data that is not used to train it). We randomly assigned the violators into five groups so that all of the violations associated with a violator are in the same group. For each fold, we trained the model on the other four groups and tested the model on the kth group. We calculated the area under the

ROC curve for each fold and averaged the five ROC statistics. The area under the ROC curve for the logistic model is 0.81 for in-sample and 0.79 for out-of-sample data, which tells us the binary logistic model we estimated has very good and stable predictive power.

Research Question 2. What is the predicted duration of delinquency for a violation?

The purpose of this question is to predict duration of delinquency for a given delinquent violation. We used a parametric survival model to determine which characteristics of delinquent violations and their mines and operators are associated with longer delinquency durations and predict the median time to payment on new delinquent violations. Survival models estimate the expected length of time until an event will occur. A useful feature of this model is that it accounts for right censoring, or those operators that never pay off their fines during the period of observation. Using right-censored data in other models, like linear regression, can lead to biased estimates [7].

The survivor model specifies violation i's time until payment as a non-linear function of its characteristics:

$$t_i = \exp(X_i'\beta + \sigma\varepsilon_i), \tag{2}$$

t_i is time until payment occurs, X_i is a vector of violator, mine, and violation characteristics, β is a vector of regression coefficients, σ is a scale parameter, and ε_i is a random error term that comes from an unspecified distribution (typically non-normal). Survival analysis methods can be used to estimate β based on the observed data.

Violation i's survivor function is

$$1 - \mathrm{F}(t_i|X_i), \tag{3}$$

where, $\mathrm{F}(t_i|X_i)$ is the probability that a firm pays off its delinquent debt by time t_i. After estimating the survivor function, we can predict the estimated median time to payment for each violation.

The disturbance term (ε) in the MSHA data follows a log-logistic distribution[4], and violation i has characteristics X_i. Thus the estimated probability that the violation will be paid off by time t_i is:

$$S(t_i|X_i) = \frac{1}{1+(te^{-X_i\beta})^{1/\sigma}}, \tag{4}$$

and the estimated median time to repayment is:

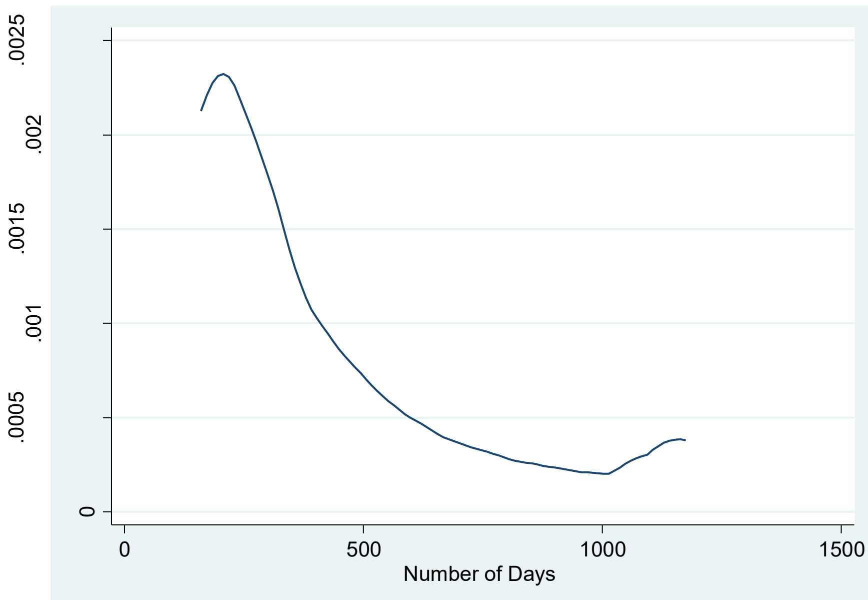
$$t_{med} = \min(t_i|S(t_i|X_i) \leq 0.5). \tag{5}$$

In this portion of the analysis, we also used the violations issued from 2011 Q1 through 2013 Q4, but only kept the ones with delinquent penalty payment more than 90 days, which includes 40,197 observations.

We first analyzed the hazard curve on the data. Figure 4 shows the hazard curve from 2011 Q1 through 2013 Q4. The hazard is the probability that a firm will pay for a violation at a given number of days, given that the violation is at least 90 days delinquent. Starting at the 90th day of delinquency, a positive trend occurs, meaning that with each day after the 90th day of delinquency, a firm is more likely to pay its penalty. After about 250 days in delinquency, the hazard rate declines until about 1,000 days in delinquency, meaning that penalties are less likely to be paid with each day. After about 1,000 days, the hazard increases again. Most likely this is not due to the delinquency behavior changing at the extreme tail, but rather the thinning of data which makes it unrepresentative of the true pattern.

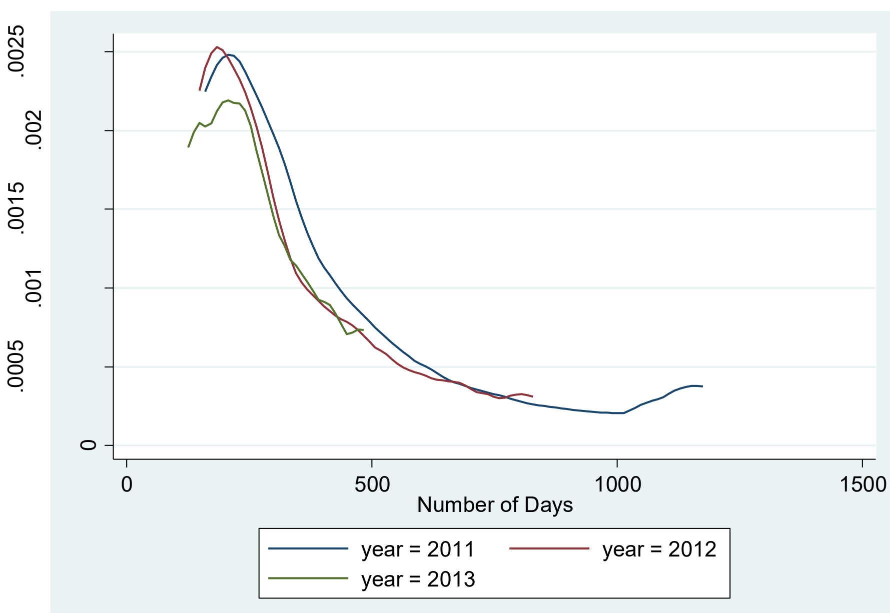
[4] We used Cox-Snell residual test to evaluate the goodness of fit of the model based on different hazard distribution and found log-logistic provides the best goodness of fit.

Figure 4: Hazard Curve, 2011 Q1 through 2013 Q4



We further analyzed the hazard curve by year to see if the hazard pattern is consistent for different time periods. Figure 5 shows the hazard curve by the year of the violation issue date for the observation period used for the survival model (2011 Q1 through 2013 Q4). The curves are similar for all three years, although the 2013 curve is shorter because the observation time for payment is shorter. For example, we can only observe a violation MSHA issued in September 2013 for one year, or about 365 days, because the data were pulled in September 2014.

Figure 5: Hazard Curve by Year, 2011 Q1 through 2013 Q4



The model estimation for the survival model follows a similar process to the binary logistic regression estimation for research question 1. After compiling a comprehensive list of covariates, we used stepwise selection and the VIF test to select covariates. Table 4 presents the coefficient signs and p-values for the selected covariates.

Table 4. Brief Covariate Specification for Survival Analysis

Rank	Variable	Sign	p-value
1	Canvass Code	Mixed	N/A
2	Violation Type	Mixed	N/A
3	District	Mixed	N/A
4	Indicator of 90-day Delinquency in Prior Year	+	0.00
5	Type of Company	Mixed	N/A
6	Rate of Injuries to Violations in Prior Year	-	0.00
7	Percent of Violations 90-day Delinquent in Prior Year	+	0.00
8	Mine Type	Mixed	N/A
9	Portable Operation Indicator	-	0.00
10	Log Number of Employees	-	0.00
11	Interaction: (% of Violations 90-day Delinquent in Prior Year) x (Log Sum Delinquent Penalties in Prior Year)	+	0.00
12	Log Sum of Delinquent Penalties in Prior Year	-	0.00
13	Log Number of Violations in Prior Year	-	0.00
14	Violator Type (Contractor)	+	0.03
15	Mine Size Points	-	0.00
16	Log penalty amount	+	0.00
17	Number of Injuries in Prior Year	+	0.00

Table 5 shows the contribution for the top five positive contributors for the survival model. Unlike the results on the binary logistic model, on average, the top five contributors (violation type - order, indicator of 90+ day delinquency in prior year, coal district 6 (eastern Kentucky), primary canvass code - coal anthracite, and percent of violations delinquent in prior year) have fairly equal influence on the delinquency duration. Violation type - order and the indicator of 90-day delinquency in the prior year each account for the 16% of the positive contributions, while coal district 6 contributes 15%. The top three contributors accounted for 47% of the positive $\boldsymbol{x\beta}$ values, and the top five contributors accounted for 70%.

Table 5: Top Five Positive Contributors to Delinquency Duration

Characteristic	Percent of Positive Contribution
Violation Type - Order	16%
Indicator of 90-day delinquency in prior year	16%
District - Coal 6	15%
Canvass Code - Coal Anthracite	13%
Percent of violations delinquent in prior year	10%
Top Three Contributors	47%
Top Five Contributors	70%

Research Question 3. Which mines and violators are at the highest risk of long-term delinquencies?

This research question combines the outputs of the first two research questions to generate a risk score for each violation. To generate a risk score for each violation, we multiplied the probability of 90-day delinquency with the predicted duration of delinquency.

$$risk\ score_i = \Pr(90\ days\ delinquency) \times predicted\ median\ time$$

In the Early Detection tool, risk scores are only calculated for violations that have not been paid. We do not estimate predicted probability for 90+ day delinquency for violations that are already 90+ days delinquent and have not been paid because the probability is known (1.00). The risk scores for these violations are the probability of delinquency (1.00) multiplied by the time delinquent, which would equal time delinquent.

Once we calculated a risk score for each violation, we aggregated risk scores to the mine and violator levels by taking the average risk score of the violations associated with a given mine or violator.

Correlation between the model predicted risk score and actual delinquency duration is the key to accurately identify the violations with long delinquencies. We tested the rank ordering correlation between the risk score and delinquency duration using the Spearman's test. Spearman's correlation coefficient (ρ) is a nonparametric measure of assessing the association between two ranked variables. The risk score and delinquency duration have a Spearman's coefficient (ρ) equal to 0.38 and a statistically significant a p-value of 0.00. Increases in the predictive model's risk score correlate significantly with increases in the actual delinquency duration of violators.

Applications

The predictive model demonstrated predictive power to discriminate operators with high delinquency risk and uncovered key relationship in the delinquency behavior in a statistical and systematic way. The model can help MSHA to detect at an early stage which operators with violations are at risk of failing to pay their fines in a timely manner before they appear on the list of Scofflaw operators and contractors with severe delinquency. The model can predict violators' risk score as soon as MSHA issues the violation.

In order to apply the model and score to MSHA's enforcement activities, we developed a business intelligence tool that implements the predictive model. We implemented this predictive model in a Microsoft Excel application so that MSHA can periodically use in-house resources to conduct the analysis and use the results for early enforcement activities. In this application, we aggregated the predicted scores to mine and operator levels. All the mines and operators can be ranked and categorized into high, medium, and low risk categories to identify the violators with highest risk of delinquency and provide warning of risk severity.

References

[1] Alison Morantz, "Does Unionization Strengthen Regulatory Enforcement- An Empirical Study of the Mine Safety and Health Administration," *NYUJ Legislation & Public Policy* 14(2011):703-704.

[2] "Metal and Nonmetal Safety and Health." *Mine Safety and Health Administration.* Accessed November 3, 2014, http://msha.gov/PROGRAMS/METAL.HTM.

[3] "MSHA's Mission." *Mine Safety and Health Administration.* Accessed November 3, 2014, http://www.msha.gov/MSHAINFO/MISSION.HTM.

[4] “Coal Mine Safety and Health.” *Mine Safety and Health Administration.* Accessed November 3, 2014, http://msha.gov/PROGRAMS/COAL.HTM.

[5] Patricia W. Silvey to Jay Mattos, January 17, 2012, Memorandum; Scofflaw Program, MSHA Office of Assessments, 3-4.

[6] “Response to NPR Series From the Mine Safety and Health Administration,” *NPR,* January 12, 2015, http://www.npr.org/2015/01/12/372490930/response-to-npr-series-from-the-mine-safety-and-health-administration

[7] Kalbfleisch, John and Ross Prentice. *The Statistical Analysis of Failure Time Data.* Hoboken: Wiley-Interscience, 2002.