**DATA PREPARATION FOR PREDICTIVE ALGORITHMS**

1. **Notation**

We relied on two sets of (unbalanced) panel data that describe mines over years:[[1]](#footnote-1) one includes information about the number of maintenance and repair injuries that occurred in these mines (***MR dataset***), and the other, the number of pinning and striking injuries that occurred in these mines (***PS dataset***). These datasets also include information about the number of violations that were issued to these mines for infractions of certain subpart denominations of the safety code (***violation types***). These violation types were selected by NIOSH as those that may be meaningfully related to each injury type; [[2]](#footnote-2) therefore, different sets of violation types are included in the MR and the PS datasets.

We proscribe the following notation to describe these datasets:

Let the injury type be represented by such that .

Let represent the number of injuries of type that occurred in mine at time . Then, the injuries in each dataset are represented by the vector,

Let there be violation types that are presumed to be meaningfully related to injuries of type , such that

Let represent the number of violations of type that were issued to mine at time (violation type is presumed to be meaningfully related to injuries of type ). Then, violations in each dataset are represented by the matrix,

The other variables in our datasets concern mine-level characteristics; these variables are represented as follows,

number of employees in mine at time

number of tons of coal produced in mine at time

operator for mine at time

time at which operator took control of mine

time at which operator relinquished control of mine

indicator for whether a safety committee was present in mine at time

district for mine

indicator for whether mine is in Appalachia

number of on-site inspection hours in mine at time

total number of violations issued to mine at time

1. **Inference: Which violation types are statistically significantly related to injuries?**
   1. **Methodology**
      1. **Theoretical Considerations**

The goal of this project was to identify whether a failure to adhere to any (and if so, to which) equipment-related safety regulations and/or procedures was significantly associated with equipment-related injuries in underground coal mines. This task probed the mechanisms undergirding injuries with the goal of discovering specific areas that could benefit from technological intervention or heightened scrutiny.

For this project, we supposed that a mine’s adherence to equipment-related safety regulations and/or procedures could be measured by its violation history; that is, by the violations that were issued to the mine for infractions of certain subpart denominations of the safety code that were deemed to meaningfully correspond with equipment-related injuries. However, we questioned whether all mines had an equal *opportunity* to be issued violations – in other words, to what extent is an endogenous variable. It is certainly possible that is a reliable metric for the adherence to equipment-related safety regulations and/or procedures. But it is also possible that reflects, instead, the amount of time an inspector spent on a given inspection. Therefore, for our analyses, we measured violation history for a mine using both the count of violations (of each violation type) that were issued, as well as the rate of violations (of each violation type) that were issued (per on-site inspection hour) (see *Section 2.1.2*).

In addition to this consideration, we questioned how the associations (if they existed) between injuries in time and violation history from before time could differ according to the operationalization of violation history and/or injuries. Specifically, we asked:[[3]](#footnote-3)

1. Is *short*- or *long*-term violation history related to injuries?; and
2. Is violation history related to whether *any* subsequent injuries occur or to *how many* subsequent injuries occur?

In our view, it was worth investigating violation-injury associations using different measures of both violation history and injuries. Consistent results across differing specifications could give us confidence in any discovered associations, and divergences might also be informative. In fact, differences could cast light on nuanced aspects of violation-injury relations. For example, if long-term violation history was found to be significantly related to subsequent injuries, but short-term violation history was not, this would suggest that injuries may be explained by long-established safety culture rather than by recent safety infractions. Therefore, for our analyses, we used two measures of violation history and injuries (each) as suggested by the above questions (see *Section 2.1.2*).

* + 1. **Model Specification**

We designed 8 general models to explain injuries in time with violations from before time .

In response to question (1) in *Section 2.1.1*, the variables of interest in the models were either, [[4]](#footnote-4)

1. The number of violations (of each violation type) issued at time ,
2. The cumulative number of violations (of each violation type) issued over the four years preceding time ,

Recall that given our consideration about *opportunity* to be issued violations (see *Section 2.1.1*), we considered that these variables of interested could be either,

1. The number of violations (of each violation type) issued, as outlined above,
2. The number of violations (of each violation type) issued per inspection hour,[[5]](#footnote-5)

In response to question (2) in *Section 2.1.1*, the outcome variable was defined as either,

1. An indicator for whether any injuries occurred at time ,
2. The count of the injuries that occurred at time ,

The models with an indicator outcome were specified using a probit model; those with a count outcome were specified using a negative binomial model.[[6]](#footnote-6) Standard errors were clustered by mine.

In all models, hours worked in mine at time was included as an exposure term. All models also included the following covariates:[[7]](#footnote-7)

* log of number of employees in mine at time ;
* log of number of tons of coal produced in mine at time ;
* number of injuries in mine at time ;[[8]](#footnote-8)
* indicator for whether a safety committee was present in mine at time ;
* district (as a factor variable);
* indicator for whether mine is in Appalachia;[[9]](#footnote-9)
* year (as a factor variable); and
* log of number of consecutive years mine has been operated by its operator at time .[[10]](#footnote-10)

The left- and right-hand sides of the models are summarized in **Figure 1**.

**Figure 1. Left- and Right-Hand Sides of Preferred Models**

|  |  |
| --- | --- |
| Left-Hand Side | Right-Hand Side |
| Injuries in mine at time ; either: | Number of violations (of each violation type) in mine ; either: |
| Indicator or Count | Count or Per onsite inspection hour |
|  | At time or Cumulatively over |
|  | Exposure term (hours worked in mine at time ) |
|  | Number of employees in mine at time |
|  | Number of tons of coal produced in mine at time |
|  | Number of injuries in mine at time |
|  | Indicator for whether a safety committee was present in mine at time |
|  | District (as a factor variable) |
|  | Indicator for whether mine is in Appalachia |
|  | Year (as a factor variable) |
|  | Number of consecutive years mine has been operated by its operator at time |

We ran these models separately on the MR data and on the PS data, resulting in 16 model specifications (***preferred models***). These models are summarized in **Table 1**.

**Table 1. Preferred Models**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Label** | **Outcome** | | **Variables of Interest** | | **Specification** | **All Models Include** |
| MR-B-VC-1 |  |  | , |  | Probit | - exposure term ()  -  -  -  -  - (as factor)  - (as factor)  - (as factor)  -  Standard errors are clustered by mine |
| MR-B-VR-1 |  |  | , |  | Probit |
| MR-B-VC-4 |  |  |  |  | Probit |
| MR-B-VR-4 |  |  |  |  | Probit |
| MR-C-VC-1 |  |  |  |  | Negative Binomial |
| MR-C-VR-1 |  |  |  |  | Negative Binomial |
| MR-C-VC-4 |  |  |  |  | Negative Binomial |
| MR-C-VR-4 |  |  |  |  | Negative Binomial |
| PS-B-VC-1 |  |  |  |  | Probit |
| PS-B-VR-1 |  |  |  |  | Probit |
| PS-B-VC-4 |  |  |  |  | Probit |
| PS-B-VR-4 |  |  |  |  | Probit |
| PS-C-VC-1 |  |  |  |  | Negative Binomial |
| PS-C-VR-1 |  |  |  |  | Negative Binomial |
| PS-C-VC-4 |  |  |  |  | Negative Binomial |
| PS-C-VR-4 |  |  |  |  | Negative Binomial |

* + 1. **Model Assessment**

For each of our preferred models, we assessed which violation types were statistically significantly () related to injuries and the magnitude and direction of the estimated effects. Moreover, we explored pattern across and divergences between specifications. Specifically, for each of the following groups of models, we evaluated which violation types were consistently statistically significantly () related to injuries, as well as the ranges of the magnitudes and the consistency of the directions of the estimated effects:

1. All models for each injury type, ;
2. Models for each injury type, , that use an indicator outcome variable (;
3. Models for each injury type, , that use a count outcome variable ();
4. Models for each injury type, , that use the number of violations (of each violation type) at time () as the variables of interest;
5. Models for each injury type, , that use the (scaled) number of violations (of each violation type) per on-site inspection hour at time () as the variables of interest;
6. Models for each injury type, , that use the cumulative number of violations (of each violation type) over the four years preceding time () as the variables of interest;
7. Models for each injury type, , that use the (scaled) cumulative number of violations (of each violation type) per cumulative on-site inspection hour over the four years preceding time () as the variables of interest
8. Models in either (4) *or* (5);
9. Models in either (6) *or* (7);
10. Models in either (4) *or* (6);
11. Models in either (5) *or* (7);
12. Models in (2) *and* (8);
13. Models in (2) *and* (9);
14. Models in (2) *and* (10);
15. Models in (2) *and* (11);
16. Models in (3) *and* (8);
17. Models in (3) *and* (9);
18. Models in (3) *and* (10); and
19. Models in (3) *and* (11).
    * + 1. **Falsification Test**

We endeavored to test the null hypothesis that injuries are independent of violation types. (In other words, we wanted to test whether the significant associations that we identified were false positives). Toward this end, we employed a two-stage randomization inference procedure informed by Ho & Donahue. This method tests the null hypothesis that an outcome is independent of a treatment, and is traditionally carried out by repeatedly randomizing the treatment (effectively enforcing a scenario in which the null hypothesis is true) and re-estimating the chosen model. Then, one compares the coefficient of the non-randomized treatment to the distribution of the coefficients estimated from the randomized treatments and generates a p-value for the test of the null hypothesis.

We executed our two-stage randomization inference procedure for each of our preferred models separately. In the first stage of our procedure, we considered the treatment to be the *joint distribution* of the variables of interest in the given preferred model. We sampled, without replacement, *every* one of these variables and re-estimated the preferred model; we repeated this step 1,000 times. For every variable of interest that had appeared as significant in the preferred model, we calculated the p-value for the test of the null hypothesis.

In the second stage of our procedure, *each* variable of interest in the given preferred model was considered to be a *separate* treatment. Therefore, for every variable of interest that had appeared as significant in the given preferred model and had also generated a p-value to reject the null hypothesis in the first stage of the randomization inference procedure (), we sampled, without replacement, this variable, and re-estimated the preferred model; we repeated this step 1,000 times. We then calculated the p-values for the tests of the null hypotheses.

* + - 1. **Specification Tests**

We performed two specification tests for all of our preferred models. The first test was designed to check whether our exclusion of certain covariates dramatically changed the “story” told by the preferred models. As noted in footnote *g*, we considered additional covariates that were ultimately excluded due to incompleteness of data. However, we had enough data to run a specification test for two of these variables: union status and longwall indicator. We therefore re-fit each of the preferred models with these two variables included on the right hand side. (However, these models were only fit on data covering 2000-2013 because the additional covariates are lacking for 2013-2016). We then assessed similarities and differences between each preferred model and its associated specification test model along which violation types were statistically significantly () related to injuries and the magnitude and direction of the estimated effects.

The second specification test was designed to check whether relaxing the definitions of prior violations and injuries dramatically changed the “story” told by the preferred models. As noted in *Section 1*, we work with unbalanced panel data; therefore, we lack rows in our datasets describing some mine at some time . When we predict , we use , , and/or . As a specification test, if is missing, we substitute , as long as is non-missing; if is missing, we substitute as long as is non-missing; and if one for is missing, we substitute , treating the missing as 0, and as long as is non-missing. We then re-fitted each of the preferred models and assessed similarities and differences between each preferred model and its associated specification test model along which violation types were statistically significantly () related to injuries and the magnitude and direction of the estimated effects.

* + 1. **Supplemental Analyses**

For every set of outcomes and variables of interest that were probed by our preferred models, we executed a Random Forest decision analysis and assessed whether the variables of interest deemed important by this method matched the variables of interest that we found to be significantly related to the outcome in the given preferred model.

1. **Prediction: Can violation types predict injuries?**
   1. **Methodology**
      1. **Theoretical Considerations**

The purpose of this project was to identify whether a failure to adhere to any (and if so, to which) equipment-related safety regulations and/or procedures was significantly associated with equipment-related injuries. Therefore, we investigated which prior violation types were significantly associated with subsequent injuries; this task was one of *inference*. A related objective was to uncover whether violations, in particular those types that we believe play a meaningful role in the occurrence of injuries, could be used to *predict* subsequent injuries. More specifically:

1. How well do the violation types that we believe to be meaningfully related to injuries predict those injuries?; and
2. Are these meaningful violation types better predictors of injuries than a simple count of all violations?

It is essential to note that this predictive approach *does not* answer the question, “Can we predict injuries from violations, and if so, how well?”. We *do not* create an algorithm driven by the sole goal of accurate prediction. Rather, we seek to answer the questions, “Can the violation types that we believe play a meaningful role in the occurrence of injuries be used to predict those injuries, and if so, how well?”. The important difference between the questions we answer and the question we do not is our focus on certain selected violation types. The fact that this method only allows the machine to access certain violation types that humans deemed to be relevant is a severe limitation for predictive power. Indeed, were the goal of this project to produce a predictive algorithm based on violation history, say, for the identification of mines to have further inspection – not, as it is, to identify the tractability of technological innovation – we would employ machine learning methods that utilize every violation-related variable to which we have access.

Critically, we treated this prediction task as separate from the inference task. Though the motivating questions for these analyses were clearly related, we used no information from our inference procedures to inform our predictive approach. This separation is necessary because the predictive task requires us to split the data into training and testing sets, whereas the inference task utilizes all of the data. (When we refer to violations that we believe *play a meaningful role* in the occurrence of injuries, we refer to the fact that the violation types that make up our datasets were selected by NIOSH at the start of this project).

* + 1. **Model Specification**

Given the predictive goal, we split our datasets into training and testing sets. The training set(s) contained all observations describing 2000-2012. The testing set(s) contained all observations describing 2013-2015.

On the training set(s), we trained the 16 preferred models described in *Section 2.1.2*. We also specified a set of null models to be used as a comparison for our preferred models. In essence, we set up a horse race between our preferred models and several null models.

In response to question 1 in *Section 3.1.1*, we specified a set of ***weak null models*** that perfectly matched the preferred models except that they excluded all of the variables of interest (i.e., the violation type variables) on the right-hand side. The left- and right-hand sides are outlined explicitly in **Figure 2**.

**Figure 2. Left- and Right-Hand Sides of Weak Null Models**

|  |  |
| --- | --- |
| Left-Hand Side | Right-Hand Side |
| Injuries in mine at time ; either: | Exposure term (hours worked in mine at time ) |
| Indicator or Count | Number of employees in mine at time |
|  | Number of tons of coal produced in mine at time |
|  | Number of injuries in mine at time |
|  | Indicator for whether a safety committee was present in mine at time |
|  | District (as a factor variable) |
|  | Indicator for whether mine is in Appalachia |
|  | Year (as a factor variable) |
|  | Number of consecutive years mine has been operated by its operator at time |

We ran these models separately on the MR data and on the PS data, resulting in 4 model specifications (***weak null models***). These models are summarized in **Table 2**.

**Table 2. Weak Null Models**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Label** | **Outcome** | | **Specification** | **All Models Include** |
| MR-B-W-NULL |  |  | Probit | - exposure term ()  -  -  -  -  - (as factor)  - (as factor)  - (as factor)  -  Standard errors are clustered by mine |
| MR-C-W-NULL |  |  | Negative Binomial |
| PS-B-W-NULL |  |  | Probit |
| PS-C-W-NULL |  |  | Negative Binomial |

In response to question 1 in *Section 3.1.1*, we specified a set of ***strong null models*** that perfectly matched the weak null models except that they included, on the right-hand side, a count of the total violations (i.e., not just the violations that were deemed meaningfully related to the injury types of interest) issued to mine at the most recent time during which the mine was active.

Given our consideration about *opportunity* to be issued violations (see *Section 2.1.1*), we considered that this additional variable could be either,

1. The total number of violations issued,
2. The total number of violations issued per inspection hour,

The left- and right-hand sides of these models are outlined explicitly in **Figure 3**.

**Figure 3. Left- and Right-Hand Sides of Strong Null Models**

|  |  |
| --- | --- |
| Left-Hand Side | Right-Hand Side |
| Injuries in mine at time ; either: | Total number of violations in mine ; either: |
| Indicator or Count | Count or Per onsite inspection hour |
|  | Exposure term (hours worked in mine at time ) |
|  | Number of employees in mine at time |
|  | Number of tons of coal produced in mine at time |
|  | Number of injuries in mine at time |
|  | Indicator for whether a safety committee was present in mine at time |
|  | District (as a factor variable) |
|  | Indicator for whether mine is in Appalachia |
|  | Year (as a factor variable) |
|  | Number of consecutive years mine has been operated by its operator at time |

We ran these models separately on the MR data and on the PS data, resulting in 8 model specifications (***strong null models***). These models are summarized in **Table 3**.

Clearly, not every “horse” competed against every other “horse” in the race; rather, each of the 16 preferred models corresponded to one weak null model and two strong null models for comparison. These associations are explicitly detailed in **Table 4**.

**Table 3. Model Comparisons**

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Compare** | | |
| **Weak Null Model** | **Strong Null Models** | |
| MR-B-VC-1 | MR-B-W-NULL | MR-B-VC-S-NULL | MR-B-VR-S-NULL |
| MR-B-VR-1 | MR-B-W-NULL | MR-B-VC-S-NULL | MR-B-VR-S-NULL |
| MR-B-VC-4 | MR-B-W-NULL | MR-B-VC-S-NULL | MR-B-VR-S-NULL |
| MR-B-VR-4 | MR-B-W-NULL | MR-B-VC-S-NULL | MR-B-VR-S-NULL |
| MR-C-VC-1 | MR-C-W-NULL | MR-C-VC-S-NULL | MR-C-VR-S-NULL |
| MR-C-VR-1 | MR-C-W-NULL | MR-C-VC-S-NULL | MR-C-VR-S-NULL |
| MR-C-VC-4 | MR-C-W-NULL | MR-C-VC-S-NULL | MR-C-VR-S-NULL |
| MR-C-VR-4 | MR-C-W-NULL | MR-C-VC-S-NULL | MR-C-VR-S-NULL |
| PS-B-VC-1 | PS-B-W-NULL | PS-B-VC-S-NULL | PS-B-VR-S-NULL |
| PS-B-VR-1 | PS-B-W-NULL | PS-B-VC-S-NULL | PS-B-VR-S-NULL |
| PS-B-VC-4 | PS-B-W-NULL | PS-B-VC-S-NULL | PS-B-VR-S-NULL |
| PS-B-VR-4 | PS-B-W-NULL | PS-B-VC-S-NULL | PS-B-VR-S-NULL |
| PS-C-VC-1 | PS-C-W-NULL | PS-C-VC-S-NULL | PS-C-VR-S-NULL |
| PS-C-VR-1 | PS-C-W-NULL | PS-C-VC-S-NULL | PS-C-VR-S-NULL |
| PS-C-VC-4 | PS-C-W-NULL | PS-C-VC-S-NULL | PS-C-VR-S-NULL |
| PS-C-VR-4 | PS-C-W-NULL | PS-C-VC-S-NULL | PS-C-VR-S-NULL |

* + 1. **Model Assessment**

After specifying the preferred and null models on the training set, we used the models to predict the outcome variable in the testing set; we then compared the model predictions to the true outcomes and generated a series of measures to evaluate performance.

The models that probed an indicator outcome (i.e., those specified with a probit model) reported the probability that each observation in the testing set would experience any injuries. For an observation with expected probability (based on the given preferred model) we defined the prediction to be,

Then, we calculated,

1. The number of True Positives (TP): mine-quarters that the given model correctly predicted would have injuries;
2. The number of True Negatives (TN): mine-quarters that the given model correctly predicted would have no injuries;
3. The number of False Positives (FP): mine-quarters that the given model incorrectly predicted would have injuries; and
4. The number of False Negatives (FN): mine-quarters that the given model incorrectly predicted would have no injuries.

From these measures, we calculated the Correct Classification Rate (CCR), the False Positive Rate (FPR), and the False Negative Rate (FNR), as follows:

The models that probed a count outcome (i.e., those specified with a negative binomial model) reported the predicted number of injuries that each observation in the training set would experience; however, given that we included an exposure term, these were not all integer values. We assessed these models in two ways.

First, we calculated the difference between the true outcomes and the predicted outcomes for each model. We then generated the sum of the squares of these differences (SSD). We also took the sum of squares of the positive and negative differences separately (SSPD and SSND, respectively).

Second, we transformed the predicted outcomes from each model to be binary. That is, for an observation with predicted outcome (based on the given preferred model) we defined the prediction to be,

Then, we calculated the number of true positives, true negatives, false positives, and false negatives, as well as the correct classification rate, the false positive rate, and the false negative rate, just as we did with the models that probed an indicator outcome.

For all of the preferred models, we calculate the difference between the model and each of its associated null models along all measures listed. These differences were examined to evaluate the predictive performance of our preferred models – that is, of the violation variables of interest.

* + - 1. **Robustness Checks**

We performed a series of specification tests for all of our preferred models. These tests were designed to check whether our results were sensitive to the definition of the training and the testing sets; that is, to check whether using different year cutoffs for dramatically changes the “story” told by the preferred models. We therefore let the cutoff between the training and the testing sets bet 2010, 2011, 2013, and 2014, and for each new cutoff, we re-fit each of the preferred models. We then assessed the predictive performance of each of the preferred models (compared to the null models, which were also run on the new training and testing sets), and checked similarities and differences differences between each preferred model and its associated specification test model along each of the predictive performance measures.

1. The mines in our dataset are not active and producing for every year in , and the years of inactivity do not fall solely on the boundaries. [↑](#footnote-ref-1)
2. These decisions were made in a meeting between Linda McWilliams, Miguel Reyes, John Heberger, Ellen Rubinstein, Alison Morantz, Sarah Levine, and Nikhil Saifullah at NIOSH in Pittsburgh on February 12th, 2016. [↑](#footnote-ref-2)
3. In preliminary and exploratory phases of this project, we considered three additional questions:

   1. Are injuries related to the number of violations (of each violation type) issued, the number of “significant and substantial” violations (of each violation type) issued, and/or the number of penalty points (for each violation type) issued?;
   2. What constitutes *long*-term violation history?; and
   3. Are injuries related to certain *part* or *subpart* denominations of the safety code?

   We chose not to pursue answers to these questions for the following reasons, respectively:

   1. We believe this question deserves investigation; however, our data prevents us from seeking answers. The barriers to analyzing the two alternative conceptions of violation history are, (a) when we probe “significant and substantial” violations rather than any violations, our data is too sparse to fit a valid model; and (b) given that the manner in which penalty points are assessed has changed over time (with the most recent change in 2007), probing penalty points requires us to abandon nearly half of our observations, resulting in too few observations on which to model our many variables of interest.
   2. We specifically considered two methods of operationalizing long-term violation history: over several recent years *or* over all the years for which we have data on mine *i*. We ultimately decided to measure long-term violation history using only the former definition; we did not think that performing parallel analyses for the latter definition would help to illuminate any nuance in the mechanisms undergirding injuries.
   3. This question was posed to identify which *groups* of violations to the subpart denominations of the safety code are meaningfully related to injury occurrence (with violations to the *part* denominations of the safety code representing *groups*). However, exploratory analyses revealed that the subpart violation types represented in our dataset were skewed across the part violation types represented in our dataset. That is, the part violation types were composed of a substantially different number of subpart violation types. Therefore, an investigation of part violation types would likely only show which part violation types were composed of more subpart violation types (i.e., there are more subpart violation types comprising part violation type , so part violation type has a higher variance and comes up significant). We were skeptical of such analyses and we thought they would not speak to our main question about violation type clustering, so we abandoned this question.

   [↑](#footnote-ref-3)
4. As noted in *Section 1*, we work with unbalanced panel data. Therefore, we lack some to explain the corresponding (i.e., there is no row in our data describing mine in time ). If any for is missing, we treat as missing, as well. As a specification test (see *Section 2.1.3.2*), if is missing, we substitute to explain , as long as is non-missing. Relatedly, if one for is missing, we substitute , treating the missing as 0, and as long as is non-missing. [↑](#footnote-ref-4)
5. These variables were scaled by 1,000 to aid coefficient estimation. [↑](#footnote-ref-5)
6. This model was selected because our datasets feature overdispersion. [↑](#footnote-ref-6)
7. We also considered the following covariates: union status of mine at time ; longwall indicator for mine at time , coal bed thickness in mine at time ; number of coal beds in mine at time ; subsidiary company indicators for mine at time ; captive production of mine at time ; recoverable reserves of mine at time ; and age of mine at time . However, we had insufficient data for the inclusion of these variables in our models. Though we lacked enough data on union status and longwall indicators to include in our preferred models, we include these variables in the models as a specification test (see *Section 2.1.3.2*). For these tests data on 2014-2016 was dropped. Future investigations may benefit from access to and examination of all of these additional covariates. [↑](#footnote-ref-7)
8. As noted in *Section 1*, we work with unbalanced panel data. Therefore, we lack some to explain the corresponding (i.e., there is no row in our data describing mine in time ). As a specification test (see *Section 2.1.3.2*), if is missing, we substitute to explain , as long as is non-missing. [↑](#footnote-ref-8)
9. This indicator takes on a value of 1 for the states VA, WV, KY, and PA and takes on a value of 0 otherwise. [↑](#footnote-ref-9)
10. As noted in *Section 1*, we work with unbalanced panel data; therefore, for some , there are no rows in our data describing mine . Consider the following case: at time , mine has been operated by operator for consecutive years, at time , mine is inactive, and at time , mine is active and operated by operator . We allow this variable to take on the value at time , effectively ignoring the inactive time in mine ’s history. [↑](#footnote-ref-10)