**Classifying Bad Mines**

***Selected sample:***

1. Mines in the 90th percentile of average, maximum, *and* median number of violations in the first 8 active quarters of mine history (N = 111)
2. Mines in the 90th percentile of average, maximum, *and* median number of injuries in the first 8 active quarters of mine history (N = 131)

***Robustness Checks:***

* Mines in the 90th percentile of average number of violations in the first 8 active quarters of mine history (N = 199)
* Mines in the 90th percentile of maximum number of violations in the first 8 active quarters of mine history (N = 161)
* Mines in the 90th percentile of median number of violations in the first 8 active quarters of mine history (N = 198)
* Mines in the 90th percentile of average, maximum, *or* median number of violations in the first 8 active quarters of mine history (N = 206)
* Mines in the 90th percentile of average, maximum, *and* median number of violations in the first 6 active quarters of mine history (N = 107)
* Mines in the 90th percentile of average number of violations in the first 6 active quarters of mine history (N = 199)
* Mines in the 90th percentile of maximum number of violations in the first 6 active quarters of mine history (N = 160)
* Mines in the 90th percentile of median number of violations in the first 6 active quarters of mine history (N = 198)
* Mines in the 90th percentile of average, maximum, *or* median number of violations in the first 6 active quarters of mine history (N = 209)
* Mines in the 90th percentile of average number of injuries in the first 8 active quarters of mine history (N = 167)
* Mines in the 90th percentile of maximum number of injuries in the first 8 active quarters of mine history (N = 189)
* Mines in the 90th percentile of median number of injuries in the first 8 active quarters of mine history (N = 168)
* Mines in the 90th percentile of average, maximum, *or* median number of injuries in the first 8 active quarters of mine history (N = 226)
* Mines in the 90th percentile of average, maximum, *and* median number of injuries in the first 6 active quarters of mine history (N = 135)
* Mines in the 90th percentile of average number of injuries in the first 6 active quarters of mine history (N = 171)
* Mines in the 90th percentile of maximum number of injuries in the first 6 active quarters of mine history (N = 207)
* Mines in the 90th percentile of median number of injuries in the first 6 active quarters of mine history (N = 162)
* Mines in the 90th percentile of average, maximum, *or* median number of injuries in the first 6 active quarters of mine history (N = 234)

***Methodology:***

To identify bad mines, we use both (total) injuries and violations hours as a proxy for mine behavior.

For every mine in the dataset, we extract the first 4, 6, and 8 active quarters (quarters in which the mine had non-zero production) in the mine history. For each subset of data, for every mine, we calculate:

* Average number of violations and injuries
* Maximum number of violations and injuries
* Median number of violations and injuries

For each of these measures, we select the mines in the 90th, 95th, and 97th percentile.

We check whether using violations or injuries identifies different bad mines. Indeed, we see a considerable lack of overlap in the mines that are selected by specifications using injuries vs. violations (all else being equal). Therefore, we proceed with two sets of analyses – one using violations and the other injuries as the measure of mine behavior.

We want to ensure that we use an adequate subset of data to identify bad mines. Therefore, we calculate the lack of overlap in the mines that are selected by specifications using the first 4, 6, and 8 active quarters of mine history (all else being equal). Across violations, injuries, and average, maximum, and median measures, we observe a minimal number of non-overlapping mines identified using the first 4 and 6 or the first 6 and 8 active quarters of mine history; we do, however, see a considerable number of non-overlapping mines identified using the first 4 and 8 active quarters of mine history. We believe using 8 quarters of mine history to detect mind behavior is the safest route. Given the similarities between mines identified using 6 and 8 quarters of mine history, we will use the mines identified using 6 quarters of mine history as a robustness check.

We now decide which percentile cutoff to use. We want an adequate sample size for subsequent analyses, so we are inclined to use the 90th percentile cutoff. The other measures give a non-viable sample; therefore, we must use this cutoff.

We next assess the differences between the mines in the 90th percentile of badness according to each measure of badness. The number of mines that are not shared between the following groups of mines are as follows:

1. Using violations as the measure of mine behavior
   * Average & Maximum – 29
   * Average & Median – 18
   * Maximum & Median – 49
2. Using injuries as the measure of mine behavior
   * Average & Maximum – 25
   * Average & Median – 12
   * Maximum & Median – 57

This is a considerable number of non-overlapping mines; however, we want to be strict. To be conservative, we thus select mines that are in the 90th percentile of size for *all* measures of size (i.e., average, maximum, *and* median). We use individual measures, as well as mines classified as large by *any* measure as robustness checks.