

# Problem Set 2

for Empirical Analysis of Firm Behavior

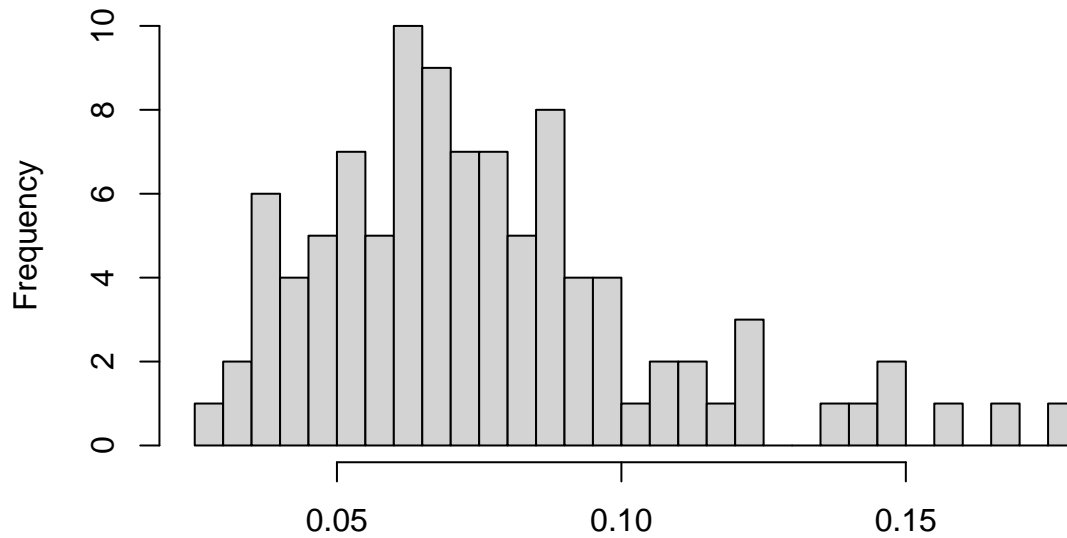
Mason Ross Hayes

11 November 2021

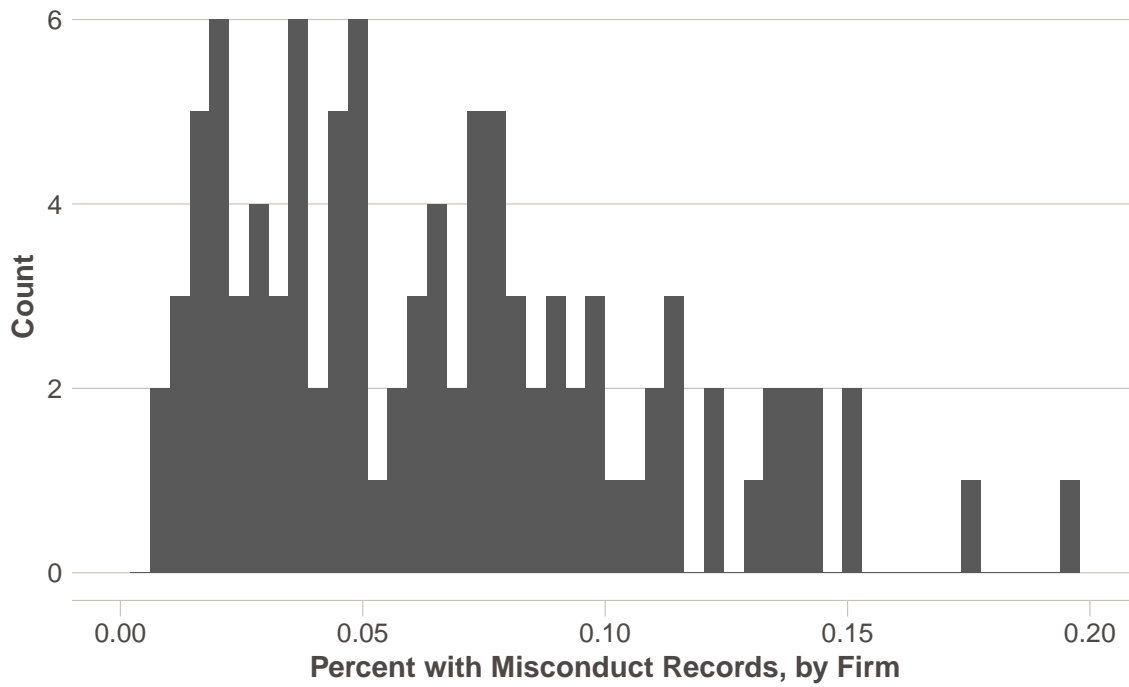
## Question 1

The distribution of percent misconduct by firm could be described by an Inverse- $\chi^2$  distribution or an Inverse-Gamma distribution, depending on the parameters we choose. Below I show a histogram of 100 random samples from an Inverse- $\chi^2$  with 14 degrees of freedom to compare to the firm misconduct histogram. The Inverse- $\chi^2$  can have a longer tail, but on the support of firm misconduct, the histograms are very similar.

**Figure 1: Inv Chi-Squared Probability Density**

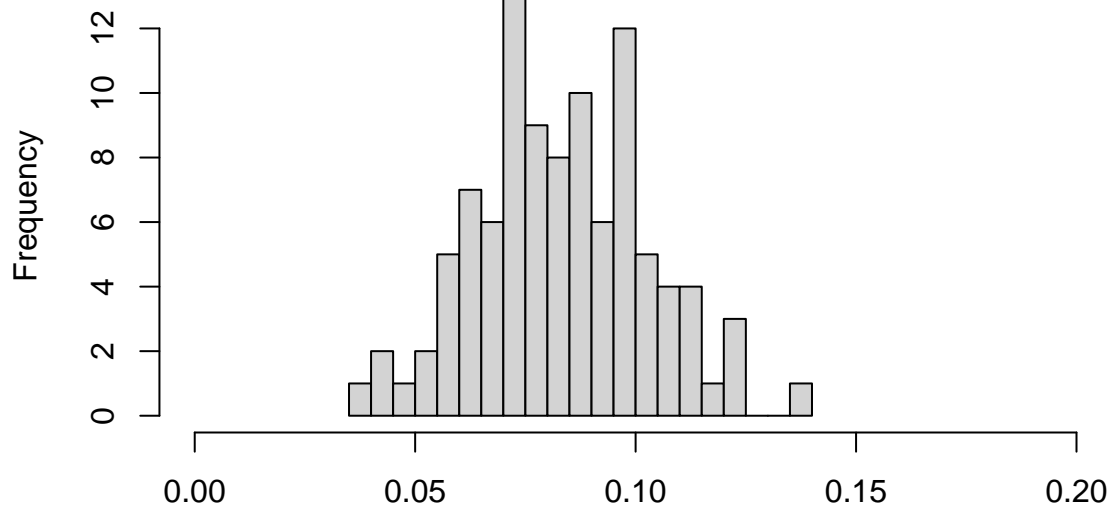


**Figure 2: Firm Level Misconduct**

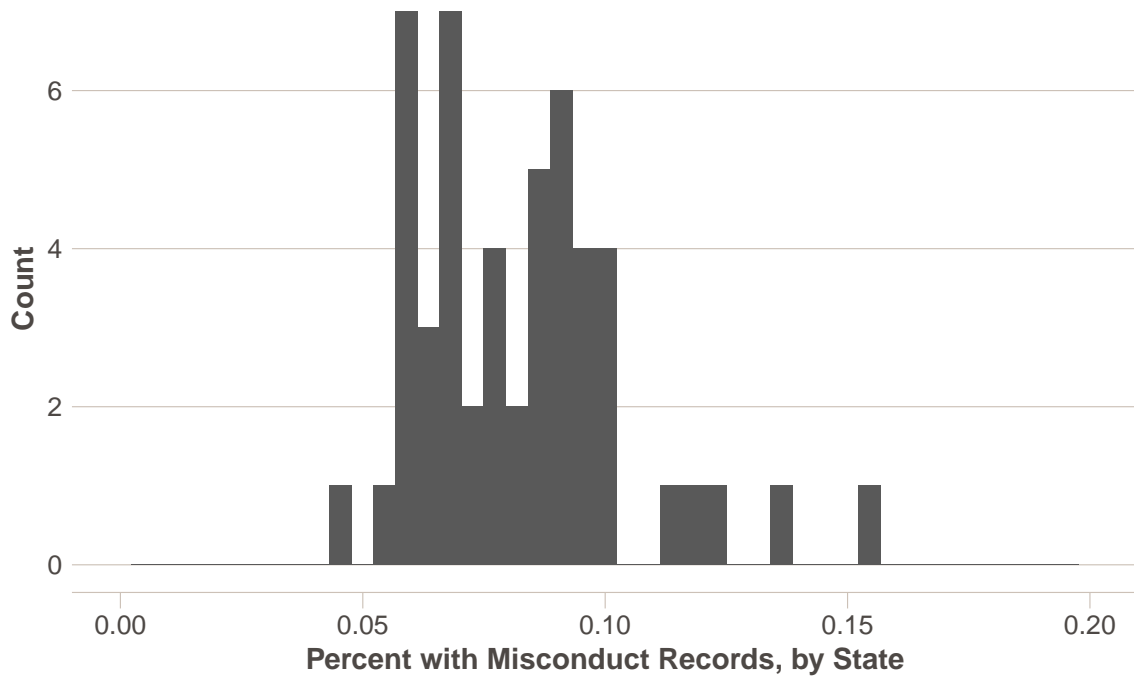


The distribution of percent misconduct at the state level looks more like a normal distribution. Using the mean and standard deviation of state percent misconduct and sampling 100 times from a normal distribution, we get:

**Figure 3: Normal Probability Density**



**Figure 4: State Level Misconduct**



Compared to the firm-level distribution, the state-level distribution shows less variance; misconduct is much more concentrated at the state level. This makes sense — at the firm level we would expect greater variance, as some firms may have a propensity to engage in misconduct that influences the (mis)conduct of their employees.

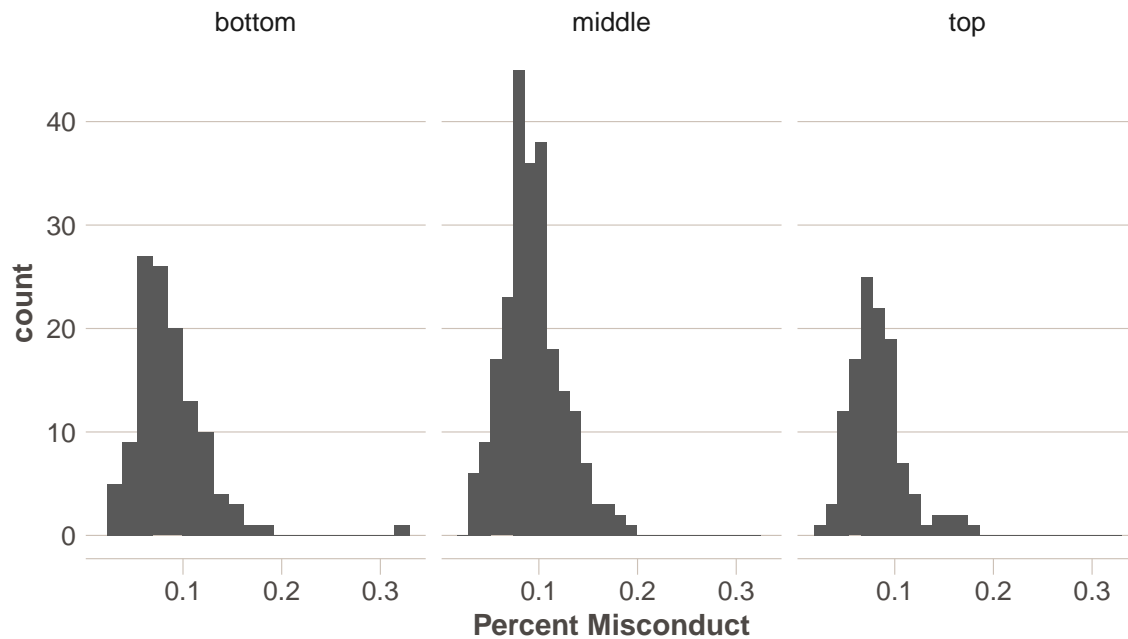
Or, firms may be in some industry where misconduct is more likely for some structural or organizational reason. Similarly, some firms may be especially strict in regulating misconduct or may emphasize virtues such as honesty and integrity in the workplace, leading both to employment self-selection and to peer effects. These different qualities would vary greatly across specializations and across firms, but we would expect that the composition of firms is relatively similar in different states.

At the state-level, where we aggregate across firms that exist in a given state, it makes sense that the average misconduct will show less variance between states than between firms. To grossly generalize, the distribution of aggregated data will always show less variance than individualized data.

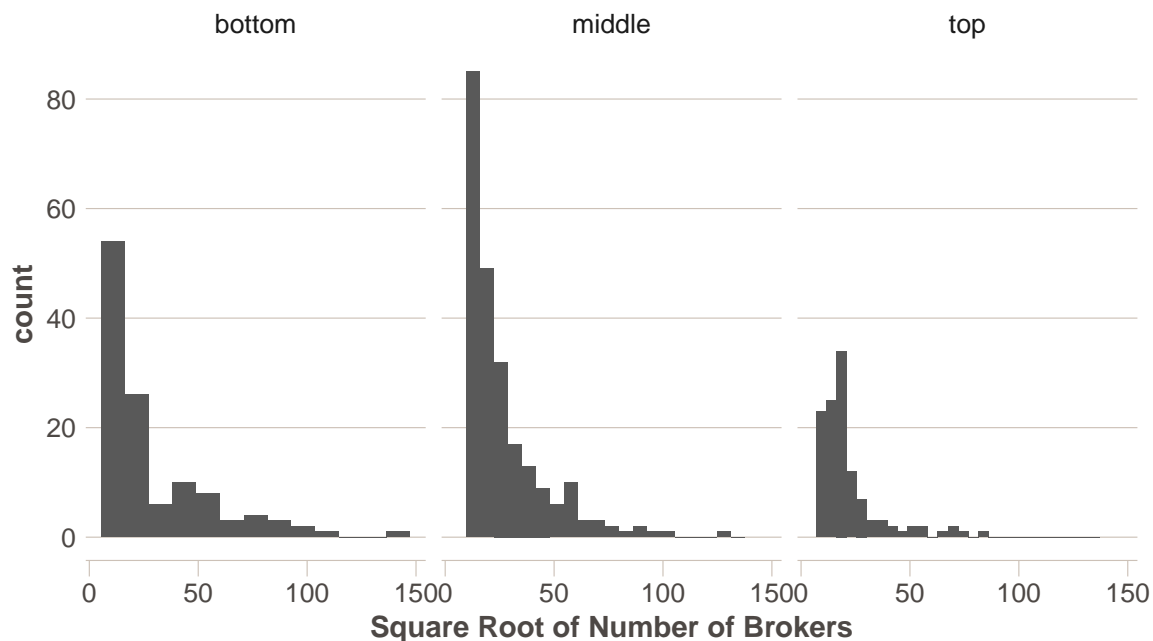
## Question 2

Before running this regression, there are a couple useful graphs to look at: what is the distribution of the percent misconduct conditional on the number of brokers, and what is the distribution of the number of brokers conditional on the percent misconduct?

**Figure 5: Percent Misconduct by County**  
**By Quantile of Number of Brokers**



**Figure 6: Number of Brokers by County**  
**By Quantile of Percent Misconduct**

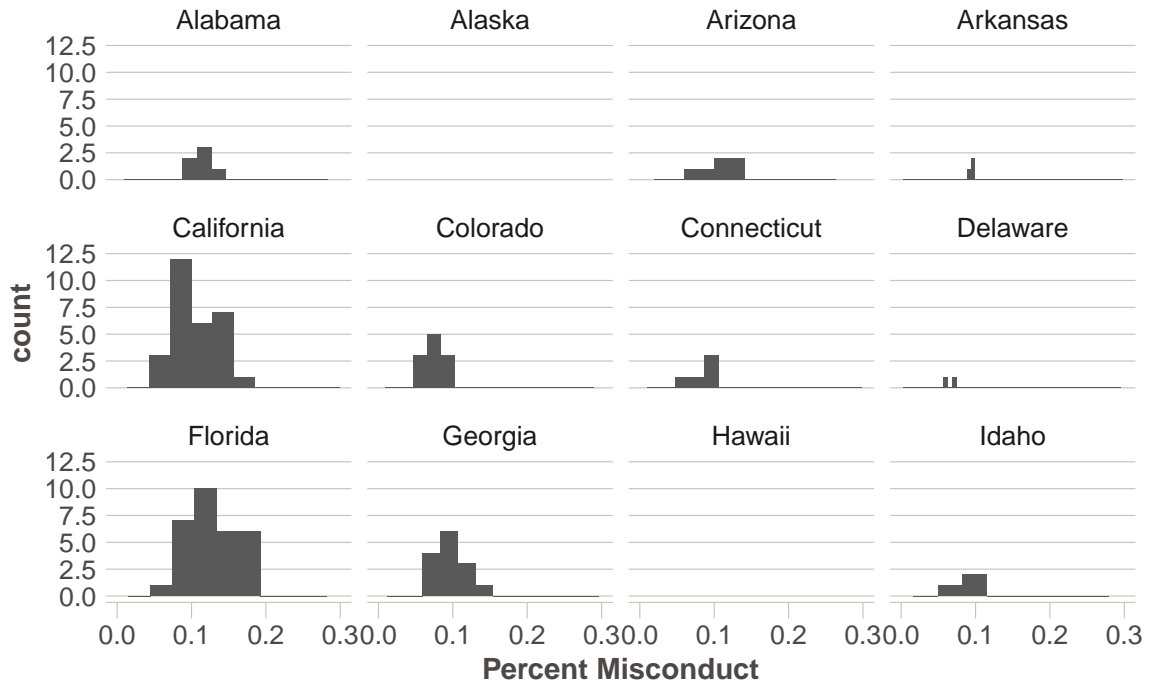


The distribution of the number of brokers for those counties falling in the bottom quantile of percent misconduct has a longer tail compared to those in the top quantile: the likelihood of a high number of brokers is higher. Without this long tail, the distributions of bottom, middle, and top quantiles all look remarkably similar.

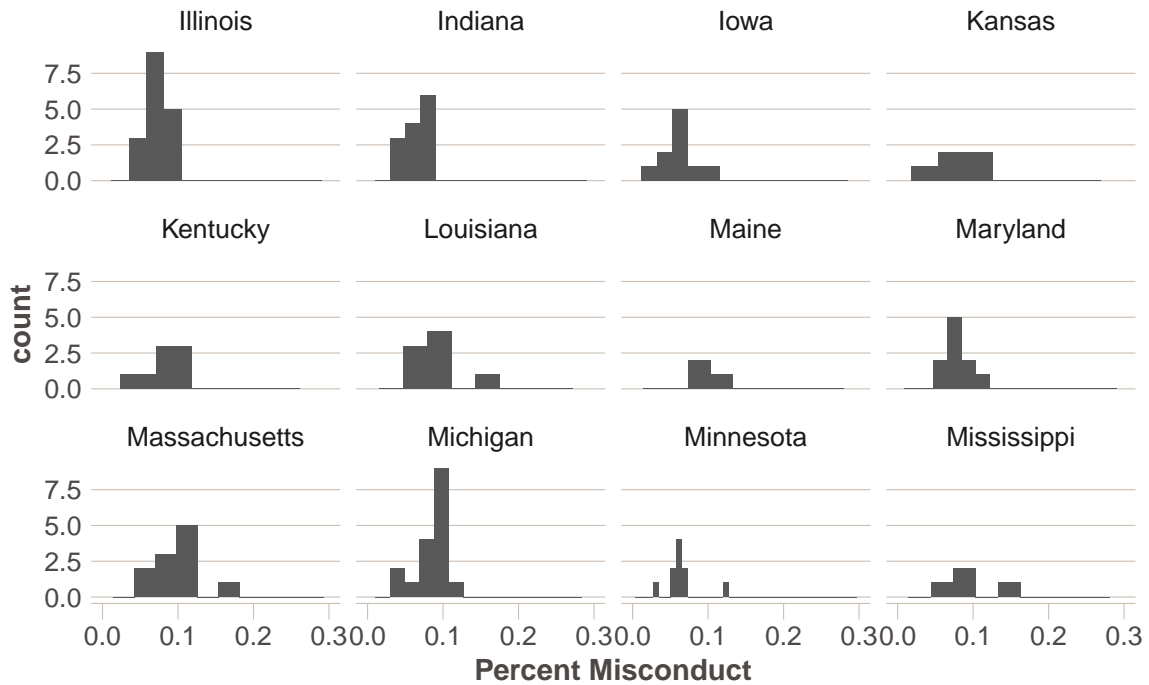
The estimation of state level fixed effects will depend on the quality of data that we have; for many states, we have only a handful of observations. This may be insufficient to truly estimate state effects if we only observe one county — and in this scenario, we are probably only observing the most populous or wealthiest counties.

In the following figures I show the distribution of misconduct by state:

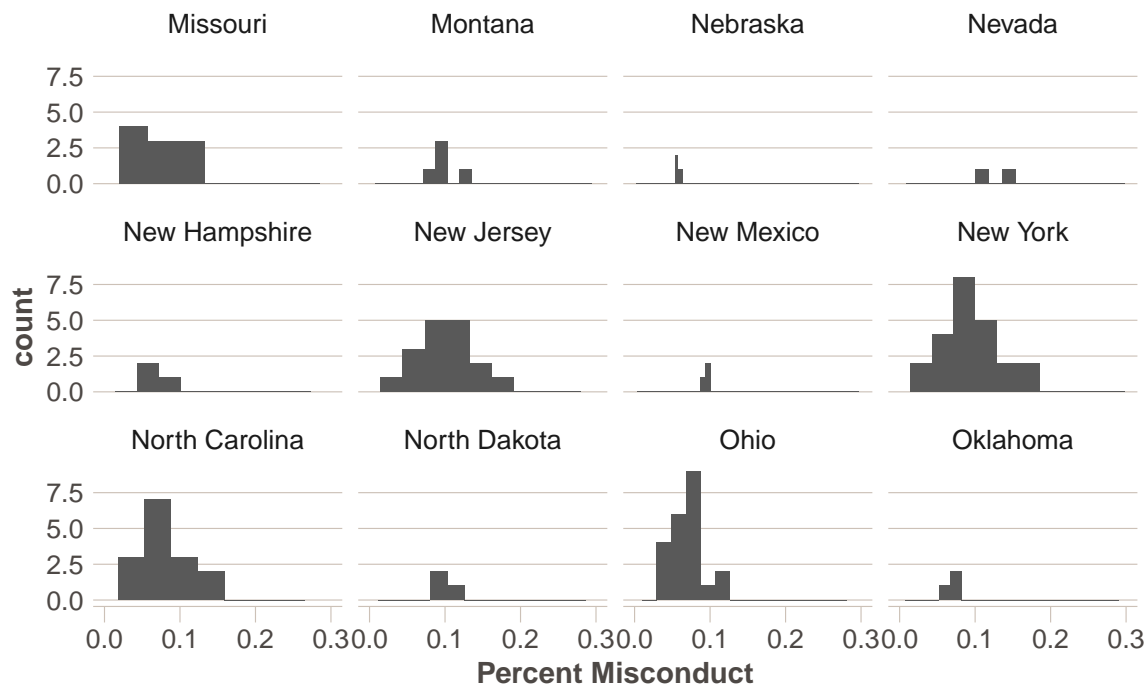
## Percent Misconduct by State



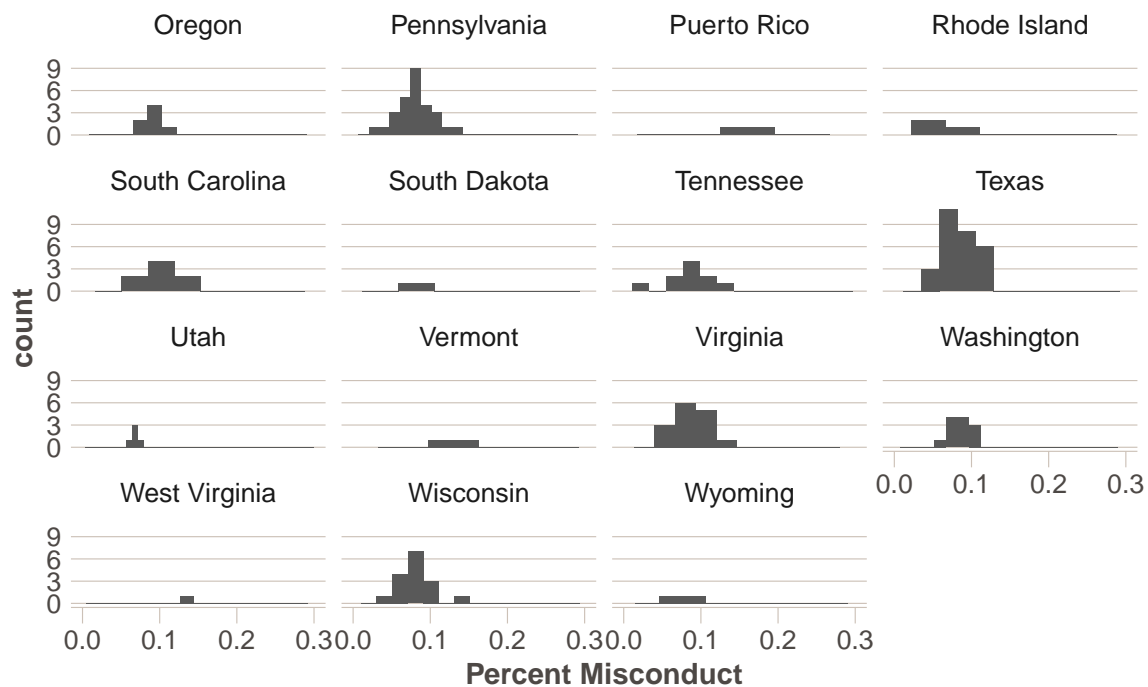
## Percent Misconduct by State



## Percent Misconduct by State



## Percent Misconduct by State



I have quite a strong prior belief that the distribution of misconduct is not inherently different across states. However, the fixed effects we estimate will probably vary greatly, especially since we have few observations per state. For example, we only observe 3 counties in all of Oklahoma; this may lead to a biased fixed effect

estimate for Oklahoma.

Table 1: Q2 Model: State FE

	<i>Dependent variable:</i>
	Percent Misconduct
nbrokers	−0.000003*** (0.000001)
sq_nbrokers	0.000000** (0.000000)
Observations	473
R <sup>2</sup>	0.349816
Adjusted R <sup>2</sup>	0.269317
Residual Std. Error	0.027364 (df = 420)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

In Table 1 I show the results of a state-fixed effects regression of misconduct on the number of brokers and the square of number of brokers, since we may expect that an increase in the number of brokers has a diminishing impact. The estimated broker coefficient is statistically significant but does not seem very meaningful; an increase of 5000 brokers is associated with a 0.0137 lower percent misconduct.

This coefficient is heavily skewed by outliers — for example, dropping New York county from the data leads to a 64% increase in the magnitude of the estimated coefficient.

## Question 3

### The Research Idea

Educational attainment may help explain misconduct at the county level. For example, it may be the case that better education is associated with greater ethics — this seems unlikely however, as college education is an intellectual pursuit and not a character-building endeavor; greater knowledge does not imply better ethical judgment. There may be some other factors that are correlated with both education and (mis)conduct; for example, stronger social ties in the community may be associated both with greater educational attainment and lower levels of misconduct. This, however, could be too subjective of a measurement and is in any case hard to measure.

The more likely and more straightforward relationship between educational attainment and misconduct is that counties with low education could see higher misconduct, not because less-educated brokers are committing misconduct, but rather that less-educated consumers may have less knowledge of how to report misconduct, less awareness that it is taking place, or fewer resources to challenge misconduct. Financial services in counties with a higher level of educational attainment may also be in wealthier areas; and these could be more competitive, attracting only the best brokers, with “lower quality” brokers defaulting to more remote locations that may have lower educational attainment. This is somewhat reminiscent of the claim made by Egan et al. 2016 that financial advisory firms who engage in misconduct are “cater(ing) to unsophisticated consumers”.

### The Data

I pull county-level educational attainment data from the Economic Research Service of the USDA. This includes the number of people at with a certain level of educational attainment (less than high school, high school diploma, some college, or bachelor’s degree and more), as well as percent of the population with a



certain level of educational attainment, from 1970 to 2019. It is hard to see why educational attainment in the current year would be directly related to the financial misconduct; if it is the case that firms locate in an area of lower educational attainment to target “unsophisticated consumers”, then this location decision would likely be made in advance. Since the county level data is from 2015, combined with the fact that we do not have educational data specific to 2015 but rather aggregated from 2015-2019, I choose to use educational data from 2000.

County data for the states of Alabama to Connecticut (in alphabetical order) is not available after joining by FIPS codes. This is likely because some counties are included in the original data for which the FIPS code does not match or does not exist in the educational data.

## Regression Results

With this regression, I hope to show if there exists a correlation between educational attainment and financial misconduct in US counties. The proposed hypothesis is that counties with lower educational attainment will show higher levels of misconduct.

The coefficients for number of brokers and for the percent of county residents with some college education are both significant at the 5% level. The effect of the number of brokers seems to be slightly dampened after adjusting for education levels in the county; it changes from  $-2.73 \times 10^{-6}$  to  $-1.94 \times 10^{-6}$ . For an increase of one percentage point in those with less than a high school diploma, there is an decrease of 0.11 percentage points of misconduct; in those counties with a higher fraction of residents who have lower than a high school education, there is a higher likelihood of financial misconduct. This may be misleading however, as areas with low educational attainment may simply be poorer areas that have lower demand for financial services; I should add county-level GDP or perhaps county-level poverty rates to the regression to account for such a possibility.

Furthermore, the estimated coefficients are not robust to the inclusion of different educational measures; after controlling for *some college*, for example, the coefficient for `pct_less_than_hs` is no longer significant, nor are coefficients of the other educational variables. Clearly, however, as these educational attainment categories are mutually exclusive and cover the entire county population, they are collinear; if a county has a low percentage of population with a bachelor’s degree or higher, it necessarily has a relatively higher population of people with some college, a high school diploma, or less. So, we have to have caution in simply adding more variables in the regression; this is an additional reason that I’ve chosen to focus only on those with lower than a high school education and those with higher than a bachelor’s degree.

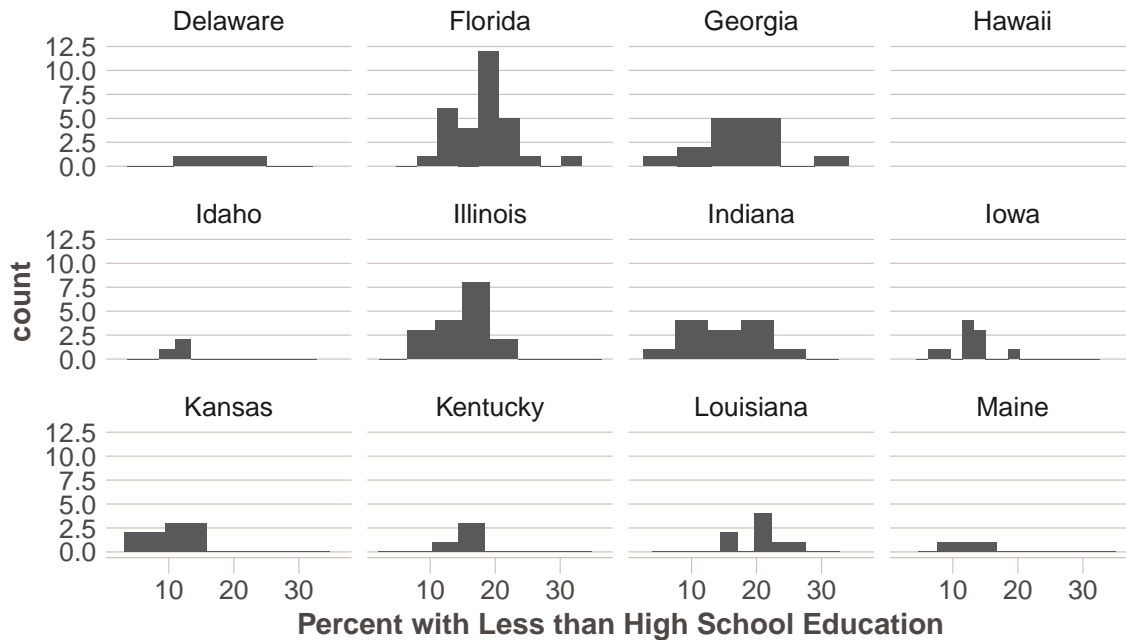
For counties for which we observe the percent with less than high school education, I display the distributions of county data by state:

Table 2: Q3 Model: State FE

	<i>Dependent variable:</i>
	Percent Misconduct
nbrokers	−0.000002** (0.000001)
sq_nbrokers	0.000000 (0.000000)
pct_less_than_hs	−0.001109*** (0.000402)
pct_bachelors	−0.000427 (0.000261)
Observations	415
R <sup>2</sup>	0.354814
Adjusted R <sup>2</sup>	0.272187
Residual Std. Error	0.027628 (df = 367)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

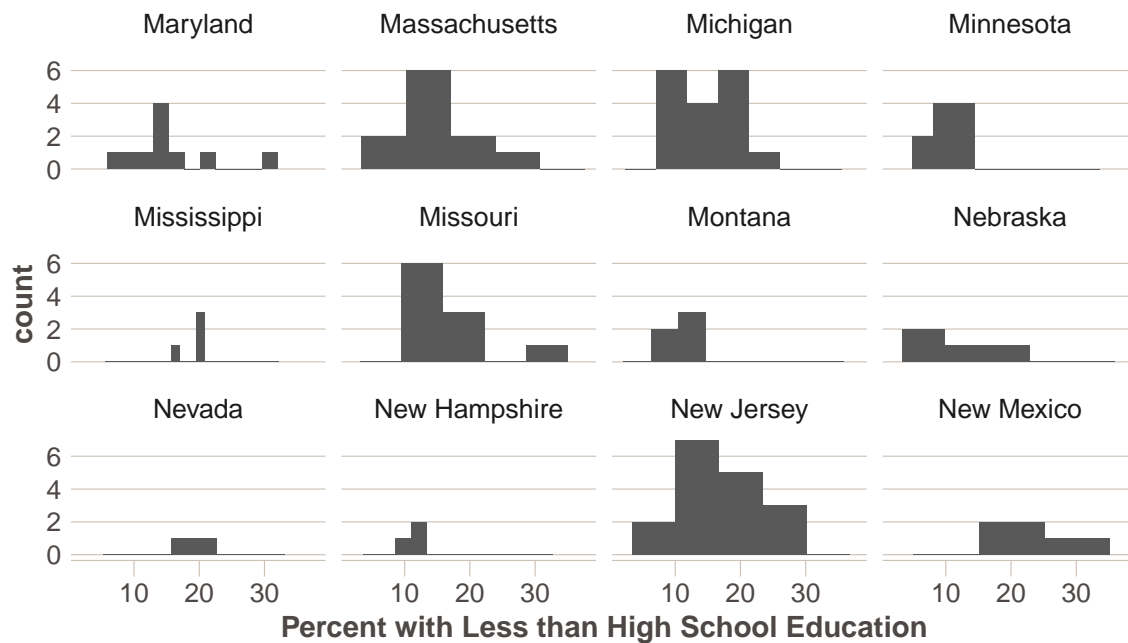
## Educational Distribution by State

### Distribution of Percent with Less than HS Education



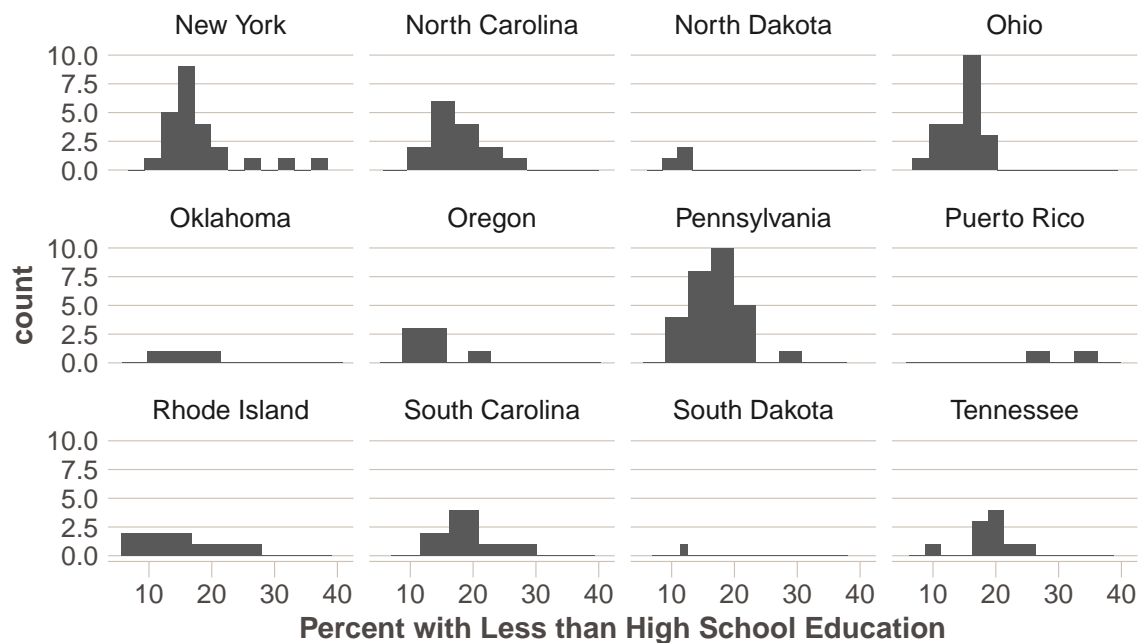
## Educational Distribution by State

### Distribution of Percent with Less than HS Education



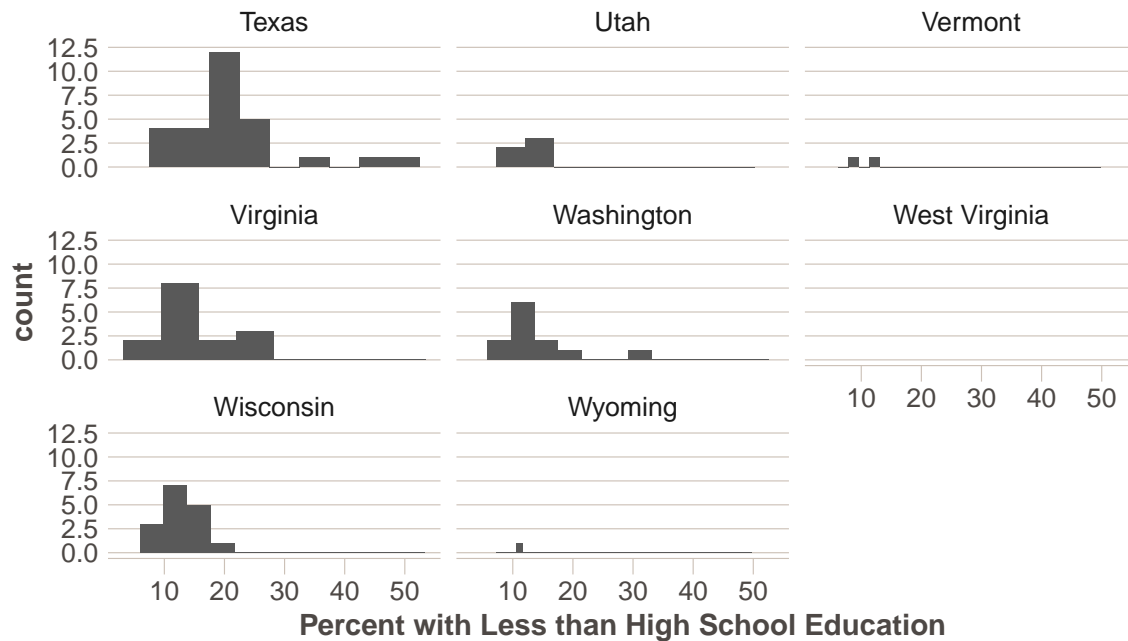
## Educational Distribution by State

### Distribution of Percent with Less than HS Education



## Educational Distribution by State

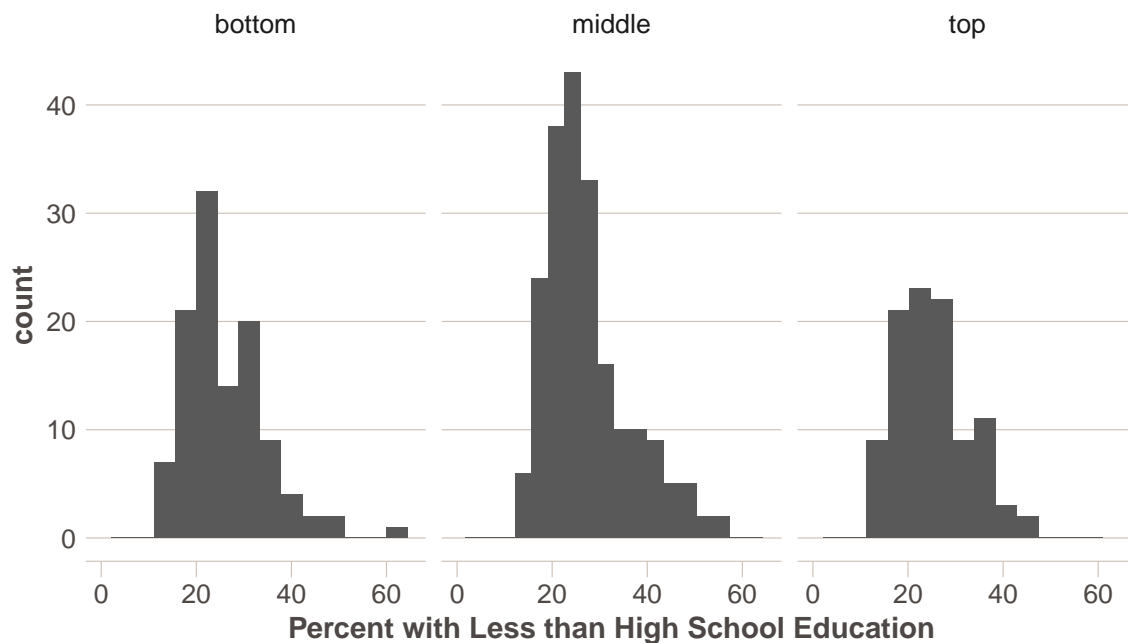
### Distribution of Percent with Less than HS Education



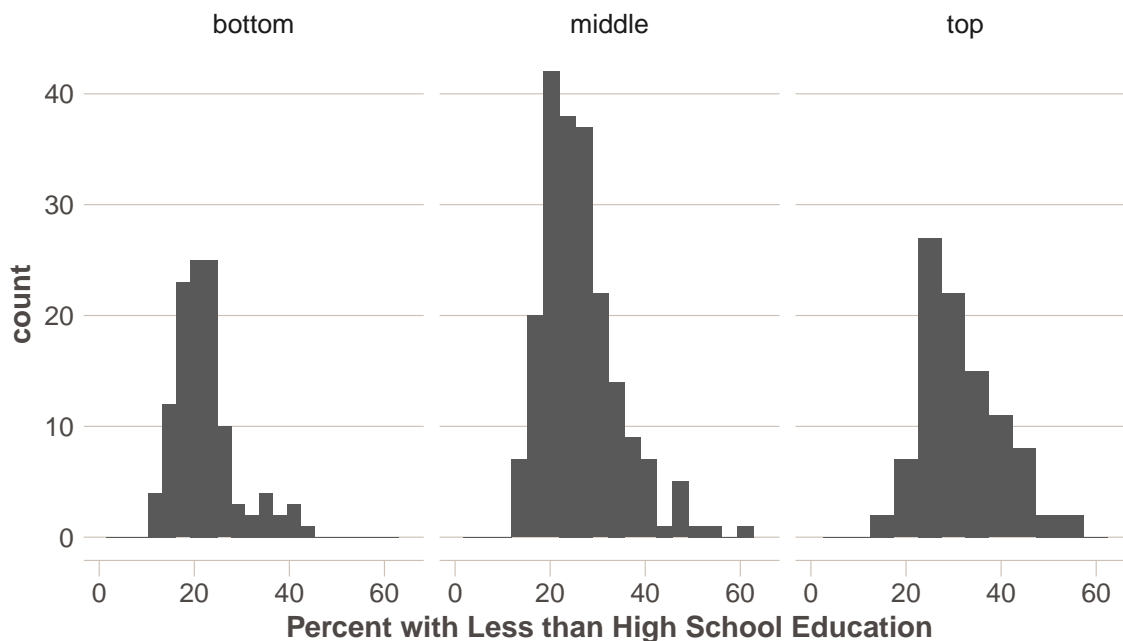
Additionally, I look at the distribution of percent less than HS education conditional on where the county falls in number of brokers or in percent misconduct:

## Distribution of Percent with Less than HS Education

### By Quantile of Percent Misconduct



## Distribution of Percent with Less than HS Education By Quantile of Number of Brokers



Notice that, conditional on greater than or equal to 40% of the population of a given county having less than a high school education, it is most likely that that county falls in the top quantile number of brokers.

## Question 4

### Research Proposal

These results are not very illuminating for my initial curiosity of how does educational attainment interact with financial misconduct. With this data, there is very little, if any evidence of a statistically significant or economically meaningful relationship between educational attainment levels and financial misconduct rates.

While the typical broker would have a Bachelor's degree in finance, business, economics, or other fields, there is no formal educational requirement to become a broker. From the current data, there is little evidence to support the idea that financial misconduct is more frequent in counties with lower educational attainment. However, we have not answered the question of whether brokers who have *ever* committed financial misconduct are more likely to later locate in these areas.

In Question 3, I suggested that perhaps “lower quality” brokers may go to less-educated areas — this could be a combination of the fact that areas with lower educational attainment are generally less developed as well and are home to fewer and less prestigious financial firms. In this case, a broker's financial misconduct could serve to financial firms as a signal that that broker is of low quality. If this is the case, we may expect brokers to begin their career in a highly developed area and/or at a large financial firm and to then move to a less developed area/smaller firm after being found guilty of misconduct, since the more influential financial firms would not be willing to hire them.

It would be insightful to have data on broker educational attainment, performance (returns), and career trajectories (relocations); the issue would be collecting this data which could span across multiple counties, states, and firms.

Every broker in the United States has a Central Registration Depository (CRD) number; this number may

change when they change positions, but is always tied to their unique ID. By using this identifier, we could track where particular brokers move over time — we could see for example if, after being found guilty of misconduct, a broker is likely to move to a smaller county, a smaller firm, or a county with lower educational attainment. Most of this data is already public, but the misconduct data we have is aggregated. I would need the time, area, firm, and CRD number associated with the misconduct, and with this data I could attempt to answer the question: after committing financial misconduct, are brokers more likely to relocate to an area with a lower level of educational attainment?

With data on broker education and performance, we could also attempt to answer the secondary questions: does the distribution of broker misconduct vary based on brokers' educational attainment ? Are brokers with a history of low performance more likely to later be found guilty of misconduct?

## References

Egan, Mark, Gregor Matvos, and Amit Seru. “The Market for Financial Adviser Misconduct.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, September 1, 2017. <https://doi.org/10.2139/ssrn.2739170>.

USDA ERS. “Educational attainment for the U.S., States, and counties, 1970-2019.” Accessed November 8, 2021. <https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/>.