”Rage against the Machines: Labor-Saving Technology and Unrest in Industrializing England” by Caprettini and Voth, AER, 2020
Replication and Simulations

Mateo Bortoli[a], Elie Malhaire[a]
[a]ENS de Lyon

Abstract

1 Context

The paper we replicate analyze the “Captain Swing” riots in 1830s England in order to investigate the impact of the adoption of new technologies (threshing machines) on social and political outcomes, such as riots.

The threshers were invented in 1786 and they spread rapidly from 1810 onwards. These machines replaced the work of many rural laborers, leaving them with little income in winter. This came on top of a wider movement to enclose land, which considerably reduced the amount of communal land, and thus reduced the proportion of income from the land that could be directly consumed by the peasants.

The first Swing Riots broke out in 1830 in Kent and then spread. In all, more than 3,000 riots broke out in 45 counties, and 514 threshing machines were attacked. The riots were eventually put down by the army and 252 people were sentenced to death.

2 Data

Diffusion of threshing machines:

- 60 regional newspaper over the period 1800-1830 : 549 new threshing machines on 466 parishes;
- complementary information from the General Views of Agriculture.

Unrest:

- Swing Riots data from the Family and Community Historical Research Society;
- British Newspaper Archive : 610 actual arson incidents and 69 attacks on machines between 1758 and 1829.

Soil composition, population:

- Geological Map of Great Britain;
- British population censuses of 1801–1831.

3 The model

$$\text{Riots}_p = \beta_0 + \beta_1 \text{Machines}_p + \beta_2 \text{density}_{p1801} + \beta_X X_p + \theta_r + \epsilon_p,$$

where:

- Riots_p: number of riots.

- Machines_p: number of machines introduced.
- density_{p1801}: population density in 1801.
- X_p: other explanatory variables.
- θ_r: region fixed effects.
- ϵ_p: error term.

3.1 Equations

```
# OLS with region fixed effects
model_OLS <- lm(SWING ~ thresh + log_density + agri_share +
                log_sex_ratio + log_distel + log_distnews +
                factor(REGION), data = swing_cross)


# IV with region fixed effects
model_IV <- ivreg(SWING ~ thresh + cer + log_density + agri_share +
                  log_sex_ratio + log_distel + log_distnews
                  |.-thresh + heavysh + factor(REGION), data = swing_cross)
```

The Instrumental Variable used in the paper is the heaviness of the soil. Heavy soils are negatively and significantly correlated with the introduction of threshold machines, and are only correlated to the riots through that channel.

3.2 Replication results

	Estimate	Std. Error	Significance
Intercept	1.60027	0.12530	***
thresh	0.38855	0.03795	***
log_density	0.10061	0.01307	***
agri_share	-0.06483	0.04433	
log_sex_ratio	-0.18142	0.06068	**
log_distel	-0.32473	0.01838	***
log_distnews	0.02237	0.01610	
Residual Std. Error	1.076 (df = 9667)		
Multiple R^2	0.05681		
Adjusted R^2	0.05623		
F-statistic	97.05 (df = 6, 9667)		
p-value	$< 2.2e-16$		
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 1: OLS Model Results

	Estimate	Std. Error	Significance
Intercept	1.47285	0.24405	***
thresh	6.36131	1.68165	***
cer	-0.18552	0.26772	
log_density	0.01049	0.03480	
agri_share	0.02418	0.08782	
log_sex_ratio	-0.03533	0.12200	
log_distel	-0.29406	0.04082	***
log_distnews	0.02505	0.03058	
Residual Std. Error	2.03 (df = 9666)		
Multiple R^2	-2.36		
Adjusted R^2	-2.362		
Wald Test	21.4 (df = 7, 9666)		
p-value	$< 2.2\text{e-}16$		
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2: IV Model Results

4 First simulation: OLS model

Using the shape of the data provided by the publication's authors, we simulate several key variables that are crucial to our analysis of social protest in parishes during the Captain Swing riots. Variables of interest include:

- Population density: logarithm of population density, sampled as rnorm(n, mean = 3.65, sd = 0.96).
- Sex ratio: logarithm of the sex ratio, sampled as rnorm(n, mean = -0.03, sd = 0.19).
- Share of agricultural workers in the parish: modeled as rgamma(n, shape = 2.11, scale = 0.18).
- Distance in km from the county of origin of the first insurrection: logarithm of this distance, sampled as rnorm(n, mean = 5.33, sd = 0.62).
- Distance in km from sources of information: logarithm of this distance, sampled as rnorm(n, mean = 2.95, sd = 0.73).
- Climatic conditions: include variables such as heavysh, an indicator for heavy machinery, modeled as rbeta(n, shape1 = 0.71, shape2 = 0.66).
- Soil quality: also related to agricultural productivity, using cer as a proxy, sampled as rnorm(n, mean = 0.63, sd = 0.1).
- Region: a categorical variable indicating the region, generated as sample(1:5, n, replace = TRUE).

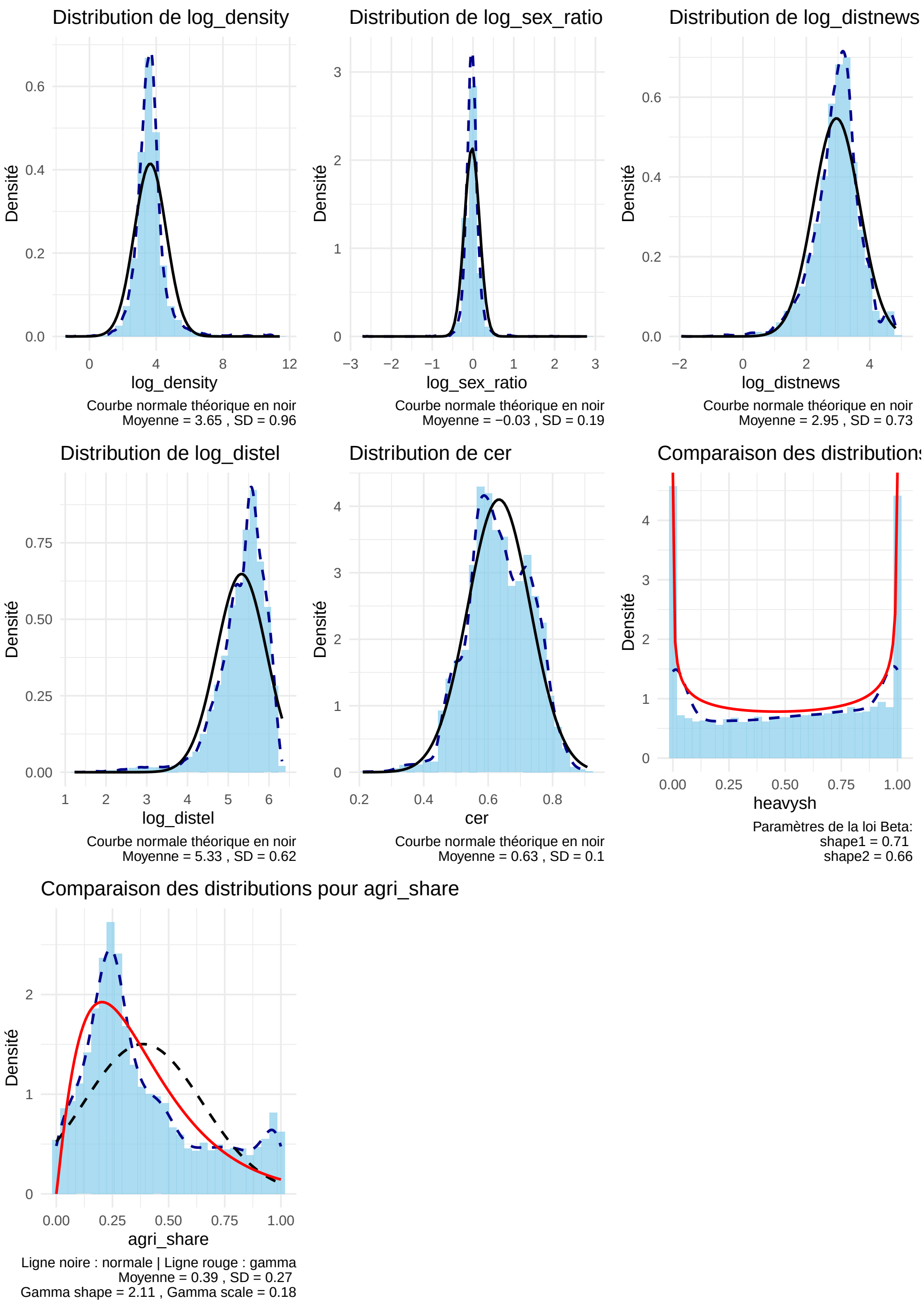


Figure 1: Distribution of the variables used during the replication

4.1 Strategy

In this first part of the simulation, we test the OLS regression with two different theoretical "true effects," which are the coefficients found by the authors in their OLS ($\beta_{\text{thresh}} = 0.35306$) and IV ($\beta_{\text{thresh}} = 6.557$) regressions. We then simulate the following equation with the two β_{thresh}:

$$\text{Riots}_p = \alpha_0 + \alpha_1 \text{Machines}_p + \alpha_2 \text{Log\_Density}_{p1801} + \alpha_3 \text{Log\_Sex\_Ratio} \quad (1)$$

$$+ \alpha_4 \text{Log\_Agri\_Share} + \alpha_5 \text{Log\_Distel} + \alpha_6 \text{Log\_Distnews} + u_p \quad (2)$$

It allows us to have a measure of the "predicted Riots" variable. Each simulation returns 5000 observations, then each is repeated 100 times. We use it to fully simulate our dataset, adding standard deviations to create heterogeneity in our treatment. We then run our regression again to find our two simulated β_{thresh}, enabling us to test them.

Indeed, the central issue in this analysis lies in assessing the true effect of machine adoption on the number of protests. Although both estimated coefficients in the publication (OLS and IV) are significant and positive, their values differ considerably: the effect determined by the IV approach is 16 times higher than that obtained by OLS. This divergence raises doubts about the robustness of the instrumentalization used.

The aim is to understand what factors might be responsible for this potential exaggeration.

4.2 Results

Table 3: Résultats pour $\beta_{\text{thresh}} = 0.35306$

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Estimates	0.1125	0.2729	0.3446	0.3478	0.4155	0.6076
P-value	0.0000	0.00002187	0.0004284	0.01057	0.005626	0.2291

Table 4: Résultats pour $\beta_{\text{thresh}} = 6.557$

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Estimate	6.308	6.485	6.551	6.550	6.610	6.755
P-value	0	0	0	0	0	0

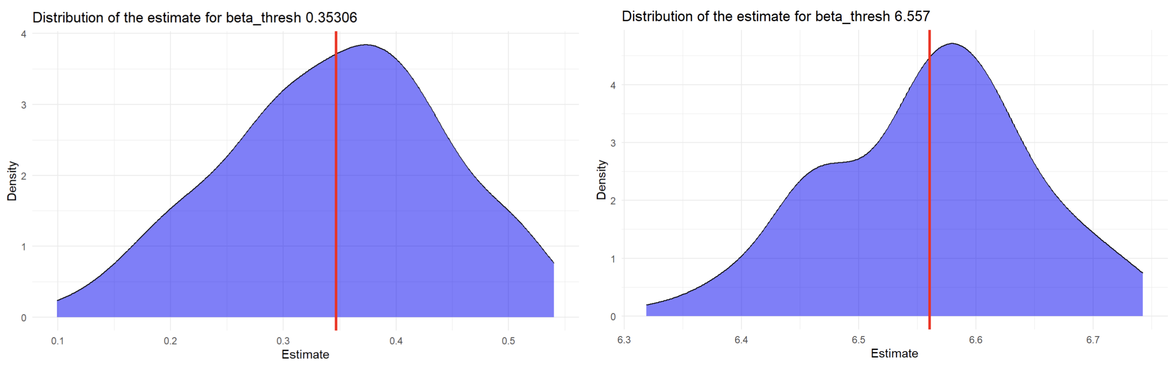


Figure 2: Distributions of the two beta estimates (from OLS and IV coefficients)

For the OLS beta, the coefficients are consistent with the implemented true effect. The mean coefficient is 0.3478 and the significance is high on average, although there is still some variability in this case.

For the IV beta, the estimated coefficients are very stable and the mean of 6.550 is consistent with the true effect. It makes sense that the p-values are all at zero given that the implemented effect is very large.

However, this can also lead us to believe that the effect is greatly exaggerated. Indeed, the statistical power is an increasing function of precision and effect size, and then, when the theoretical effect is overestimated, the statistical power becomes so high that even minor estimation errors are systematically significant. It could be that the instrumental variable used in the paper captures effects that are not those questioned by the paper, and that would mechanically increase the magnitude of the coefficients.

In parallel, we implement a test from the retro-design package of Gelman et al's, that give us the following results:

Table 5: Retrodesign results for $\beta_{\text{thresh}} = 0.35306$

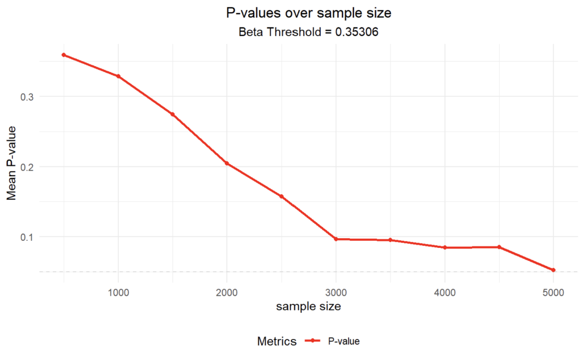
Metric	Value
Power	0.9443
Type S Error Rate	1.88×10^{-8}
Type M Error (Magnification Ratio)	1.0382

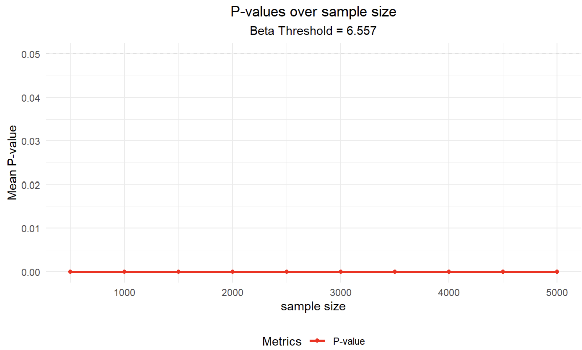
Table 6: Retrodesign results for $\beta_{\text{thresh}} = 6.557$

Metric	Value
Power	1.0000
Type S Error Rate	0
Type M Error (Magnification Ratio)	1.0001

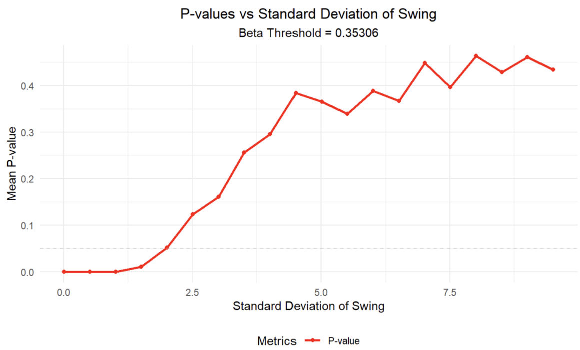
In order to analyze our p-values in more detail, we then make our DGP more complex:

- by varying the sample size:





- by varying the standard deviations used for the creation by simulation of the SWING variable (Riots):



It appears that, for the simulation based on the true effect of the OLS, the sample size must be at least 3000 observations and that the standard deviation applied to our dependent variable must not exceed 2.2 in order to reject the null hypothesis.
Concerning the simulation based on the true effect of the IV, the original coefficient is so large that neither the sample size nor the standard deviation influences the p-value, which remains constantly zero. It is then appropriate to investigate whether or not we are facing exaggeration.

5 Second simulation: IV model

5.1 Data Generating Process

5.1.1 Endogeneity and random noise simulation

We're now going to our data generating process more complex. Indeed, threshers and riots may be endogeneous. This means that their idiosyncratic error term is not random but correlated together. To model this issue, we use a two-dimensional multivariate normal distribution. We simulate random noise in the thresh generation, which follows a 0-centered normal distribution with variance 1. Then we simulate a noise correlated with it in the generation of the SWING variable. The coefficient of correlation is 0,5.

5.1.2 Model Setup

The outcome variable ($SWING$) is modeled as a function of the treatment (*thresh*), covariates ($\mathbf{X}$), and a normally distributed error term. correlated with the thresh errors. The functional form ensures positivity by taking the exponential of a scaled linear combination of predictors enabling to avoid excessive large or small values.

$$\text{SWING} = \exp\left(\frac{\beta_0 + \beta_{\text{thresh}} \cdot \text{thresh} + \beta_{\mathbf{X}} \cdot \mathbf{X} + \text{error}_2}{4}\right)$$

Where, β_0 represents the intercept, β_{thresh} is the coefficient of interest measuring the causal effect of *thresh*, and $\beta_{\mathbf{X}}$ are coefficients for covariates.

5.1.3 Instrument variable and first stage Probit Model

Soil quality is a factor that may have played a role in the introduction of new machines as heavy soils make them useless or too costly, on the other hand, the authors hypothesize that there is no link between soil heaviness and protests. The first step is to simulate the variable that measures the quality of soils by computing the share of heavy soils in a parish. This variable is in the interval between 0 and 1, since it is a percentage. As the replication work has shown, the probability density is concentrated on the edges of the definition interval [0;1], which leads us to use a Beta distribution with parameters alpha = 0.71 and beta = 0.66.

The thresh variable is constructed in two stages. It is a linear combination of soil quality weighted by a negative and weak coefficient (in the replication the regression coefficient between these two variables is -0.037, so we choose a coefficient of -0.1) and a stochastic component (the first column of the variance / covariance matrix) which reflects the impact of unobserved or random factors on machine introduction. This continuous variable is then transformed into a binary variable, respecting the proportion present in the authors' original data where for 95% of observations there is no threshers. Indeed, it allows us to take into account the non-linearity of this variable: you can't own half a machine. To make it simpler, we've turned it into a binary variable that takes the value 1 if the continuous variable exceeds the 95th percentile, and 0 otherwise.

First stage probit model: To address potential endogeneity, we first estimate the probability of receiving treatment ($thresh = 1$) using a Probit regression model with the instrument (*heavysh*) and observed covariates ($\mathbf{X}$). The Probit model is given by:

$$P(\text{thresh} = 1 \mid \text{heavysh}, \mathbf{X}) = \Phi\left(\beta_0 + \beta_{\text{heavysh}} \cdot \text{heavysh} + \beta_{\mathbf{X}} \cdot \mathbf{X}\right)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. This step produces predicted probabilities $\widehat{\text{thresh}}$ to be used in the second stage.

5.1.4 Second Stage: IV Regression (2SLS)

In the second stage, we estimate the causal effect of *thresh* on $SWING$ using a linear regression model. The predicted values of *thresh* from the first stage serve as instruments to address endogeneity. The regression model is specified as:

$$\text{SWING} = \beta_0 + \beta_{\text{thresh}} \cdot \widehat{\text{thresh}} + \beta_{\mathbf{X}} \cdot \mathbf{X} + \epsilon$$

Where, $\widehat{\text{thresh}}$ represents the fitted values from the first-stage Probit model, and ϵ is the error term. This two-stage least squares (2SLS) approach provides a consistent estimate of the causal effect of *thresh*.

5.2 Analysis

First stage Our first objective is to evaluate the first stage of our IV, i.e. how the heavysh instrument predicts the probability that the dependent and binary variable thresh is equal to 1. The probit model is the most suitable for modeling such variables, as it allows predicted probabilities to be bounded between 0 and 1. We then calculate McFaden's pseudo $R\hat{2}$, which is a measure of fit for probit models. The higher the $R\hat{2}$, the greater the thresh variance explained by the model, i.e. the relationship between heavysh and thresh is significant.

Our results show that with an R2 close to 0, the quality of soil is a poor instrument for estimating threshers. It could be either the results of the low prevalence of thresh = 1 in the simulated sample (only 5%) or the weakness of the true relationship between heavysh and thresh, which we argue is -0.1 (yet almost three times bigger than the one they find in their study: where the estimated coefficient between thresh and heavysh was -0.037). A graphical analysis clearly shows the small difference between parishes where the soil is more heavy in terms of threshers.

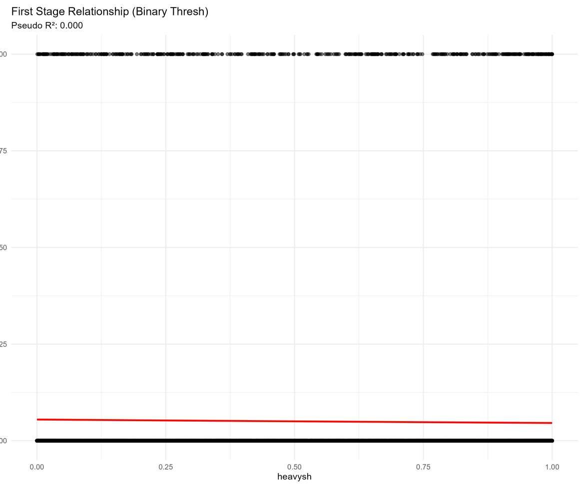


Figure 5: first stage regression

Second stage In this second section, we focus on statistical analysis. As expected from the weakness of the instrument, our simulated data fail to show the true effect, whether we take the true effect of the OLS or the much larger effect of the IV: the model's variance is extremely large as we can see below:

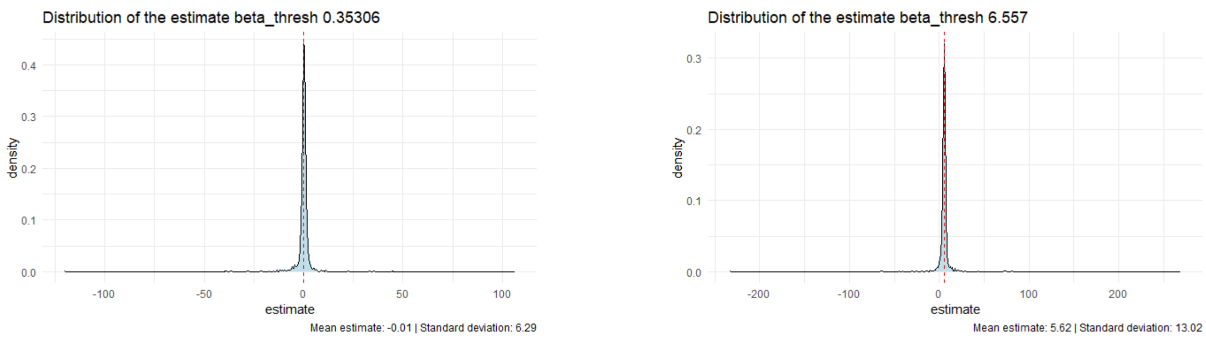


Figure 6: Distributions of estimates (true beta from OLS and IV coefficients)

The results from the retro design analysis show very low power values (around 5%), a Type S error rate close to 0.5 (indicating a high probability of getting the sign wrong), and a very high Type M factor (meaning a significant overestimation of the true effect).

Table 7: Retrodesign Results for β_{thresh}

$\beta_{\textbf{thresh}}$	**Power**	**Type S Error Rate**	**Type M Error (Magnification Ratio)**
0.35306	0.05000001	0.4996603	8037.562
6.557	0.0509768	0.3937331	25.11011

6 Alternative Scenario: does gender influence riots?

To further explore our theoretical model, we want to study the relationship between the gender variable and protests. Monte-Carlo simulation simulation is used to estimate how the gender ratio influences the riots.

We use the OLS Data Generating Process (Part II), which we feel is the most relevant and closest to reality. To simulate a gender ratio ranging from almost 100% female to almost 100% male, we vary the mean of the binomial distribution simulating it from -5 to 5. As our variable of interest is the logarithm of the gender ratio, this ensures this result.

As we do not know the true effect of the gender on social unrest we decide to also vary this estimator between -1 and 1. We then explore the sensitivity of SWING to variations in this ratio and in the true effect of gender. Figure X display the 3D representation. As we can see there is no pattern visible. With our modification it seems that gender ratio does not have any clear effect on social unrest.

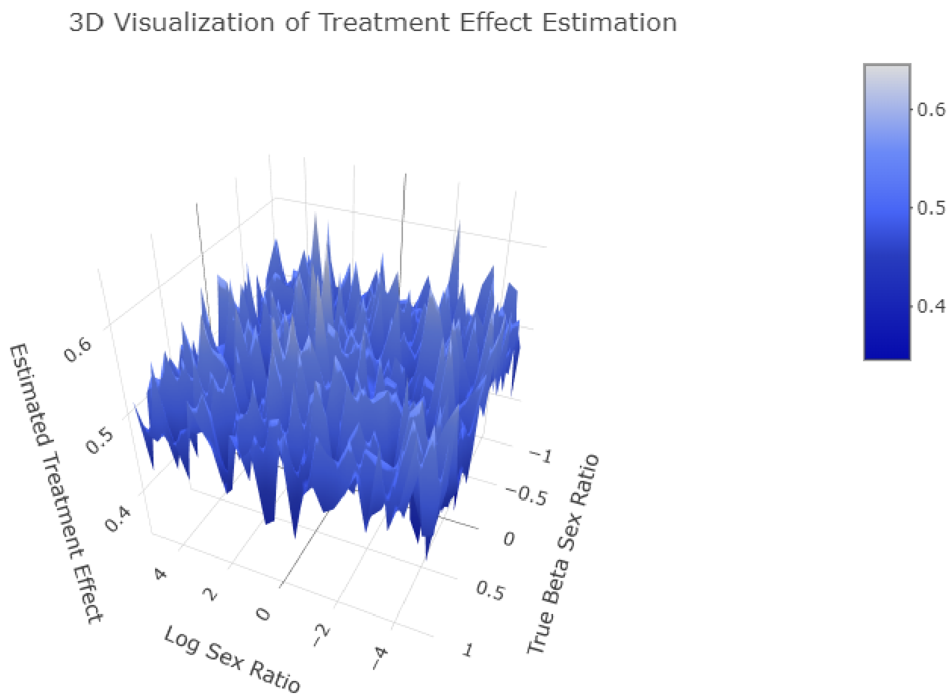


Figure 7: Distribution of estimates over gender ratio and gender true effect