

AN ANALYSIS OF US WORK STOPPAGES AND EMPLOYEE WAGES

NYSSA CORNELIUS AND CHRIS JENNINGS-SHAFFER

1. INTRODUCTION

The GitHub repository for this project is located at <https://github.com/NyssaCornelius/FProj>. The Binder link for this project is [Binder link](#).

The data sets are taken from the U.S. Bureau of Labor statistics (<https://www.bls.gov/>). For work stoppages, we used the excel file at <https://www.bls.gov/web/wkstp/monthly-listing.xlsx>. This file contains information on each work stoppage in the US from 1993 on that involved at least 1000 workers. Information about the columns is given in the following table (all columns are stored as text).

column	basic type	description
Organizations involved	nominal	The company or branch of government where the work stoppage occurred.
States	nominal	The states where the work stoppage occurred.
Areas	nominal	The geographic areas where the work stoppage occurred.
Ownership	nominal	This is either private industry (private sector) or state/local government (public sector).
Industry code	numerical	The 2017 NAICS code describing the industry associated to the work stoppage.
Union	nominal	The name of the union associated with the work stoppage.
Union acronym	nominal	The acronym for the union.
Union Local	nominal	A number or word description of the local union associated with the work stoppage.
Bargaining unit	nominal	A description of the bargaining unit for the work stoppage (usually empty).
Work stoppage beginning date	ordinal	Start date of the work stoppage.
Work stoppage ending date	ordinal	End date of the work stoppage.
Number of workers	numerical	The number of workers involved in the work stoppage.
Days idle cumulative for this work stoppage	numerical	The total duration of the work stoppage.
Note	nominal	Any notes about the work stoppage (e.g., the union changed names or the number of workers involved changed during the work stoppage).

Employment statistics at the national level are taken from <https://download.bls.gov/pub/time.series/ce/>, specifically from `ce.industry`, `ce.series`, and `ce.data.0.AllCESSeries`. Since a full description of these data sets are given in `ce.txt`, let us be a bit brief in the description. The file `ce.series` contains an entry for each type of data stored at the national level (for example: “Average hourly earnings of all employees, stone mining and quarrying, seasonally adjusted”), whereas the actual data values for this are stored in `ce.series`; they are joined on the column `series_id` (e.g., `CES1021231003`). The file `ce.series` also contains a column for `industry_code`, which is matched to the description of the industry code in `ce.industry`. In `ce.industry`, there is partial information to match an industry code of `ce.series` with an NAICS industry code of work stoppage data. This is discussed in greater detail in Section 3.

Similarly, employment statistics at the state level are taken from <https://download.bls.gov/pub/time.series/sa/>, specifically `sa.series`, `sa.data.0.Current`, `sa.industry`, and `sa.state`, with a detailed description of the data sets given in `sa.txt`. The file `sa.series` contains an entry for each type of data stored at the state level and the actual data for the entries are stored in `sa.data.0.Current`, again joining on the column `series_id`. We use `sa.state` to decode the state for a series from numeric to string. However, this time `sa.industry` does contain even partial information to match industry codes, but it does at least give an English description of the codes. This is also discussed in greater detail in Section 3.

[...still need to add explanation for min wage data used in chloropleths ...]

The choice of using these data sets is that they contained the greatest amount of information we could locate. Upon reviewing the literature, they are the standard sets to use (and have been for a long time). To load in the data, we downloaded the text files and then imported them into Pandas dataframes in our jupyter notebook. The question we ask and answer in this project is the following. With the major economic and political changes of the last few decades, do strikes still have a negative correlation with income inequality? Specifically, do strikes correlate with an increase in wages of the associated workers? How do geographical regions affect the frequency of strikes? Does the minimum wage play a role in the frequency of strikes? The inputs for this project are the data sets described above, the outputs are the data visualizations and conclusions that we draw below.

The remainder of this report is as follows. In section 2 we briefly discuss some of the existing literature (both research and non-research articles) on the subject, which will put our project in perspective of the larger picture. In section 3 we go over the data cleaning necessary for our project and the challenges that this presented. In section 4 we both describe how we visualized the data as well as give some of the most relevant visualizations. We conclude this report in section 5, where we make our concluding remarks.

2. LITERATURE REVIEW

Articles discussing work stoppages are almost exclusively about strikes (as shut outs initiated by management are very rare in comparison) and usually focus on union representation of the work force. While this last part is usually true, it is not universally so, as seen in [Rub88] where the author argues that strikes and unions do not go together in terms of correlation with employee wages and wealth inequality. In particular, they make the case that union representation decreases income inequality among specific groups but increases overall inequality, whereas strikes decrease overall income inequality (e.g. union representation may reduce income inequality among low and middle income white families while increasing income inequality between black and white families, whereas strikes reduce income inequality at the aggregate level). That being said, this article is an outlier in this respect and considers data only from 1949 to 1976. In the rest of our analysis, we take the more common approach of grouping unions and strikes together.

We will not spend much time on non-research literature, as for this subject it tends to be very clearly biased. However, we do note that while opinions that are pro-union and pro-strike do sometimes cite relevant statistics, this is far less the case with the opposing side. The latter is largely relegated to opinion pieces in local newspapers and blogs (and so we do not include references to them). While the pro-labor side also appears in similar sources (which we also omit), non-research literature also comes out of think tanks such as Economic Policy Institute and Washington Center for Equitable Growth. These articles are more likely to back up their claims with data and references (although they are often self-citing). For example, in [Bah19] the author makes the case that strikes have and still do empower workers and reduce economic inequality. The article discusses some of the political history surrounding unions and strikes in US, such when unions suffered a serious blow in 1981

when then President Ronald Reagan fired 11,000 air traffic controllers for striking for higher pay and reduced hours; describes summaries of polls on what workers do and do not like about unions and strikes [HF19]; and gives specific examples of recent strikes that have and have not paid off for workers [Yan18]. While not statistical data, the article also describes some of the current anti-union practices of major corporations (see [Har14]). Similarly, while the opinions in [SP20, PMS21] are not always backed by hard data, there the authors do correctly point out trends in the number of strikes and issues with the main source of data for work stoppages. The main source of data for US work stoppages is the U.S. Bureau of Labor Statistics (and this has been the case for over a century) and one issue with the data is that it only includes work stoppages that involve at least 1000 workers. As the authors point out, according to the Bureau of Labor Statistics, nearly 60% of workers in the private sector are employed by companies with fewer than 1000 workers. Additional issues with this data set are regularly brought up in research articles.

There are of course think tanks that are not pro-union, but their arguments avoid statistics of economic inequality and the historical correlation with unions and strikes. One example of this is [Eps20] from the Hoover Institution on War, Revolution, and Peace. Here the author argues against unions with pro-capitalism claims of free trade and competition benefiting the worker. For statistics, the author instead looks to the overall performance of the US economy and unemployment levels [Fit19, Coh19, Has20]. Another approach, as seen in [Wat14] is to examine economic inequality in terms of skilled labor and education of the workforce. However, the authors of [Mis18] make the claim that the data and analysis in [FHKN18] shows that the levels of training and education do not adequately explain include inequality. However, our project is not investigating statistics about these other sources.

In terms of research articles, strikes and unions are studied in many different respects. Lighter on the numbers are studies on public opinion and ethical issues. In [BK70] the authors give a lengthy account of the opinions for and against the legality of strikes in the public sector strikes. Generally speaking, most people at the time of the article accepted that private sector employees should be able to strike, but were divided when it came to the the public sector. Claims against public sector strikes are based on public sector work being essential, that the cost of increasing the collective bargaining power of public sector workers is higher and with lower returns than in the private sector, and that public sector strikes are inappropriate because they may affect public policy. Also there is the dubious claim (which is disputed by the authors [BK70]) that low pay in the public sector, such as for teachers, is due to public opinion on the importance of the service, whereas low pay in the private sector reflects a misallocation of resources. Ultimately the authors take the stance that strikes should be legal for areas of the public sector that would not lead to immediate public danger (e.g., fire fighters should not have the option to strike as part of their collective bargaining). Related to the issue of public sector strikes, [TS06] examines the ethical issue of medical workers being able to strike, which is assumed to be an issue of major importance as doctors move away from autonomous positions to the employee-employer model common to modern health care. Another common topic about the public sector is teacher strikes. By examining one specific case study, the authors of [BW10] draw the unsurprising conclusion that long term teacher strikes have long term (decades long) negative affects on students, but enter the political realm by implying this should be used as a talking point against the legality of teacher strikes. In [HFNR21] the authors examine public opinion polls about teacher strikes and education unions, and make the claim that public opinion is generally pro-labor, with first hand knowledge of the strike greatly increasing this (i.e., parents are more likely to side with their children’s teachers than to blame the teachers for the strike).

Taking statistics in mind, there are a large number of articles attempting to study and model the occurrence of strikes [Ken86, Ken85, Mau82, Nap87], usually based on strikes appearing in

waves over time. While present in articles over a century old [Cro08], there has been an increased interest in how relevant the recorded data actually is and how appropriately it is being analyzed. For example, in [PP74] it is pointed out that the data from the Bureau of Labor statistics records the the number of days of the work stoppage and the number of people involved, but this does not properly measure the actual cost of the strike. The authors attempt to correct this by examining the work force involved in each strike and weighting the strike with a cost associated to the particular type of work. Related to the issue of actual economic cost, the authors of [McH91] attempt to measure the impact of a strike based not just on the individual firm, but also on the cost to associated businesses. The authors of [SB87] suggest that the the variance in strikes is better explained by the cost of losing a job than the unemployment rate, which is commonly used. For the modern era, it is suggested in [MD10] that unions should be studied separately for institutionalized unions and social movement unions, whereas the authors of [KT89] think that the old models and explanations are no longer relevant to due to changes in business practices and public policy stemming from technology and globalization. There is also the claim, as seen in [WLR99], that strikes and unions are now irrelevant to income inequality.

The point to take from this is that historically it has been widely accepted that union representation and strikes correlate with improved conditions for the labor force. It is also accepted that in recent decades, union representation and the number of strikes have significantly declined. This is attributed to the change in power held by companies due to technology, globalization, and political power. In 2018 and 2019 there was a sudden and major uptick in strikes, but like most things this slowed in 2020 due to covid-19. Our research is to verify that despite this loss in power of the labor force, strikes do indeed still correlate with positive gains, as this should be verified and not be taken for granted. [...To be changed, if our data doesn't show that...]

3. DATA CLEANING

First we detail the basic data cleaning and type conversion done when importing the data sets. As mentioned earlier, all of the data is initially stored in plain text format. For the work stoppage data, the columns of interest are the states, industry code, work stoppage beginning date, and work stoppage ending date. We specify that industry code is an int, states is a string and that missing values should be the empty string, and otherwise we leave everything up to pandas. The national level data comes from `ce.industry`, `ce.series`, and `ce.data.0.AllCESSeries`. For `ce.industry`, we load this csv file directly with pandas with tab as the column separator. For `ce.series`, there are some issues with white space as the separator when loading this file directly with pandas, because of this we specify the column names explicitly and use the string strip method as a converter on the `series_id` column. Because `ce.series` contains a large amount of data not relevant to this project, we restrict to entries corresponding for weekly earnings of employees (`data_type_code` 11) and use the seasonally adjusted data (season 'S'). The data in `ce.data.0.AllCESSeries` has the same white space issue as `ce.series` and we handle it in the same way. The state level data comes from `sa.series`, `sa.data.0.Current`, `sa.states`, and `sa.industry`. The data in `sa.series` is white space delimited and this is handled properly by pandas. We restrict to entries for average weekly earnings (`data_type_code` 4). The data in `sa.data.0.Current` is white space delimited and this is handled properly by pandas. As `sa.state` is a short file, we record the information as a dictionary and do so by hand since this file translates an integer state code to the full state name, but the work stoppage data records states as their abbreviations. The file `sa.industry` is not loaded in the notebook and is instead something we use for data cleaning done by hand, which we discuss below. Despite the similar information being stored at the national and state level and the similar naming conventions, the actual data follows different conventions.

[...to add description of minimum wage data...]

For this project, missing data usually means that it is either completely unavailable or difficult to find. We do not encounter missing data due to bad or missing readings where we could try to fill in based on nearby data. To explain the ways this can happen, we separately discuss data at the national level and data at the state level.

At the national level we want to join the work stoppage data with the national data based on industry and then look up wage data for that industry around the time of the work stoppage. This leads to missing data in three consecutive stages. The work stoppage data uses NAICS codes and the national data uses something not well documented. The file `ce.industry` provides partial information to match these codes, with a `ce` industry code having none or more associated NAICS codes. At least for the subset corresponding to our data, an NAICS code is matched with at most one `ce` industry code. This process leaves a little over 100 unmatched NAICS codes in the work stoppage data that can still be match with `ce` industry codes. This additional matching is done by hand for two reasons. The first is it is a small enough number to easily do by hand and can be done by looking at the description of the NAICS codes and the description of the `ce` industry code. For example, the NAICS code 721110 is for “Hotels (except Casino Hotels) and Motels” and the `ce` industry code 70721110 is for “Hotels and motels, except casino hotels”, but the `ce.industry` file only matches 70721110 with the NAICS code 72111 (which is another valid NAICS code for “Hotels (except Casino Hotels) and Motels”). This example might suggest that the NAICS codes are embedded as substrings of the `ce` industry codes, but this is often not the case. The `ce` industry code 31327390 matches the NAICS codes 32731, 32733, and 32739, of which only one is a substring; the `ce` industry code 20000000 matches the NAICS code 23, which is not a substring of 20000000 but is a substring of a large number of `ce` industry codes. An NLP approach would yield matches, but it is very questionable that these would be accurate matches. If the process used to construct the `ce` industry codes was not a black box, then a programmatic approach would very likely work.

Once the industry code for a work stoppage is matched (either via `ce.industry` or by hand), there are two additional issues. A valid `ce` industry code may not have any associated `ce` series entries (i.e., the Bureau of Labor Statistics records that industry code, but tracks no wage data at the national level for that industry). Once the `ce` industry code is matched to a `ce` series, we can check for data on the `ce` series in `ce.data.AllCESSeries`. Often there is no data for the `ce` series around the time of the work stoppage. We detect this by using the work stoppage begin and end dates, the `AllCESSeries` year and period columns (period is basically month), and doing calculations with time deltas. For this project, we require wage data six months before the work stoppage began (the six months is set in a variable that can be changed to whatever time delta we want). Generally speaking, once wage data is collected for a `ce` series, it is collected in the future. This means we run into missing wage data for a work stoppage in that BLS did not yet record any data, but we do not encounter missing wage data where only a select number of months are missing (i.e., something we could try to fill in). For work stoppages that get past these three issues of missing data, the wage data is generally clean and we have valid time series to analyze.

In terms of information, missing values at the state level are similar to that at the national level, but are rather different in terms of data. The main difficulty is that `sa.industry` does not give any matches of `sa` industry codes with NAICS codes. Although the documentation in `sa.txt` says that the `sa` industry codes are SIC codes, this is simply not the case. To make this clear, SIC codes are four digits but `sa` industry codes are six digits. As a specific example, “Bituminous Coal And Lignite Mining” is `sa` industry code 112203, NAICS code 212111, and SIC code 1221. The reason this is so disappointing is that there are well established crosswalks for NAICS and SIC codes (<https://www.naics.com/wp-content/uploads/2014/10/NAICS-to-SIC-Crosswalk.pdf>). We match NAICS codes with `sa` industry codes by hand based on the descriptions. If the construction of `sa` industry codes was not a black box, then a programmatic approach would likely work. Once

a work stoppage is matched to an sa industry code, there are the same issues with the national level data. That is to say, an sa industry code may not have any associated sa series entries and an sa industry code that matches an sa series entry may not have data around the time of the work stoppage. As additional issue with missing data is that an sa industry code may have an associated sa series entry with wage data, but not in the state of the work stoppage (i.e., the Bureau of Labor Statistics recording data for an industry in one state does not mean they collect data for that industry in all states). For work stoppages that get past these missing value issues, we generally have clean time series data.

The process of connecting a work stoppage with wage data requires quite a bit of feature engineering and calculations as discussed above. Additional feature engineering and statistical summaries comes directly after this. For a give work stoppage, we have national level wage data as a time series and state level wage data as a time series. With this we compute the data for the visualizations discussed in the next section.

4. VISUALIZATIONS

...

5. CONCLUSIONS

...How did this project show us the information?...

...Did this give the results we wanted?...

...Future research...

In terms of future research, there are several obvious directions based on additional available data and the granularity of the data. These are in line with ideas mentioned in the existing literature. The work stoppage data includes information about the number of workers and the duration of the work stoppage, so we could analyze how these numbers impact the benefits to the striking workers. We matched NAICS industry codes with industry codes from the BLS data, so we could group similar industries (both on the work stoppage side of the data and on the BLS side) to look for patterns there. Related to the previous two ideas, we could try to account for the cost of that type of labor. Since we have information on the unions associated with the work stoppages, we could also look for patterns when grouping them by private versus public sector and also by institutional versus social movement unions. We used data at the national and state level, there is similar data available for metro areas (<https://download.bls.gov/pub/time.series/sm/>) that we could also analyze. This would require matching industry codes similar to the state level data. Our analysis considered wage data, but there is other relevant employment data such as employee injuries and illness (<https://download.bls.gov/pub/time.series/cd>, CF, CH, FI, FW, HC, HS, II, SH, SI), employee benefits (EB), unemployment and job openings (JL, JT, LA), and various costs to the employer (CC, CI). While there is a wealth of data available, each individual data set may require us to match industry codes, which is a time consuming process, especially since we cannot even trust the documentation of these data sets.

REFERENCES

- [Bah19] BAHN, Kate: *The once and future role of strikes in ensuring U.S. worker power*. <https://equitablegrowth.org/the-once-and-future-role-of-strikes-in-ensuring-u-s-worker-power/>. Version: 2019
- [BK70] BURTON, John F. ; KRIDER, Charles: The role and consequences of strikes by public employees. In: *The Yale Law Journal* 79 (1970), Nr. 3, S. 418–440
- [BW10] BELOT, Michèle ; WEBBINK, Dinand: Do teacher strikes harm educational attainment of students? In: *Labour* 24 (2010), Nr. 4, S. 391–406

- [Coh19] COHN, Kevin Gary D. and H. Gary D. and Hassett: *Tax Reform Has Delivered for Workers*. <https://www.wsj.com/articles/tax-reform-has-delivered-for-workers-11577045463>. Version: 2019
- [Cro08] CROSS, Ira: Strike Statistics. In: *Publications of the American Statistical Association* 11 (1908), Nr. 82, 168–194. <http://www.jstor.org/stable/2276104>. – ISSN 15225437
- [Eps20] EPSTEIN, Richard: *The Decline Of Unions Is Good News*. <https://www.hoover.org/research/decline-unions-good-news>. Version: 2020
- [FHK18] FARBER, Henry S. ; HERBST, Daniel ; KUZIEMKO, Ilyana ; NAIDU, Suresh: Unions and inequality over the twentieth century: New evidence from survey data / National Bureau of Economic Research. 2018. – Forschungsbericht
- [Fit19] FITZGERALD, Maggie: *Black and Hispanic unemployment is at a record low*. <https://www.cnbc.com/2019/10/04/black-and-hispanic-unemployment-is-at-a-record-low.html>. Version: 2019
- [Har14] HARRIS, Jessica: *The once and future role of strikes in ensuring U.S. worker power*. <https://onlabor.org/captive-audience-meetings-a-backgrounder/>. Version: 2014
- [Has20] HASKINS, Justin: *Trump's big reelection weapon: A remarkable manufacturing jobs boom*. <https://thehill.com/opinion/campaign/479579-trumps-big-reelection-weapon-a-remarkable-manufacturing-jobs-boom>. Version: 2020
- [HF19] HERTEL-FERNANDEZ, Alexander: *What kind of labor organizations do U.S. workers want?* <https://equitablegrowth.org/what-kind-of-labor-organizations-do-u-s-workers-want/>. Version: 2019
- [HFNR21] HERTEL-FERNANDEZ, Alexander ; NAIDU, Suresh ; REICH, Adam: Schooled by Strikes? The Effects of Large-Scale Labor Unrest on Mass Attitudes toward the Labor Movement. In: *Perspectives on Politics* 19 (2021), Nr. 1, S. 73–91. <http://dx.doi.org/10.1017/S1537592720001279>. – DOI 10.1017/S1537592720001279
- [Ken85] KENNAN, John: The duration of contract strikes in U.S. manufacturing. In: *Journal of Econometrics* 28 (1985), Nr. 1, 5–28. [http://dx.doi.org/https://doi.org/10.1016/0304-4076\(85\)90064-8](http://dx.doi.org/https://doi.org/10.1016/0304-4076(85)90064-8). – DOI [https://doi.org/10.1016/0304-4076\(85\)90064-8](https://doi.org/10.1016/0304-4076(85)90064-8). – ISSN 0304–4076
- [Ken86] KENNAN, John: The economics of strikes. In: *Handbook of labor economics* 2 (1986), S. 1091–1137
- [KT89] KEERAN, Roger ; TARPINIAN, Greg: Public Policy and the Recent Decline in Strikes. In: *Policy Studies Journal* 18 (1989), Nr. 2, S. 461
- [Mau82] MAURO, Martin J.: Strikes as a result of imperfect information. In: *ILR Review* 35 (1982), Nr. 4, S. 522–538
- [McH91] MCHUGH, Richard: Productivity effects of strikes in struck and nonstruck industries. In: *ILR Review* 44 (1991), Nr. 4, S. 722–732
- [MD10] MARTIN, Andrew W. ; DIXON, Marc: Changing to win? Threat, resistance, and the role of unions in strikes, 1984–2002. In: *American Journal of Sociology* 116 (2010), Nr. 1, S. 93–129
- [Mis18] *Examining the links between rising wage inequality and the decline of unions*. <https://equitablegrowth.org/examining-the-links-between-rising-wage-inequality-and-the-decline-of-unions/>. Version: 2018
- [Nap87] NAPLES, Michele I.: An analysis of defensive strikes. In: *Industrial Relations: A Journal of Economy and Society* 26 (1987), Nr. 1, S. 96–105
- [PMS21] POYDOCK, Margaret ; McNICHOLAS, Celine ; SHIERHOLZ, Heidi: *Downturn in strike activity reflects COVID-19 recession but fails to capture many walkouts for worker safety*. <https://www.epi.org/publication/2020-work-stoppage-report/>. Version: 2021
- [PP74] PALOMBA, Neil A. ; PALOMBA, Catherine A.: The economic costs of strikes: a suggested measure. In: *Review of Financial Economics* 10 (1974), Nr. 1, S. 12
- [Rub88] RUBIN, Beth A.: Inequality in the working class: The unanticipated consequences of union organization and strikes. In: *ILR Review* 41 (1988), Nr. 4, S. 553–566
- [SB87] SCHOR, Juliet B. ; BOWLES, Samuel: Employment Rents and the Incidence of Strikes. In: *The Review of Economics and Statistics* 69 (1987), Nr. 4, 584–592. <http://www.jstor.org/stable/1935952>. – ISSN 00346535, 15309142
- [SP20] SHIERHOLZ, Heidi ; POYDOCK, Margaret: *Continued surge in strike activity signals worker dissatisfaction with wage growth*. <https://www.epi.org/publication/continued-surge-in-strike-activity/>. Version: 2020
- [TS06] THOMPSON, Stephen L. ; SALMON, J W.: Strikes by physicians: a historical perspective toward an ethical evaluation. In: *International Journal of Health Services* 36 (2006), Nr. 2, S. 331–354
- [Wat14] WATSON, William: *If American wages are determined by skills, it's hard to see how inequality in wages is unfair*. <https://financialpost.com/opinion/william-watson-if-american-wages-are-determined-by-skills-its-hard-to-see-how-inequality-in-wages-is-unfair?r>. Version: 2014

- [WLR99] WALLACE, Michael ; LEIGHT, Kevin T. ; RAFFALOVICH, Lawrence E.: Unions, Strikes, and Labor's Share of Income: A Quarterly Analysis of the United States, 1949–1992. In: *Social Science Research* 28 (1999), Nr. 3, 265–288. <http://dx.doi.org/https://doi.org/10.1006/ssre.1999.0647>. – DOI <https://doi.org/10.1006/ssre.1999.0647>. – ISSN 0049–089X
- [Yan18] YAN, Holly: *What kind of labor organizations do U.S. workers want?* <https://www.cnn.com/2018/05/29/us/what-teachers-won-and-lost/index.html>. Version: 2018

Email address: `nyssa.cornelius@du.edu`

Email address: `christopher.jennings-shaffer@du.edu`