Dataset on Union Elections in the U.S., 1980-2020

Laurent Bouton[*], J.J. Naddeo[†‡], Ben Solow[*], Lee Tucker[§]

May 15, 2023

This document provides a description of the data sources and cleaning procedures for our new dataset covering union elections in the US for the period 1980 to 2020. It also provides a definition for each variable in the dataset.

An observation in the dataset is a full description of an NLRB election. The election information includes the election results (votes for and against unionization) of a bargaining unit within a firm including the number of eligible voters which corresponds to the size of the bargaining unit. Along with the electoral results we also know the address of the firm, date of election, and the date the results were tallied. Separately, via publicly available NLRB information, we construct the partisan leaning of the members of the National Labor Relations Board (NLRB) at the time of the election. The main novelty of our dataset in comparison to existing ones is that we extend the coverage period to 2010-2020.

1 Data Sources and Cleaning

In this Section, we discuss the sources we used to build our dataset and how we cleaned the data.

We collect our data from four different sources:

[*]Georgetown University, Department of Economics.
[†]McCourt School of Public Policy, Georgetown University
[‡]Georgetown University Law Center
[§]U.S. Census Bureau

1. Elections occurring from 1980-1999 were collected by Thomas J. Holmes;[1]

2. Elections occurring from 2000-2009 were collected by J.P. Ferguson;[2]

3. For election occurring from 2010-2020, we submitted a Freedom of Information Act for NLRB election data for that time period;[3]

4. The partisan leaning of the NLRB was determined by using the list provided on the NLRB website.[4] A row in the list represents the composition of the board from one date to another. This list was processed to label each election occurring during a Republican, Democratic, or Independent/Split board.

Although the first two sources have been utilized in previous work, this dataset is the first to extend the union election data beyond 2009. This was done through a FOIA request.[5] Adding coverage from 2009-2020 adds about 6,000 elections to our dataset. A short discussion of each cleaning decision is in order.

To begin, we removed outlier elections in which the election date is not within the time frame of each source (e.g. data from before 1980 was removed from all sources) as these are most likely the result of data entry error. Along the same lines we removed elections in which the election date is recorded as happening after the recorded tally date.[6]

There are three major types of elections within our data. First, there are the Certification of Representative Petitions (RC), which are secret ballot elections in which employees seek to unionize a workplace. Employees can also file an RC Petition to raid, or challenge the authority of another union to represent a particular bargaining unit.[7] Second, a Rep-

[1]Used in Holmes (2006) with data provided here: http://users.econ.umn.edu/~holmes/data/geo_spill/

[2]Used in Ferguson (2008) with data found here: https://github.com/jpfergongithub/nlrb-cats

[3]FOIA request NLRB-2021-000031

[4]Accessed here: https://www.nlrb.gov/about-nlrb/who-we-are/the-board/members-of-the-nlrb-since-1935

[5]In fact, the first FOIA request we submitted revealed internal errors in the underlying NLRB database. After some correspondence, we were able to rectify the issues.

[6]The tally date is the date in which ballots are counted.

[7]These elections will have multiple unions on the ballot. We exclude these elections because the electoral rule is different than in single union elections (i.e., they may require a second round). Because of the different

resentation Petition (RM) happens when an employer wants to demonstrate to the NLRB that the union has lost the support of a majority of the employees. A "successful" RM election ends with decertification of the current union. Finally, a Decertification Petition (RD) is filed by the employees and must have signatures of at least 30% of the bargaining unit expressing an interest to get rid of their union. Of course, employer cannot be involved meaning instigating, encouraging, soliciting, distributing, etc. in the employees decision to obtain, complete, and file an RD Petition.

The latter two election types (RM and RD) happen in workplaces in which there is already an existing union. There are various reasons why the effect of decertification of an existing union is not symmetric to the effect of unionization, which thus justify dropping RM and RD elections from our dataset. For instance, any downward rigidity of wages, which is prevalent in most labor markets (Fehr and Goette, 2005; Holden and Wulfsberg, 2008), would prevent the anticipated effect of decertification on wages. Also, even without wage rigidity, the wage-employment trade-off for incumbent employees is different in unionization and decertification because incumbent employees suffer when employment level in the firm decreases, but do not necessarily gain when it increases. Finally, we removed elections with missing electoral data, duplicates[8], and, following the existing literature, elections with fewer than 20 participants.

Tables 1-3 show the impact each filter has on the number of total rows within the data, as well as the unique number of casenumber-bargaining unit-election date triple. After cleaning each source of data we then simply append the datasets together. The result is harmonized coverage of NLRB RC elections from 1980-2020 representing 49,878 unique elections.

electoral rule, it is not clear that there exists a single threshold vote share for unionization. RDD estimation would then be difficult to justify.

[8]Duplicates are mainly caused by multiple address entries, and or re-runs that also duplicate the original election information (possibly an artifact of how the NLRB database is merged).

Cleaning of Source #1

Filter	Number of obs	Number of unique elections
Raw	105,173	104,062
Remove pre-1980 elections (and those missing election dates)	83,804	82,904
Remove elections marked closed before election	83,365	82,469
Remove non-RC cases	68,094	67,280
Drop small elections (<20 total votes)	36,865	36,544
Drop missing voting data (for or against)	36,865	36,544
Drop multi-union elections	36,287	35,970
Drop duplicates due to address info	35,970	35,970

Table 1: Impact of cleaning process on Holmes data

Cleaning of Source #2

Filter	Number of obs	Number of unique elections
Raw	345,036	28,977
Remove pre-2000 elections (and those missing election dates)	327,464	27,460
Remove elections tallied before election	326,271	27,345
Remove duplicate re-runs	303,838	26,002
Remove non-RC cases	226,480	20,720
Drop small elections (<20 total votes)	124,180	10,831
Drop missing voting data (for or against)	123,501	10,806
Drop multi-union elections	98,272	10,020
Drop duplicates due to address info	10,020	10,020

Table 2: Impact of cleaning process on JP Ferguson data

Cleaning of Source #3

Filter	Number of obs	Number of unique elections
Raw	18,391	16,250
Remove pre-2009 elections (and those missing election dates)	18,351	16,217
Remove elections tallied before election	18,239	16,122
Remove duplicate re-runs	17,602	16,040
Remove non-RC cases	14,641	13,365
Drop small elections (<20 total votes)	7,236	6,589
Drop missing voting data (for or against)	7,234	6,588
Drop multi-union elections	6,567	6,088
Drop duplicates due to address info	6,088	6,088

Table 3: Impact of cleaning process on FOIA data

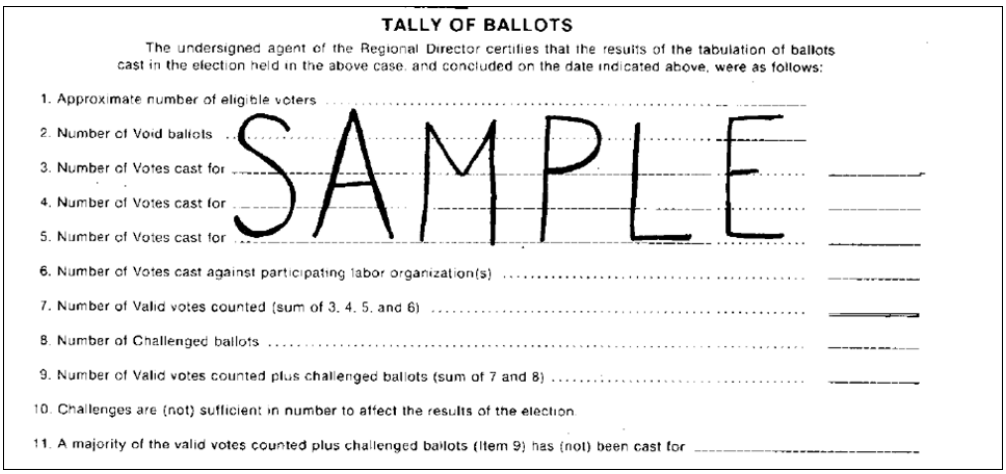
2 Data Accuracy

In each dataset there are variables that measure the number of votes for and against unionization as well as the number of votes that were challenged and the number of valid votes counted. The election procedure outlined by the NLRB[9] details how votes can be challenged/determined valid. To ensure that our measurement of votes cast for and against unionization are correct we construct a variable total votes and compare it to the number of valid votes. As each data source may be the result of various different data collections processes we perform this data check on each source separately.

For the data coming from Holmes (2006) (1980-1999) we find that 100% of the time the number of votes for plus votes against equals the number of valid votes counted. In the next segment of data coming from Ferguson (2008) (2000-2009) we find that 99% of elections pass this test. Finally, in the data acquired using a FOIA request (2010-2020) we find that only

[9]https://www.nlrb.gov/sites/default/files/attachments/pages/node-195/form-nlrb-5547-election-and-post-election-procedures.pdf accessed 8/31/21

86% of the elections pass the test. We satisfy this test for 97% of elections if we allow the number of valid votes to equal either the total number of votes cast for/against or the total number of votes cast for/against plus the number of challenged votes. This indicates that sometimes the challenged votes may not be recorded in the number of valid votes. We are discussing with our FOIA officer how to interpret these findings. She has provided us with a sample tally of ballots that is used to populate the columns in our current dataframe. As one can see, 86% of the time the number of valid votes variable measures field 7 while 10% of the time it measures field 9. According to our FOIA officer, the best explanation is simply data entry error that occurs when one transcribes the handwritten numbers from Figure 1 into the NLRB's database.


TALLY OF BALLOTS

The undersigned agent of the Regional Director certifies that the results of the tabulation of ballots cast in the election held in the above case, and concluded on the date indicated above, were as follows:

1. Approximate number of eligible voters
2. Number of Void ballots
3. Number of Votes cast for
4. Number of Votes cast for
5. Number of Votes cast for
6. Number of Votes cast against participating labor organization(s)
7. Number of Valid votes counted (sum of 3, 4, 5, and 6)
8. Number of Challenged ballots
9. Number of Valid votes counted plus challenged ballots (sum of 7 and 8)
10. Challenges are (not) sufficient in number to affect the results of the election.
11. A majority of the valid votes counted plus challenged ballots (Item 9) has (not) been cast for

SAMPLE

Figure 1: Sample tally ballot provided by FOIA officer. The values are handwritten by officer in charge of tallying the election results and then subsequently imputed into the NLRB's database, which produces the data we have obtained.

Another important metric that we used to test the accuracy of our data is an indicator of whether the unionization attempt was successful. All of our data sources provide variables that indicate a union "win" or "loss". We then create a variable that is 1 when total votes for the union is greater than total votes against and 0 otherwise. In the datasets that span 1980-1999, 2000-2009, and 2010-2020 there is agreement between both indicators in 99.9%, 98.6%, and 98.1% of the elections we record, respectively. This gives us confidence in our data cleaning procedure.

3 Definitions of the Variables

Table 4 provides the main variables harmonized across all data sources for each election.

Table 4: Important Variables and Descriptions

Variable Name	Description
casenumber	case number unique to election
votesforlaborорg	number of votes for unionization
votescastagainstlaborорg	number of votes against unionization
source	data source
employer_state	state of employer
employer_zip	zipcode of employer
month	month of election
day	day of election
year	year of election
vote_diff	Difference between votes for and against unionization
total_voters	total number of eligible voters
vote_share	share of votes for unionization
partyincontroloftheboard	political party that appointed majority of NLRB board at the time of election
repseats	number of NLRB seats held by republicans
demseats	number of NLRB seats held by democrats
emptyseats	number of empty seats on NLRB board at the time of elections
indseats	number of NLRB seats held by independents
repcontrol	republican controlled board at time of election
demcontrol	democratic controlled board at time of election
indcontrol	independent controlled board at time of election
challenged_votes	number of votes challenged in the election
employer_name	name of employer

References

Ferguson, John Paul, "The eyes of the needles: A sequential model of union organizing drives, 1999-2004," *Industrial and Labor Relations Review*, 2008, *62* (1), 3–21.

Holmes, Thomas, "Geographic Spillover of Unionism," *National Bureau of Economic Research*, 2006.