Explorations in Economic History 90 (2023) 101531

Contents lists available at ScienceDirect

Explorations in Economic History

journal homepage: www.elsevier.com/locate/eeh

Research Paper

The economic impact of social distancing: Evidence from state-collected data during the 1918 influenza pandemic☆

Benjamin Bridgman[a], Ryan Greenaway-McGrevy[b,*]

[a] Bureau of Economic Analysis, Washington, DC, 20233, United States
[b] The University of Auckland, New Zealand

ARTICLE INFO

JEL classification:
I18
N12
O47

Keywords:
Pandemics
1918 influenza
Non-pharmaceutical
Interventions
Employment

ABSTRACT

What are the long-run economic impacts of the policy responses to control pandemics? We investigate this question by exploiting state-collected data spanning one of the most consequential global pandemics in centuries, the 1918 influenza pandemic. Specifically, we use a difference-in-differences framework to examine the effects of non-pharmaceutical interventions (NPIs), ultimately finding no long-run impact of NPIs on employment, positive or negative. Employment trends prior to 1918 suggest that World War One is an important confounding factor in analyses of the pandemic, since cities with tighter NPIs grew rapidly between 1914 and 1918. We identify new control variables that account for war production and find that social distancing did not have long run employment impacts. The evidence underscores the importance of accounting for confounding economic and policy factors for understanding the impact of pandemics on economic outcomes.

1. Introduction

COVID-19 has increased interest in the economic impact of pandemics. The 1918 influenza pandemic was a major global event with no peer in the intervening years, so it has become a laboratory to understand the impact of such events (Arthi and Parman, 2021). A feature of the 1918 flu in the United States that makes it attractive to study these questions is there were significant differences in the policy response across cities. Some cities had lengthy closures (called Non-Pharmaceutical Interventions, or NPIs) while others did not. Studies have used this variation to examine the impact of NPIs on mortality (Barro et al., 2020; Markel et al., 2007), innovation (Berkes et al., 2023), and educational outcomes (Ager et al., 2022).

Is there a longer-run trade-off between NPIs and the economy? Many potential effects, positive and negative, have been put forth in the context of COVID-19. NPIs could have a negative impact on the economy by weakening firms through lost sales (Elenev et al., 2022) and disrupting schooling (Fuchs-Schundeln et al., 2022; Psacharopoulos et al., 2021). On the other hand, NPIs could have a long-run positive effect relative to the counterfactual of widespread infection. Influential work has argued that exposure to the disease can have long-lasting impacts on labor market outcomes and educational outcomes (Almond, 2006). NPIs could improve outcomes by protecting firm managers and workers from succumbing to the disease (Atolia et al., 2021) or increase long-run sales by reducing

☆ We thank Stephan Maurer and Ferdinand Rauch for sharing their market access indicator and Kim Ruhl for aid with collecting data. Adam Copeland, Abe Dunn, Jeremy Moulton, Scott Wentland, and the audience at the Econometric Society NASM (Montreal) provided valuable comments. Bridgman is Assistant Chief Economist at the Bureau of Economic Analysis, Washington D.C. Greenaway-McGrevy is Associate Professor at the Univeristy of Auckland. The views expressed herein are those of the authors and do not necessarily reflect the views of the Bureau of Economic Analysis or the Department of Commerce.

* Corresponding author.
E-mail addresses: Benjamin.Bridgman@bea.gov (B. Bridgman), r.mcgrevy@auckland.ac.nz (R. Greenaway-McGrevy).

https://doi.org/10.1016/j.eeh.2023.101531
Received 15 August 2022; Received in revised form 17 May 2023; Accepted 23 May 2023
Available online 25 May 2023

infection in public spaces (Goolsbee and Syverson, 2021). Any persistent effects would amplify (or mitigate) the contemporaneous effects, potentially changing the calculus on the use of NPIs.

We use a difference-in-differences framework to examine the impact of NPIs during the 1918 pandemic on economic outcomes. A major contribution is introducing newly digitized annual (or near annual) employment data collected by U.S. states for 26 cities. These relatively high-frequency data allow us to observe potential pre-trends in each year between 1914 and the pandemic in 1918.[1] They also allow us to identify and account for short-lived and fast-moving factors that may otherwise confound the quasi-experimental analysis, such as those related to World War One (WWI).

We find that there is no long run impact of NPIs on employment, positive or negative. In our baseline empirical specification, we find a marginally positive treatment effect for our measure of NPIs duration in a single year – 1920 – that is significant at a ten percent level. However, the positive treatment effect is sensitive to samples and specification. We do not find an effect for NPIs response speed. (We define these NPIs measures precisely below.) Overall, we conclude that the direct economic impact was small and short-lived, if it existed at all, highlighting the importance for researchers to account for confounding economic and geopolitical effects when analyzing the economic impact of pandemics more generally.

Our work also shows that WWI is an important confounding variable that should be accounted for in 1918 pandemic policy evaluation exercises. Cities with tighter NPIs grew faster between 1914 and 1918, which impairs a causal interpretation of observed differences in outcomes after the policy intervention in 1918. Moreover, these pre-trends persist after conditioning on the conventional set of controls employed by the extant literature, which account for differences in population growth, production composition and historic flu susceptibility between cities. The pre-trends coincide with a major expansion of manufacturing exports to European allies prior to U.S. entry into WWI. We add controls to account for both foreign and domestic manufacturing demand caused by the war, finding that these controls substantially attenuate the pre-trends to the point where they are statistically insignificant, facilitating a causal interpretation of the estimated treatment effects. In other words, including WWI covariates results in the fitted DID model satisfying the (conditional) parallel trends assumption.

Our paper contributes to the literature on economic impact of infectious disease on the economy. Despite the size of event, there was very little analysis of the 1918 flu's economic impact in the United States prior to the COVID-19 pandemic. A handful of papers examine the direct impact of the flu on the economy: Brainerd and Siegler (2003); James and Sargent (2006), and Garrett (2009). Our paper expands on this literature by bringing in more frequent U.S. economic data. We also examine the impact of NPIs, which these papers do not. Almond (2006) and Beach et al. (2022a) examine the effects of the flu on future generations, a much longer time frame than our work. There are a few papers looking at other countries such as Bloom and Mahal (1997) for India and Karlsson et al. (2014) for Sweden. COVID–19 has led to a renewed interest in this event. Beach et al (2020) provide a survey of the literature that has grown since COVID–19. In addition to the those cited above, papers have examined the economic impact in other countries, such as Denmark Dahl et al. (2020), Japan (Noy et al., 2020) and Spain (Basco et al., 2020). Fluegge (2022) examines city population growth. Anderson et al. (2020) examine banking liquidity provision in New York State counties. Our analysis focuses on U.S. employment.

The closest work to this paper are Correia et al. (2022) and Lilley et al. (2020) (henceforth CLV and LLR respectively), which uses a similar empirical approach on Federal Census of Manufactures (FCM) data. CLV find evidence of a positive effect from NPIs while LLR argue that this result is due to longer run pre-trends. The FCM data have sparse coverage and do not cover 1918, the peak year of the flu.[2] Like CLV, we find that cities with tighter NPIs grew between the FCM census years 1914 and 1919. However, all of this growth occurs *prior to the onset of the pandemic* in 1918. Observed differences in outcomes in the post-pandemic period therefore do not reflect the causal impact of variation in the NPIs treatment, but differences in growth rates over the period spanning WWI. We find that the positive employment effects attributed to NPIs by CLV are substantially attenuated and become statistically insignificant once we account for the early war export boom. These results help reconcile the Velde (2022) finding of little impact of NPIs on economic activity with the CLV marginally positive long-run effects.

The rest of the paper proceeds as follows. We begin by describing the state-collected data. We then present our empirical strategy and results, and compare our finding to those of the extant literature. Section 4 concludes.

2. Data description

This section describes our dataset. We begin by detailing the state statistical collections that form the basis of the data. We then summarize our sample coverage of cities.

2.1. State data collections

The Federal statistical program was much more limited during the 1910s (Rockoff, 2020). There were few annual data series and no annual measures of output. To fill in these gaps, many states conducted their own statistical programs that were collected much more frequently. The data fall into two categories: data derived from surveys and those from administrative records. Most of our data

[1] Following a convention in many difference-in-differences papers, we refer to a violation of the parallel trends assumption as "pre–trends." For example, see Rambachan and Roth (2022).

[2] CLV (2022) do have more frequent data for another quantitative indicator of economic activity (bank deposits) but their main results rely on the FCM. Velde (2022) and Bodenhorn (2020) use high frequency data to examine the contemporaneous impact of the flu but their indicators are limited in scope.

Table 1
Data coverage.

State	Cities	State-collected data	Federal data
CA	Los Angeles, San Francisco	1916–18, 21	1914, 19, 23, 25
CT	New Haven	1916, 18, 20–22	1914, 19, 23, 25
IA	Des Moines	1915, 17, 19, 21	1914, 23, 25
MA	Boston, Cambridge, Fall River, Lowell, Worcester	1910–25	
MI	Detroit, Grand Rapids	1910–19	1921, 23, 25
MO	Kansas City, St. Louis	1910–12, 14–21, 23–25	
NJ	Newark	1914–16	1919, 21, 23, 25
NY	New York City	1915–25	1914
OH	Cincinnati, Cleveland, Columbus, Dayton, Toledo	1910–12, 14–21, 23–25	
PA	Philadelphia, Pittsburgh	1914–25	
RI	Providence	1910–11, 1913–21	1923, 25
SC	Charleston	1910–1920, 23-5	
WA	Seattle, Spokane	1910, 12, 14–15, 18	1919, 21, 23, 25

come from economic surveys. We augment these traditional surveys with those collected from administrative records. These data were collected from the files of state programs such as workmen's compensation insurance or as a result of regulation enforcement on workplace safety or restrictions on child labor. These data have not been used much.[3] Velde (2022) uses data from the Ohio, New York, and Massachusetts sources we use to study the 1918 pandemic. He does not use them to examine the impact of NPIs.

While these data have not been used extensively, we believe they are of remarkably high quality for this era. We only include data that are derived from mandatory Government collections that covered a substantial sample of manufacturing establishments. We exclude collections that were not mandatory and had poor response rates (e.g. Nebraska) and those that collected data from a small portion of establishments (e.g. Kentucky). The state collections had some advantages over Federal collections at that time. State surveys were conducted frequently (generally annually) so respondents were used to filling out surveys and enumerators were practiced in conducting collections. In contrast, the FCM was collected every 5 years in the 1910s and had inconsistent coverage from survey to survey (Rhode, 2001).

Most of our data come from economic surveys. All of our data for 12 cities come from such surveys and Ohio's five cities use a combination of survey and administrative sources. Massachusetts and Missouri, states that make up a third of our sample of cities, had annual surveys of manufactures that extended back to the 1890s. Massachusetts worked closely with the U.S. Census Bureau and the enumerators were the same as those that collected the FCM. These surveys are quite comparable to the FCM. The state surveys generally followed Federal surveys both in industry coverage and data items collected. Deviations from the Federal template usually involved collecting more data items with greater detail than the FCM. For example, most states broke out employment by gender when the FCM did not.

The remaining 5 cities from three states draw exclusively on administrative data. These 5 cities' data use data collected by factory inspectors, where enumerators physically visited establishments. These three states (Michigan, Rhode Island, and Washington State) report employment levels comparable to the FCM. Rhode Island's factory inspections had a long history, beginning in 1894. Ohio's administrative data come from its mandatory workmen's compensation insurance program. Its coverage was extraordinarily complete, covering all employers with 5 or more employees. These data were of such high quality that the U.S. Bureau of Labor Statistics published these data to aid in studying the Great Depression (for example, see Croxton and Croxton, 1932).

2.2. Sample

Our core dataset cover 26 cities in 11 states reported in Table 1. These states are geographically diverse and contain the major industrial centers at the time, including half of the top 20 cities in value of manufacturing product in 1919 (New York, Newark, Philadelphia, Detroit, Cleveland, St. Louis, Pittsburgh, Boston, Cincinnati, Kansas City, and San Francisco). These cities also have a wide variety of disease experiences, with the city with the highest and lowest mortality in Markel et al. (2007) (Pittsburgh and Grand Rapids respectively). In the Appendix, we compare our sample with the FCM data used by CLV and LLR and find they display similar patterns.

The table reports the years that employment data are available for each state's data collection between 1910 and 1925. We augment the dataset with information from the FCM in order to fill out observations through to 1925 for cities in Connecticut, New Jersey, Iowa, Rhode Island, Michigan, California, and Washington State. The table also reports what FCM observations we use. Specifically, we use growth rates in manufacturing employment from the census to extrapolate employment numbers for 1914 (Des Moines, New Haven, San Francisco, Los Angeles, New York), 1919 (Newark, San Francisco, Los Angeles, Seattle, Spokane), 1921 (Newark, San Francisco, Los Angeles, Seattle, Spokane, Grand Rapids and Detroit), 1923 and 1925 (Des Moines, Newark, San Francisco, Los Angeles, Seattle, Spokane, Grand Rapids, Detroit and Providence). We have employment figures at annual frequencies over contiguous time spans for most of our cities. However, when data are missing for a given year, we linearly interpolate to fill in the missing observations. This results in 48 interpolations out of a total of 286 data points in our baseline sample. Empirical results based on the non-interpolated sample are presented as a robustness exercise in the Appendix.

[3] An example of previous use of state-collected data is Sundstrom (1990), which examines wage rigidity with data from Ohio.

Table 2
City sample statistics.

City	State	NPI days	Response time
Los Angeles	CA	154	5
San Francisco	CA	67	11
New Haven	CT	39	−22
Des Moines	IA	56	10
Boston	MA	50	13
Cambridge	MA	49	14
Fall River	MA	60	10
Lowell	MA	59	11
Worcester	MA	44	15
Detroit	MI	47	17
Grand Rapids	MI	62	17
Kansas City	MO	170	0
St. Louis	MO	143	1
Newark	NJ	33	−10
New York	NY	73	−11
Cincinnati	OH	123	2
Cleveland	OH	99	−2
Columbus	OH	147	5
Dayton	OH	156	−5
Toledo	OH	102	2
Philadelphia	PA	51	8
Pittsburgh	PA	53	7
Providence	RI	42	19
Charleston	SC	69	8
Seattle	WA	168	5
Spokane	WA	164	1

Our measures of NPIs are based on the work of Markel et al. (2007)). We use two of their measures. “NPI days” is the total number of days the city mandated three NPIs: School closures, ban on large gatherings, and isolation of ill people and their contacts. This is our measure of NPI duration. “NPI speed” is the number of days after flu infections accelerated that one or more of the three NPI measures was imposed. This is our measure of how quickly city officials reacted to the flu. Markel et al. (2007) measure the acceleration date as when flu mortality was twice the 1910–16 baseline. Negative numbers indicate that NPIs were imposed before that date. However, we multiply by −1 so that larger values correspond to faster responses.

The original Markel et al. (2007) data only cover 23 of our 26 cities so we fill out the sample with information from the Influenza Archive.[4] Charleston, Des Moines, and Detroit have the required documentation to calculate NPI Days. We use the calculation reported in Berkes et al. (2020). These cities do not have the required flu reports to calculate NPI speed. We use Grand Rapids' speed of response for Detroit. Both cities imposed initial NPI at the same time as result of a state mandate. Primary sources indicate that Detroit's leadership argued against the state NPI mandate, so using Grand Rapids's slow response (17 days) is likely accurate. For Charleston and Des Moines, we predict NPI speed based on (log) NPI days using a simple linear regression fitted to the full Markel sample of 43 cities and Detroit. The sample correlation between the two variables is a reasonable 0.64, with cities that had longer NPI durations also having faster NPI implementation, and so we are confident that this prediction yields a reasonable approximation of NPI speed for these two cities.[5] Table 2 reports the NPI variables for our sample.

All data are available online (Bridgman and Greenaway-McGrevy, 2023).

3. Empirical strategy

This section assesses the economic impact of NPIs. The public health policy response varied significantly across cities. Our geographically disaggregated data allow us to examine the heterogeneous response to the flu on economic outcomes. To examine the impact of responses to the flu we adopt a multi-period difference-in-differences (DID) approach similar to that used by CLV and LLR. The multiperiod DID specification permits us to examine whether the parallel trends assumption holds, and to examine the effects of fast-moving confounding factors such as those related to WWI.

An important question when setting up the event study is to determine the counterfactual trend. We use 1914 to 1924 as our baseline sample period.

The beginning of our baseline sample is dictated by data constraints. Between 1910 and 1913 we have employment observations for two thirds of our sample. We therefore start our sample in 1914 but consider beginning in 1910 as a robustness check.

While our selection of 1914 is largely practical, we think this is a reasonable date. It is not too late, allowing multiple periods (five years) with which to identify pre-trends prior to the pandemic. CLV emphasize that the 1914 to 1919 trends are the most relevant to the analysis. Neither is it too early. Extrapolating too far back may pick up trends that are not germane to our analysis. This is

[4] https://www.influenzaarchive.org/index.html.

[5] Our empirical findings and conclusions are unchanged when we restrict the sample to 24 cities with NPI speed observations.

a particular concern for the turn of 20th Century when new inventions and the expansion of electricity make early 20th Century manufacturing much different than the late 19th Century. In 1900, many significant products (e.g. automobiles and radios) were in their pre-commercial infancy while others (e.g. airplanes) were not even invented. By 1914, these products had been commercialized.

We end our analysis in 1924 because we have reason to believe that trends would change after this year when immigration was strictly limited. While immigration quotas began in 1921, Moser and San (2020) date the quantitatively important quotas as after 1924, a result of the 1924 Johnson–Reed Act. Abramitzky et al. (2019) argue that this restriction had differential effects on cities related to the mix of countries sending previous immigrants to each city.

3.1. Empirical model

Let $Y_{i,t}$ be employment in city i in year t, let Q_i be a treatment variable measuring NPI stringency (NPI days or NPI speed), and let X_i be a vector of controls. Furthermore, let $t = -\underline{T}, -\underline{T}+1, \ldots, 0, \ldots, \overline{T}-1, \overline{T}$, and let the treatment occur in period $t = 0$. Thus we have $\underline{T}$ pre-treatment observations and $\overline{T}$ post-treatment observations. The model is then

$$\ln\left(Y_{i,t}\right) = \alpha_i + \tau_t + \sum_{s=-\underline{T}, s\neq 0}^{\overline{T}} \beta_s Q_i \mathbf{1}_{s=t} + \sum_{s=-\underline{T}, s\neq 0}^{\overline{T}} \gamma_s' X_i \mathbf{1}_{s=t} + \varepsilon_{i,t} \tag{1}$$

where $\mathbf{1}_{s=t}$ is a dummy variable indicator for each year before and after the treatment year, which we set to 1918. Q_i is standardized to permit interpretation of the coefficients in terms of standard deviations in the treatment. Note that controls are interacted with the time period dummies for each year before and after the intervention. To account for potential heteroskedasticity and serial dependence in the error term, we cluster the standard errors by city.[6]

Controls are important given that the treatments of interest are non-random.[7] Much of the debate has focused on the impact of fast-developing Western cities catching the flu later and implementing more stringent interventions after observing outcomes in Eastern cities (LLR, Correia et al., 2020, Barro et al., 2020). Population (or population growth) prior to the onset of the pandemic has therefore been identified as an important control that attenuates pre-trends in employment outcomes in FCM datasets (LLR). We therefore include population growth in the set of controls.

However we must also be particularly judicious in our selection of controls given the limited cross sectional dimension of our panel. We consider the set of controls employed by CLV, LLR and Correia et al. (2020) for which we have corresponding data for our sample of 26 cities, but only include variables that are found to be significantly correlated with NPIs in our baseline regression models. In subsequent robustness checks we consider additional controls that have been identified as important in the extant literature.

Our set of all potential controls includes the following city-level measurements: (log) population in 1900, (log) population in 1910, (log) population in 1914, and population density. Our state level controls include share of manufacturing employment, proportion of population in cities larger than 2500 persons (hereafter “urbanization rate”) and flu-deaths in 1917. We cannot include manufacturing share of city population (ratio of FCM manufacturing employment in 1914 to census population in 1910), as these data are lacking for Charleston. Regardless, this variable only explains 12% of the variation in NPI days when Charleston was omitted from the sample, and is statistically insignificant at the 5% level.

We reduce this larger set of potential controls to three. The controls are: manufacturing share of population, 1917 influenza deaths per capita, and population growth between 1900 and 1914. We selected these controls based on the following process.

NPI days is highly correlated with manufacturing share of employment ($p = 0.01$), 1917 flu deaths ($p = 0.01$) and urbanization ($p = 0.05$) in individual level regressions. By itself, manufacturing share of employment explains 26% of the variation in NPI days, while flu deaths explains 24%. Including urbanization in the regression model offers only a 1 percentage point increase in R-squared when manufacturing share and flu deaths are already in the regression model. We therefore omit urbanization from the controls. Population growth by itself only explains 12% of the variation in NPI days ($p = 0.08$), and does not offer much increase in explanatory power beyond that offered by state manufacturing share and 1917 flu deaths. However, as discussed above, LLR stress the need to condition on long run trends in population growth given that many of the cities with tighter NPIs are fast-growing cities. We therefore include this variable as a control.

Following the same approach to select controls for the NPI speed treatment yields manufacturing share of employment ($p = 0.04$) as the single variable with substantial explanatory power – explaining 16% of the variation in NPI speed. We use the same set of controls selected for NPI days as it includes manufacturing share of employment.

3.1.1. NPI days

Figure 1 illustrates results when NPI days is used as the treatment in the multiperiod DID model. We begin with the model without controls (left hand side). (We will introduce the controls used in the other panels below.) There is a prominent and continuous upward trend between 1914 and 1918. This pre-trend indicates that tighter NPI cities (i.e. cities with longer NPI interventions) were growing

[6] We also implemented a panel version of the Newey–West standard errors. Our findings remain unchanged and are thus not reported.

[7] Even when the research design is experimental or relies on a quasi-experimental framework, Angrist and Pishcke (2009) argue that controls can be important for precision of the model. Specifically, they state: “Inclusion of the variables Xi, although not necessary in this case, may generate more precise estimates of the causal effect of interest... [and] including these control variables therefore reduces the residual variance, which in turn lowers the standard error of regression estimates” p.23–24.

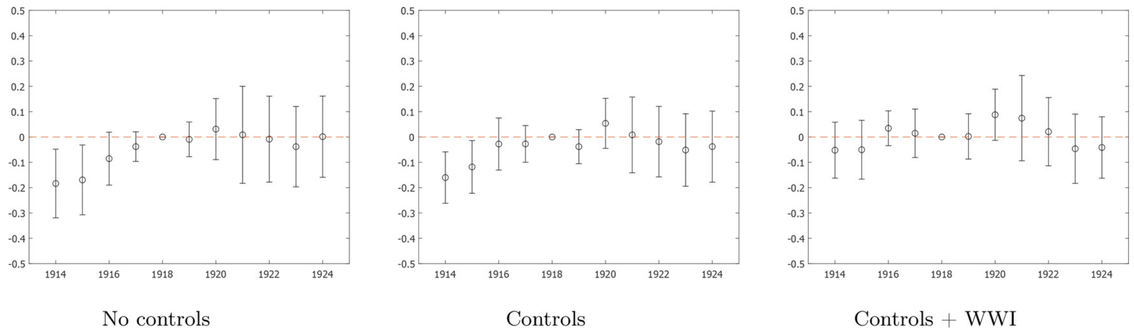


Fig. 1. NPI days treatment.
Note: Multiperiod difference-in-differences with NPI days as treatment. Point estimates (circles) and 95% confidence intervals (error bars). Outcome (y-axis) is change in log employment. “Controls” refers to the model that includes a set of controls taken form the extant literature, including population growth, production composition and historic flu susceptibility. “Controls + WWI” refers to this set plus the three WWI controls we identify below.

faster immediately before the pandemic. This violation of the PT assumption invalidates the treatment and control comparisons in the post treatment period.

Point estimates are insignificant in the post-treatment period (1919 to 1924). The confidence intervals are relatively large, ranging from between 0.12 and 0.4 in magnitude. For example, in the specification without controls, they span (−0.08, 0.06) in 1919 (the first post-treatment period) and (−0.16, 0.16) in 1924 (the final post-treatment period). This implies that the coefficients are imprecisely estimated.

Next we add in the limited set of controls taken from the extant literature: Manufacturing employment share, population growth, and 1917 flu deaths per capita (middle panel). There are two important differences once controls are included. First, prior to treatment, the coefficients are smaller in magnitude, indicating that the controls remove some of the observed pre-trend. Without controls the estimated coefficient for 1914 is −0.184. Once the controls are included the estimated coefficient shrinks to −0.160. Nonetheless, the upwards pre-trend between 1914 and 1918 remains. Second, the post-treatment coefficient for 1920 is slightly larger in magnitude, increasing from 0.031 to 0.054, though it remains statistically indistinguishable from zero (t-stat = 1.076).

The observed pre-trends in the state-collected data have implications for the interpretation of findings that are based on FCM-Markel data. Cities with tighter NPI grew between the FCM census years 1914 and 1919. However, in the state-collected dataset, all of this growth occurs prior to the onset of the pandemic in 1918 – regardless of whether the conventional controls are included or not. Observed differences in outcomes in FCM data in the post-pandemic period therefore likely do not reflect the causal impact of variation in the NPI treatment, but differences in growth rates over the period immediately prior to the pandemic and spanning WWI. We conclude that FCM data may be too sparse to distinguish the causal effects of NPI interventions from pre-trends unless suitable controls are identified that can mitigate selection into treatment in high NPI cities. We return to this issue below.

3.1.2. NPI speed

Figure 2 illustrates the results when NPI speed is used as the policy intervention. The patterns mirror those for the NPI days treatment above. There is an upwards pre-trend in the specification without controls – although in this case the coefficients are

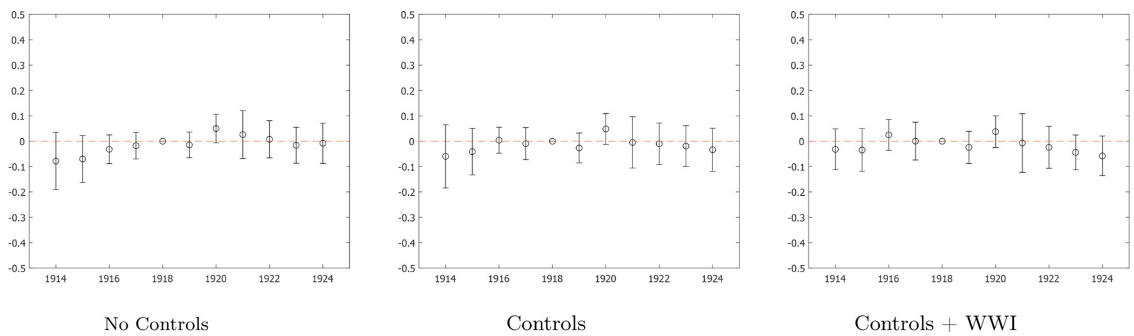


Fig. 2. NPI speed treatment.
Notes: Multiperiod difference-in-differences with NPI speed as treatment. Point estimates (circles) and 95% confidence intervals (error bars). Outcome (y-axis) is change in log employment. “Controls” refers to the model that includes a set of controls taken form the extant literature, including population growth, production composition and historic flu susceptibility. “Controls + WWI” refers to this set plus the three WWI controls we identify below.

not statistically distinguishable from zero. Notably the treatment effect in 1920 is positive and near significant at the 5% level in 1920 (t-stat = 1.742). Compared to the NPI days treatment above, both the coefficients and the confidence intervals are smaller in magnitude.

The inclusion of the three controls (manufacturing employment share, population growth, and 1917 flu deaths per capita) further attenuates the pre-trend in the point estimates. The treatment effect in 1920 is still marginally significant (t-stat = 1.571).

3.2. WWI controls

As demonstrated in the previous subsection, the set of controls used by LLR and CLV does not remove pre-trends over the short time horizon of our sample period when NPI days is used for treatment intensity.(See the Appendix for results when the full set of controls are included.) This suggests that there remain confounding effects that need to be accounted for.

A concern with examining the 1918 flu is that it coincides with World War One, so any effects attributed to the flu or policy responses to it may be due to the war. WWI has been an issue in other work examining the 1918 flu. Beach et al. (2022b) has found WWI to be confounding factor for the long-run impact of the flu on individual achievement found in Almond (2006).

In this section we motivate and identify controls that account for the confounding impact of the war. We find evidence that the war is an important confounding factor. The addition of these variables to the set of controls substantially attenuates the pre-trend in the NPI days treatment exercise. Coefficients in the pre-treatment period are statistically indistinguishable from zero and have tighter confidence intervals. We also find that estimated treatment effects immediately after the flu are positive and have marginal statistical significance in the NPI days treatment exercise. However, in robustness exercises this result is highly sensitive to changes in model specification. Estimated coefficients for NPI speed are not significant.

There are two aspects of the war that we will analyze. It is well known that U.S. involvement directly increased employment since the government purchased an enormous quantity of manufactured goods to prosecute the war. Less well known is a surge in exports during the period prior to U.S. entry as WWI combatants no longer traded with each other and their production shifted to war goods.

We begin the analysis by describing the impact of the war on U.S. manufacturing. This history will guide our choices of control variables.

The United States did not the enter the war until April 1917. However, the war led to an export boom that benefited manufacturing in particular. U.S. exports doubled as a share of GNP from 1914 to 1916 to its highest level ever recorded (Fordham, 2007). This increase was due to exports of manufactured goods. Real finished manufactured exports increased 152 percent between 1913 and 1916. The corresponding increase for crude materials and semi-manufactures was 12 percent (Lipsey, 1963). This period marked when U.S. exports shifted from being dominated by agricultural goods to manufactures. Though the U.S. was neutral, the vast majority of these exports went to the Allies in Europe[8]

This export demand led to a massive boom before U.S. entry. While all manufacturing benefited, cities outside the traditional manufacturing core in the Northeast grew faster. The massive demand and the loss of domestic civilian capacity among the allies meant that low trade-cost Northeastern cities could not fulfill that demand alone. The demand was strong enough that European buyers would accept the higher trade costs of more distant cities. For example, West Coast shipyards were at capacity and expanding to supply distant Norwegian contracts.[9] This force also increased Western employment to meet domestic consumption since Northeastern goods that would have supplied U.S. markets (absent the war) were exported.[10]

Once the U.S. joined the war, the military contracted for war materiel. Builders of navy ships and army aircraft were particularly important employers (Garrett, 2009). Military contracts were not the only important impact on manufacturers. To supply troops an ocean away, the military needed a fleet of civilian cargo ships. The Emergency Fleet Corporation (EFC) was created to contract for merchant ships. The EFC expanded shipbuilding in areas that had not been major production areas pre-war: It paid to create new shipyards and expand those that had made small boats previously, not just buying from established yards. (Established yards were already at capacity due to the export boom.) Seattle was particularly affected, with manufacturing employment tripling, with all of the gain due to shipbuilding.

To account for these effects, we include three control variables.

To account for the export boom prior to U.S. entry, we use a measure of market access to European markets developed by Maurer and Rauch (2022). As discussed above, the growth of manufacturing was strongest in areas furthest from Europe. To examine the exposure of each city to European export demand, we use the effective distance of each city to Liverpool (the main British port) developed by Maurer and Rauch (2022). This measure estimates the shortest route from the center of each county to Liverpool, taking into account the availability of waterways and railroads as measured in Donaldson and Hornbeck (2016). Effective distance accounts for the relative cost of each mode, so that overland routes are "further" than water routes. (See Maurer and Rauch, 2022 for details.) We match each city to its county and use the county's effective distance for the city.

To account for government purchases of ships during U.S. involvement in the war, we calculate a city level version of the Garrett (2009) indicator of war production. (His measure is state level). The indicator variable takes a value of one if the city was a

[8] The initial impact of the war was harmful for exports and the economy. The global financial system ground to a halt with the onset of WWI; the New York Stock Exchange was closed for 5 months (Sprague, 1915). This recession was severe enough to be included in Barro and Ursua (2008) set of GDP disasters. However, the U.S. economy rebounded in 1915.

[9] *Federal Reserve Bulletin*, December 1916, p. 710.

[10] "World's Fair Year Beaten in Every Line of Business" *San Francisco Examiner*, December 30, 1916, p. 21.

major location of war contracts for ships. The ship contracts can be either Navy ships (that are included in the Garrett, 2009 measure) or EFC contracts for merchant ships (which is not).

One limitation of the indicator is that it is not informative about the size of the contracts. As discussed above, Seattle was significantly impacted by the government contracting, while cities such as NYC were not. To better approximate the size of the contracts, we aggregate the maximum ship deadweight tonnage (DWT) of the shipyards awarded contracts. The dummy indicator is then interacted with the DWT of the city to yield a measure of the size of the contract(s) awarded.[11]

We selected these control variables because they were available and capture the impact of the war on manufacturing. To validate these controls, we examine the explanatory power of the war controls for the treatments of interest. NPI days is highly correlated with shipping costs ($p = 0.00$), as by themselves they explain 50% of the variation in NPI days. Garrett war production is uncorrelated with NPI days. However, ship building capacity (DWT interacted with ship contract dummy) explains about 11% of the variation in NPI days ($p = 0.1$), and thus we choose to include it as a control. This results in three additional controls: shipping costs, the ship contract indicator, and the ship contract indicator interacted DWT.

3.2.1. *Results with war controls*

The right hand panels of Figs. 1 and 2 exhibit treatment effects once war controls are included.

For NPI days, prior to treatment, the coefficients become smaller in magnitude, and are statistically indistinguishable from zero at the five percent level. There is no clear upward trend from 1914 to 1918, whereas there is an upward trend without the war controls. In addition, the confidence intervals of the estimated coefficients prior to treatment exclude the point estimates from the model with no controls, indicating that the inclusion of the war controls has a statistically significant impact on point estimates. For example, the 95% confidence interval for 1914 spans 0.059 to −0.163, while the coefficient in the model without controls is −0.184 (left hand panel). Meanwhile, the coefficient for the model with non-war controls is −0.160.

Post treatment, the coefficients are generally statistically indistinguishable from zero, indicating no effect of NPI days on employment. However, the treatment effect for 1920 is positive and statistically significant at a ten percent level (t-stat = 1.714), indicating that there is some weak evidence that cities with tighter NPI had higher levels of employment in 1920. The point estimate is 0.088, indicating that a one standard deviation increase in NPI days increased employment levels by about 9%. The error bounds are rather large in some years, so we cannot exclude quantitatively large effects. The point estimates cluster near zero, so these effects could be either positive or negative.

For NPI speed, we observe little effects on the estimated coefficients once war controls are included. There was no discernible pre-trend in the results and the addition of the war controls has little effect: the coefficients remain statistically indistinguishable from zero. Again, the error bounds post-treatment do not exclude some relatively large effects, though the largest effects tend to be negative for this indicator.

We conclude from these results that war production is indeed an important confounding factor when explaining the effects of NPI days on economic activity in the annual data. Controlling for WWI attenuates the pre-trends that conventional controls do not remove.

Why are NPIs and WWI-induced production correlated? There is a strong common geographical aspect to variation in both NPI intensity and war-induced production. First, the flu is thought to have originated in Boston, which meant that Eastern cities were hit by the flu first with little warning (Barro et al., 2020). Officials in those cities could not impose NPIs until the disease was widespread, so even the strictest eastern cities had relatively few NPI days. Cities further west had the advantage of early warning and the experience of the initial cities (Correia et al., 2022). Second, and as discussed above, Northeastern cities grew comparatively slowly in response to the pre-U.S. entry export boom. Despite having being correlated with geography, our controls are more effective at removing pre-trends. Geographic indicators used in the previous literature, such as distance from Boston (suggested by Barro et al., 2020), or dropping West Coast cities, do not remove the pre-trends we identify.

It is also possible that causality went the other way: Areas with major expansions of manufacturing may have imposed NPIs to protect workers from disease so that production was not disrupted. NPIs were not imposed on factories and closing other public places reduced the opportunity for exposure. There is evidence that this was a concern. For example, Seattle public officials paid for (an ineffective) vaccine for shipyard workers for this reason.[12]

While we think WWI is a reasonable explanation of these pre–trends, our main conclusion is not affected by how well the WWI controls work: There is no model specification that satisfies the parallel trend assumption and shows a statistically significant, long-run impact of NPIs on growth.

3.3. *Robustness checks*

In this subsection we provide a summary of a number of robustness checks. We examine the robustness of our results to alternative data samples and additional controls. Results are depicted in the Appendix.

We consider two alternative samples. First, we drop the interpolated observations. Forty-eight of the 286 observations in our sample are linear interpolations. These interpolations are concentrated in two years (1920 and 1922). We lose statistical significance for 1920 (at the ten percent level) in the NPI days treatment specification, but our other conclusions are otherwise unaffected. Second,

[11] The data sources are reported in the Data Appendix.

[12] Influenza Archive, Seattle essay. https://www.influenzaarchive.org/cities/city-seattle.htm

we expand the time period covered to 1910 to 1924. Between 1910 and 1914 we have 18 of the 26 cities in our baseline sample (although not all of these cities have an observation in every year between 1910 and 1913). Our results remain unaffected under this longer sample.

We also examine how robust our findings are to different controls. We examine what happens with the same set of controls used by LLR. These controls fail to remove pre-trends over the war period. Once our war controls are included in the set, pre-trends disappear. This is further evidence that the confounding impact of the war are not accounted for in the extant set of controls used by the literature.

3.4. WWI in the FCM data

In this section, we examine the impact of our WWI controls on the FCM data. Because WWI controls significantly attenuate pre-trends in the state-collected data, we examine whether these controls help remove pre-trends in analyses that rely on FCM data too. We find that including them attenuates both pre-trends and the treatment effect in the FCM data. These results suggest that our results are not driven by the fact that our set of cities differs from the FCM-Markel dataset[13] They also suggest that the WWI controls are useful for analyses that rely on FCM data.

We compare four specifications: A model without any controls, one with the controls used in CLV, one that further adds our war controls, and one that only includes our war controls. The results are illustrated in Fig. 3.

The first row of the figure depicts results when no controls are included in the model. These results illustrate the pre-trends problem in the FCM dataset: cities with tighter NPIs during the pandemic were growing faster over the previous two decades than cities with looser NPIs. This is more apparent when NPI days is used as the measure of restrictiveness.

The second row depicts results after conditioning on controls. Including controls in the model significantly attenuates the pre–trends to the point where they are statistically insignificant. This facilitates a causal interpretation of the treatment effects in the post-pandemic period.[14] It is also evidence that our results are not driven by the different set of cities.

First we discuss the NPI days treatment. Results are very similar to CLV when we include only their set of controls (although we have more cities, and go back to 1899). The treatment effects in 1919, 1921 and 1923 are 0.0652, 0.061 and 0.054, respectively, and all are significant at a 5% significance level (two-tailed). The inclusion of war controls shrinks the magnitude of the treatment effects and renders them statistically insignificant. This supports the proposition that the war is indeed an important confounding factor in FCM analyses.

Next we discuss the NPI speed treatment. All coefficients are statistically indistinguishable from zero for both sets of controls. As in the state-collected data, we conclude that the war controls are not as relevant when NPI speed is used as the indicator.

The inclusion of war controls reduce the impact of NPIs on employment in the post-treatment period and render the coefficients statistically insignificant. This evidence is consistent with the export boom and war production, not NPIs, having a positive employment effect.

Finally, before moving on, we consider what happens in the FCM-Markel dataset when *only* our WWI controls are included. The appear in the final row of the figure. Interestingly, inclusion of these three controls are sufficient to remove both the pre-trends and the post 1914 treatment effects.

3.5. Discussion

As demonstrated in preceding sub-sections, we find little evidence of a long run impact of NPIs on employment, using either state-collected annual data or FCM data. Other accounts of the 1918 pandemic accord with NPIs having a transitory impact at most. The NPIs imposed on economic production were not very restrictive, so it is not surprising that there is little evidence that they had a long-run impact. Unlike other sectors of the economy, manufacturing was not subject to closures. War production further counteracted NPIs and worked against the disruptions of the influenza (Benmelech and Frydman, 2020). War plants were unlikely to close even when there were disease outbreaks. The war may have further insulated the economy by shifting production into manufacturing and out of non-war related production. Romer (1988) argues that most war production came from crowding out other industries rather than through new production. Some of the industries that were closed by NPI, such as spectator sports, were curtailed as non-essential to the war effort prior to the flu. For example, the World Series would have occurred during the height of the flu without the war. Instead, Major League Baseball played a shortened season that ended before the flu, so games did not have to be canceled or postponed.[15]

Higher frequency data indicates NPIs had a transitory impact. Velde (2022) examines several indicators, like monthly retail trade, and finds short-lived effects. State-collected data show that industries subject to significant NPIs experienced a quick recovery once the restrictions ended. Massachusetts collected unemployment rates of union members by occupation on the last day of each quarter. Theatrical stage employees had a 48 percent unemployment rate on September 30, 1918, likely as a result of theater closures.

[13] In Appendix A.1.2, we show that our sample and the FCM data do not differ along many important dimensions.

[14] At different iterations of their work, CLV and LLR proposed different sets of controls. We obtain similar results with these alternative controls.

[15] Certainly, the NPI used in 1918 were not restrictive compared to those used in most places in response to COVID–19. This must be taken into account when extrapolating findings from studies on the 1918 pandemic to the current context.

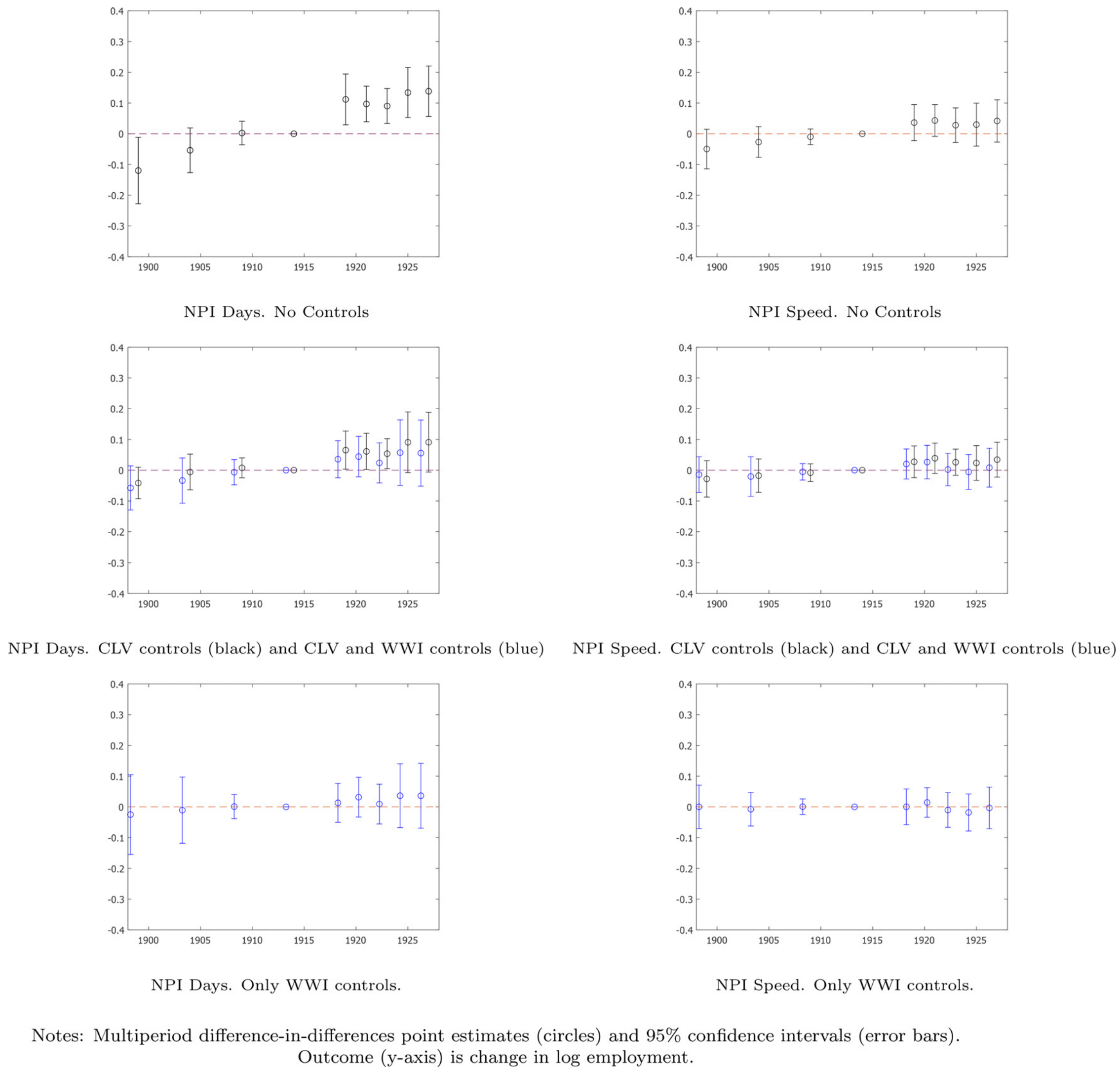


Fig. 3. FCM measures of employment.
Notes: Multiperiod difference-in-differences point estimates (circles) and 95% confidence intervals (error bars). Outcome (*y*-axis) is change in log employment.

Employment quickly recovered, and unemployment dropped to 8 percent on December 31, 1918. This rate is only slightly higher than the 5 percent reported on the same date in 1919.[16] (We compare with the same quarter to account for seasonal effects.)

4. Conclusion

We examine the long-run economic impacts of the policy responses to control pandemics using annual, city-level data and a difference-in-differences framework. We exploit cross-sectional variation in NPI during the 1918 influenza pandemic and find that there is no long-run impact of NPI on employment – either positive or negative.

Preliminary diagnostics indicate that the economic impact of WWI is an important confounding factor in quasi-experimental analyses on the 1918 pandemic. Cities with tighter NPIs grew rapidly in the years immediately prior to the pandemic, and these pre-trends remain even after controlling for variables used in the related literature, such as differences in long-run population growth, industrial composition, and prior flu susceptibility. The pre-trends are only significantly attenuated once we condition on variables related to war production.

[16] Data are reported in *Annual report on the statistics of labor for the year ending November 30, 1921*, Commonwealth of Massachusetts Department of Labor and Industries, part III, p. 42.

Having removed pre-trends, we find no long-run relationship between employment levels and NPIs. We find some short-run positive effects of marginal statistical significance in some specifications, but these are not robust to alternative specifications and samples.

We believe that the war provides a plausible mechanism to explain long run differences in employment growth between cities, as it led to massive changes in manufacturing during this period. For example, shipbuilding in Seattle went from being too small to be separately reported in the employment data to the largest single industry, making it difficult to explain its employment figures without accounting for the war.

Our work also has broader implications for related DID analyses based on FCM data. In our dataset, all of the employment increase in tight NPI cities between the FCM years of 1914 and 1919 occurs before the onset of the pandemic, and thus should not be attributed to NPI that were imposed in 1918. We show that controlling for WWI eliminates long-run positive impacts in Federal data. At minimum, we establish that research on the effect of other mechanisms affecting regional economies will have to satisfactorily confront the war, including work based on lower frequency FCM data.

Declaration of Competing Interest

Authors declare that they have no conflict of interest.

Data availability

Data are available at Bridgman and Greenaway-McGrevy (2023).

Appendix A

A1. Alternative samples

This section presents the robustness exercises summarized above in Section 3.3

Non-interpolated data Forty-eight of the 286 observations in our 1914–24 sample are linear interpolations. We re-run the baseline regression model on the non-interpolated data to examine the extent to which this affects our findings. Figure 4 illustrates our findings. We omit 1922 from the analysis as there are only 9 observations in that year and 6 explanatory variables.

Patterns are generally the same as those exhibited in the interpolated baseline sample. The most notable difference is that the coefficient 1920 in both the NPI days and NPI speed are now very close to zero and insignificant. Eight of the cities in the sample are missing data for 1920.

1910 to 1924 As an additional robustness check we extend the sample to cover 1910 to 1924. Eight of our cities lack any data between 1910 and 1913.

Figure 5 exhibits the results.

A1.1. Alternative controls

NPI regressions with LLR controls

We include the following city-level controls: (log) population in 1900, (log) population in 1914, population density, and manufacturing share of city population (ratio of FCM manufacturing employment in 1914 to census population in 1910), and the following

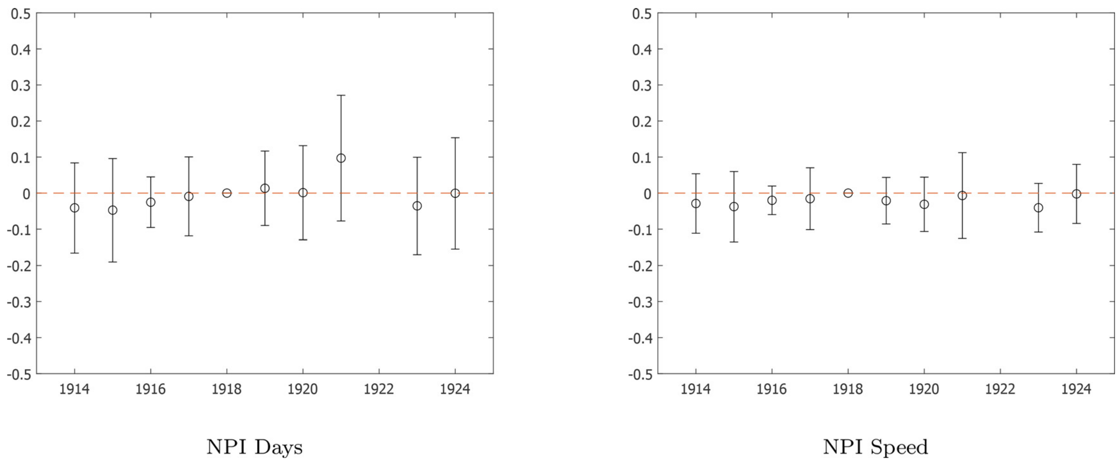


Fig. 4. Robustness check: non-interpolated sample.

Notes: Multiperiod difference-in-differences point estimates (circles) and 95% confidence intervals (error bars). Outcome (y-axis) is change in log employment. The data for 1922 are insufficient to produce an estimate. Model includes WWI controls.

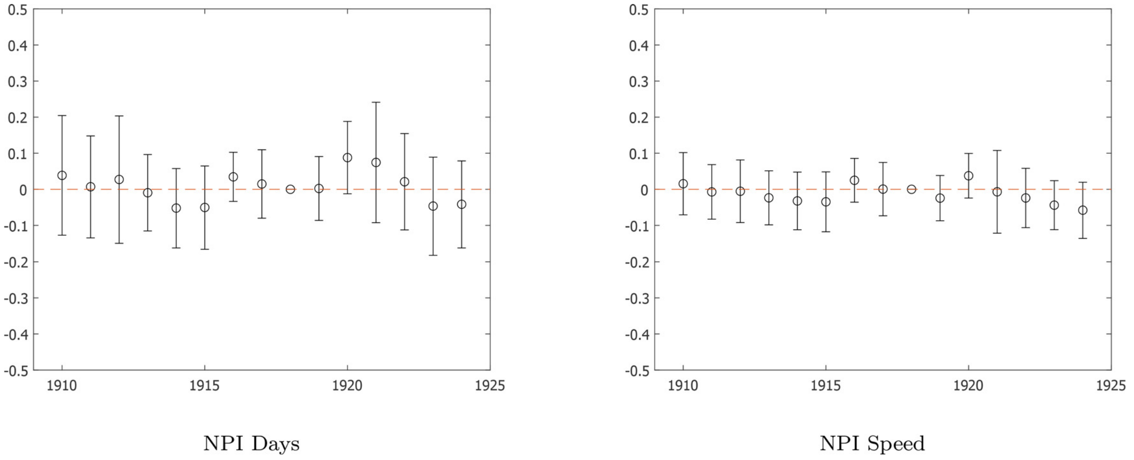


Fig. 5. Robustness check: 1910 to 1924 sample.
Notes: Multiperiod difference-in-differences point estimates (circles) and 95% confidence intervals (error bars). Outcome (y-axis) is change in log employment. Model includes WWI controls.

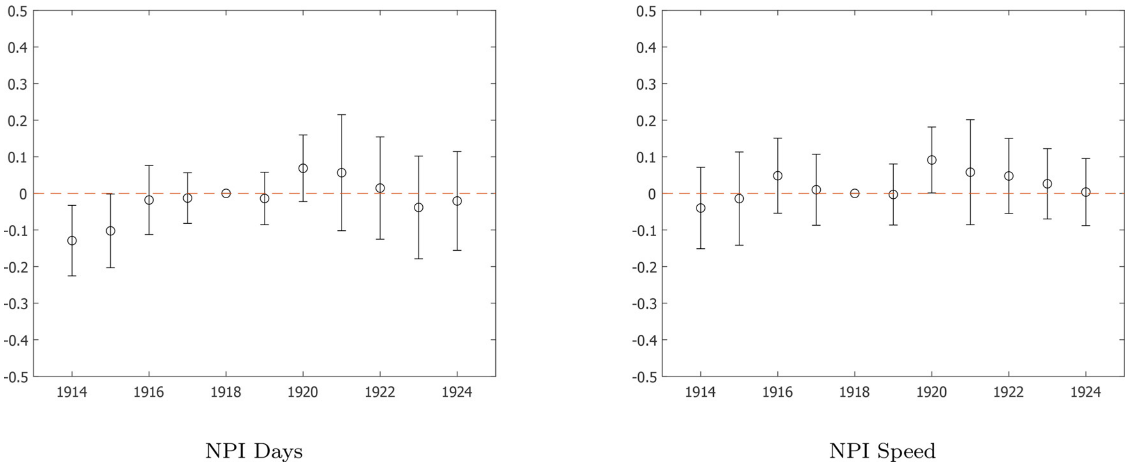


Fig. 6. Robustness check: LLR controls.
Notes: Multiperiod difference-in-differences point estimates (circles) and 95% confidence intervals (error bars). Outcome (y-axis) is change in log employment. Model includes full set of controls used in LLR.

state-level controls: share of manufacturing employment, proportion of population in cities larger than 2500 persons and flu-deaths in 1917. The missing manufacturing share of city population for Charleston is replaced with the sample average. Figure 6 exhibits the results.

NPI regressions with LLR and War controls We add the war controls to the set of LLR controls. Figure 7 exhibits the results.

A1.2. Comparison of FCM and state-collected datasets

In this subsection we investigate the extent to which our state-collected sample of 26 cities is representative of employment patterns in the larger FCM sample of cities. We compare employment treatment effects obtained from the State dataset to those obtained from (a) the FCM data and (b) the FCM data for the cities in our sample. There are 50 cities that have any NPI data, the Markel et al. (2007) sample plus the Berkes et al. (2020) data. Charleston is omitted from the 50 city dataset since it is missing FCM manufacturing employment in 1921, 1923 and 1927, leaving 49 cities in the sample when NPI days is the treatment. NPI speeds are missing for Atlanta, Des Moines, Dallas, San Antonio, Salt Lake City and Charleston. We impute NPI speed values for these cities based on NPI log days using the same method described above. The datasets overlap in 1914, 1919, 1921 and 1923, giving us three post-pandemic observations. Since the point of the exercise is to compare patterns in employment across the various samples, we omit controls from the estimated models. We find that the patterns in employment conditional on the treatment are very similar between the state-collected and FCM datasets.

The first thing we examine is how representative of the policy response the state-data sample is of the larger FCM-data sample. Both the location and scale of the distribution of NPI measures are similar in the two samples. For example, the mean and standard

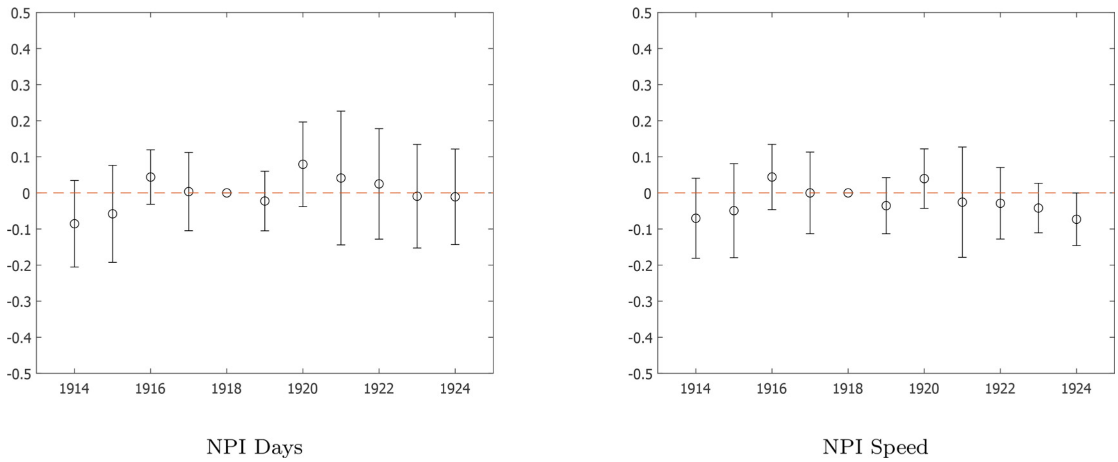


Fig. 7. Robustness check: LLR and WWI controls.
Notes: Multiperiod difference-in-differences point estimates (circles) and 95% confidence intervals (error bars). Outcome (y-axis) is change in log employment. Model includes WWI and full set of controls used in LLR.

deviation of NPI days in the FCM sample of 49 cities is 88.23 and 46.43, while our sample of 26 cities has a mean of 87.0 and a standard deviation of 48.49. The mean and standard deviation of NPI speed in the FCM sample is 7.35 and 7.84, while our sample has a mean of 7.46 and a standard deviation of 8.11. This is consistent with the cities in the state level dataset being representative of the cities in the FCM dataset at least in terms of the variation in NPI measures adopted. There is a similar negative correlation between the two NPI measures in the FCM sample (−0.56) and the State-level sample (−0.67).

Figure 8 exhibits the two NPI treatment effects for three different samples: (1) state-collected data for our original sample of 26 cities (black); (2) FCM data for our original 26 cities minus Charleston SC (pink); and FCM data for 49 cities (blue). Estimates are based to 1914 as the treatment date for comparative purposes. For clarity we only depict the point estimates for the State sample as the confidence intervals are depicted in previous figures.

We make two important observations. First, results for the 25 city FCM sample are very similar to those of the 49 city FCM sample (pink compared to blue). This includes the pre-treatment period of 1899 to 1914. Based on this we conclude that our 25 city sample does not suffer from any noticeable sample selection issues compared to the 49 city sample. There is a loss of statistical precision, with the smaller sample exhibiting larger confidence intervals in each year of the FCM. Second, patterns in the treatment effects look similar between the FCM samples and the state-collected samples in the years in which they coincide. The state-collected sample point estimates for 1919 and 1923 fall within (or very close to) the confidence intervals of the point estimates obtained from the FCM. However, the estimate for 1921 is above the confidence intervals. However, the 95% confidence interval for the 1921 observation (not depicted) does overlap that of the two FCM-based data points.

While all the data sources show an increase from 1914 to 1919, the state data show that most of this increase occurs prior to the flu in 1918. We cannot say for certain whether the FCM data would show the same pattern as they have no observations between 1914 and 1919. However, the patterns are very similar for the dates that do overlap, which suggests that long-run positive effects in the FCM data may be an artifact of growth that precedes the flu. In the state-collected data, we showed that accounting for WWI eliminates the 1914 to 1918 pre-trends.

A2. Data appendix

All data and replication code are available at Bridgman and Greenaway-McGrevy (2023).

A2.1. State-collected employment data

For years of coverage, see Table 1.

California

Employment and payroll for all manufacturing establishments from the *Biennial report of the Bureau of Labor Statistics of California*.

Connecticut

Biennial Report of the Factory Inspector to the Governor, State of Connecticut. We use New Haven County for New Haven City.

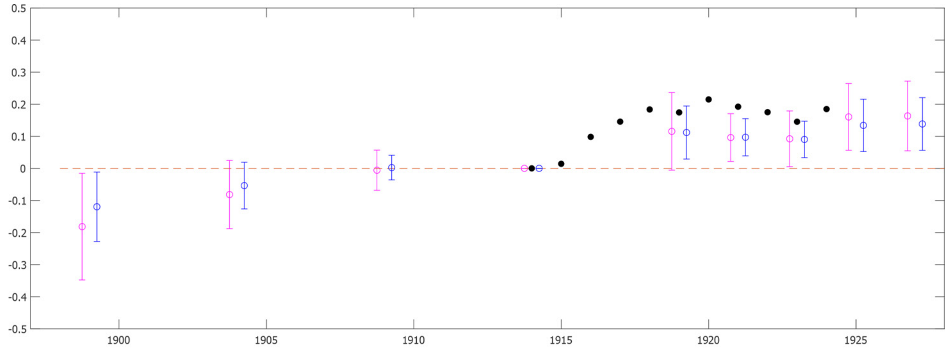
Iowa

Report of the Bureau of Labor Statistics, State of Iowa. We use Polk County for Des Moines City.

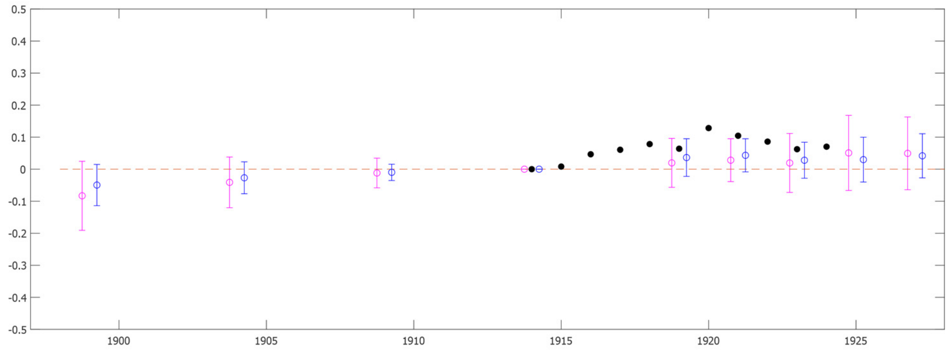
Massachusetts

Annual report on the statistics of manufactures, Commonwealth of Massachusetts.

Michigan



NPI Days



NPI Speed

Fig. 8. Treatment effects estimated from FCM and state-collected data sources.
Notes: Multiperiod difference-in-differences for (log) employment. Black is state-collected data for 26 cities. Pink is FCM data on those 26 cities. Blue is FCM data for 50 cities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Employment and average daily total payroll from *Annual Report of the Department of Labor of the State of Michigan*. Data collected by factory inspectors and cover of all manufacturing establishments.

Missouri

Annual report of the Bureau of Labor Statistics, State of Missouri.

New Jersey Annual Report of the Bureau of Industrial Statistics of New Jersey.

New York

1915–July 1921: *The Labor Market* New York Department of Labor, Bureau of Statistics and Information. August 1921–1924: *Industrial Bulletin* New York Department of Labor. Employment in a sample of "Representative Manufacturing Plants."

Ohio We combine two data collections. For 1910–12, the Bureau of Labor Statistics collected an annual Census of Manufactures. From 1914 to 1924, we use administrative data from the Workmen's Compensation program. Its coverage was extraordinarily broad, covering all employers in all industries aside from interstate railroads.

1905–1912: *Annual report of the Bureau of Labor Statistics, made to the General Assembly of Ohio*.

1914–1924: Monthly manufacturing wage-earners from Table 163 in Spurgeon Bell and Ralph J. Watkins (1928). *Industrial and Commercial Ohio*, Bureau of Business Research, College of Commerce and Administration, Ohio State University. The data are reported at the county level. We use the county in which each city is located: Cuyahoga (Cleveland), Franklin (Columbus), Hamilton (Cincinnati), Lucas (Toledo), and Montgomery (Dayton). Coverage is firms with 5+ employees (1914–1923) and firms with 3+ employees (1924). Data for 1922 are missing since the source data were lost to a fire prior to compilation.

Pennsylvania Annual Report of the Commissioner of Labor and Industry of the Commonwealth of Pennsylvania. Data are reported at the county level. We use Allegheny County for Pittsburgh. Philadelphia City and County coincide.

Rhode Island

Employment data from various issues of *Annual report of the factory inspectors*, State of Rhode Island and Providence Plantations. Data cover non-agricultural establishments employing 5 or more persons.

South Carolina

Annual Report of the Department of Agriculture, Commerce and Industries of the State of South Carolina. We use Charleston County for Charleston City.

Washington State

Employment data from various issues of *Biennial Report of the Bureau of Labor Statistics and Factory Inspection*, Department of Labor of Washington State. The 1910 and 1912 data come from a census of manufactures collection by county. We use Spokane County for Spokane City and King County for Seattle. We have factory inspection data for 1914, 1915, and 1918. These data cover the factory inspection districts centered in Seattle and Spokane, so cover establishments outside these cities. However, the vast majority of establishments covered are within city limits.

A2.2. Flu mortality control data

Alabama

Alabama State Board of Health Annual Report 1917, p. 76: 392 Flu/1951 Pneumonia deaths 1917; 1917 population: 2,354,670.

District of Columbia/Utah

1920 Mortality, Bureau of the Census, p. 30.

Georgia/Nebraska/Oregon

Use city level (Atlanta, Portland, Omaha) data from 1920 Mortality, Bureau of the Census, p. 30.

Illinois

Annual report of the Department of Public Health 1918, p.45. 8277 Pneumonia deaths, fiscal year ending June 30, 1918. Population 6,276,364.

Iowa

The flu fatality rate for Iowa in 1917 is taken from Des Moines. Source: 19th Biennial Report of the Iowa State Board of Health, Table 5 (pp.27-8) gives 1917 deaths from flu (5) and pneumonia (112) in 1917. Divided by 1917 population.

Louisiana

Biennial Report of the Louisiana State Board of Health to the General Assembly, 1916-17. 1774 Pneumonia deaths/30 Influenza deaths (Under age 1 only), calendar year 1917. (Mortality tables 3 and 5) State population: 1,864,533 (Table 3).

Texas

Texas Department of Health Biennial Report: 2890 Pneumonia/705 influenza deaths in 1917. (p.20). Population: 4515423 (July 1, 1917) p. 11.

A2.3. WWI controls

Navy ship contracts Source is the index of the US Navy Shipbuilding Contracts in the Navy Department Library (https://www.history.navy.mil/research/library/research-guides/shipbuilding-contracts/us-navy-shipbuilding-contracts-in-the-navy-department-library.html) We use the contracts in Volume 18, pt. 1, which covers the major ship contracts let out during hostilities. We omit pt. 2 since these cover smaller contracts (such as paint) and Volume 19 since they are smaller watercraft or contracts let out after the war.

The contractors we include are William Cramp (Philadelphia) and Union Iron Works (San Francisco).

William Cramp was an EFC contractor with maximum DWT of 10,000 DWT.

Union Iron Works were completed a 11,800 ton steamer in 1918. (Cordage Trade Journal, Vol. 56, no. 11, p. 242)

EFC Contracts/capacity Source is U.S. Shipping Board's 1919 Annual Report, Appendix II, Table I. Variable set to one if a city is listed as having a contract from the Emergency Fleet Corporation. Capacity is maximum deadweight tonnage that could be built.

The capacity of yards with canceled or completed contracts (as of the 1919 report) were not reported. We impute their capacity using information about ships they built in that period.

Patterson–McDonald (Seattle, WA): Built motorships sold to Australian Govt, 4400 DWT as reported in The Rudder (No. 1918, p. 522).

Elliott Bay (Seattle, WA): Built motor ships sold to Norway, 3350 DWT as reported in Motorship (January 1920, p. 45).

Kiernan & Kern (Portland, OR): Built a 2239 DWT cargo ship in 1918 (http://shipbuildinghistory.com/shipyards/19thcentury/otherpc.htm).

Southern Shipbuilding (Charleston SC): We set capacity to zero since this yard never started any of its contracted ships, leading to its contract being cancelled. (Shipping, March 9, 1918, p. 493) Therefore, the contract likely had no employment effect.

References

Abramitzky, R., Ager, P., Boustan, L.P., Cohen, E., Hansen, C.W., 2019. The Effects of Immigration on the Economy: Lessons from the 1920s Border Closure. Technical Report. NBER Working Paper 26536.

Ager, P., Eriksson, K., Karger, E., Nencka, P., Thomasson, M.A., 2022. School closures during the 1918 pandemic.

Almond, D., 2006. Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 U.S. population. J. Polit. Econ. 114 (4), 672–712.

Anderson, H., Chang, J.-W., Copeland, A., 2020. The Effect of the Central Bank Liquidity Support during Pandemics: Evidence from the 1918 Influenza Pandemic. Technical Report. New York Federal Reserve Bank.

Angrist, J.D., Pishcke, J.-S., 2009. Mostly Harmless Econometrics. Princeton University Press.

Arthi, V., Parman, J., 2021. Disease, downturns, and wellbeing: economic history and the long-run impacts of COVID-19. Explor. Econ. Hist. 79, 101381.

Atolia, M., Papageorgiou, C., Stephen, J., Turnovsky, S.J., 2021. Re-opening after the lockdown: long-run aggregate and distributional consequences of COVID-19. J. Math. Econ. 93.

Barro, R.J., Ursua, J.F., 2008. Macroeconomic crises since 1870. In: Brookings Papers on Economic Activity, pp. 255–350.

Barro, R.J., Ursua, J.F., Weng, J., 2020. The Coronavirus and the Great Infuenza Pandemic: Lessons from the Spanish Flu for the Coronavirus's. Technical Report. NBER Working Paper 26866.

Basco, S., Domenech, J., Roses, J., 2020. The redistributive effects of pandemics: lessons on the Spanish flu. Instituto Figuerola de Historia y Ciencias Sociales Working Paper in Economic History 2020-05.

Beach, B., Brown, R.R., Ferrie, J., Saavedra, M., Thomas, D., 2022a. Re-evaluating the long-term effects of in utero exposure to the 1918 influenza pandemic. J. Polit. Econ. 130 (7), 1963–1990.

Benmelech, E. Frydman, C. 2020. The 1918 influenza did not kill the US economy. VoxEU April 29, https://voxeu.org/article/1918-influenza-did-not-kill-us-economy

Beach, B., Clay, K., Saavedra, M., 2022b. The 1918 influenza pandemic and its lessons for COVID-19. J. Econ. Lit. 60 (1), 41–84.

Berkes, E., Deschenes, O., Gaetani, R., Lin, J., Severen, C., 2023. Lockdowns and innovation: evidence from the 1918 flu pandemic. Rev. Econ. Stat., forthcoming 1–30.

Bloom, D., Mahal, A.A., 1997. AIDS, flu, and the black death: impacts on economic growth and well-being. In: The Economics of HIV and AIDS: The Case of South and South East Asia. Oxford University Press, Delhi, pp. 22–52.

Bodenhorn, H., 2020. Business in a Time of Spanish Influenza. Technical Report. NBER Working Paper.

Brainerd, E., Siegler, M., 2003. The Economic Effects of the 1918 Influenza Epidemic. CEPR Discussion Paper.

Bridgman, B. Greenaway-McGrevy, R. Data and Code for "The Economic Impact of Social Distancing: Evidence from State-Collected Data During the 1918 Influenza Pandemic" Inter-University Consortium for Political and Social Research. Ann Arbor, MI10.3886/E191461V1

Correia, S. Luck, S. Verner, E. 2020. Response to Lilley, Lilley, and Rinaldi (2020)" mimeo, MIT Department of Finance

Correia, S., Luck, S., Verner, E., 2022. Pandemics depress the economy, public health interventions do not: evidence from the 1918 flu. J. Econ. Hist. 82 (4), 917–957.

Croxton, F.E., Croxton, F.C., 1932. Fluctuation in employment in Ohio, 1914–1929. Bulletin of the U.S. Bureau of Labor Statistics.

Dahl, C.M., Hansen, C.W., Jensen, P.S., 2020. The 1918 epidemic and a V-shaped recession: evidence from municipal income data. CEPR COVID Econ. 6, 137–162.

Donaldson, D., Hornbeck, R., 2016. "Railroads and American economic growth: a 'market access' approach. Q. J. Econ. 131 (2), 799–858.

Elenev, V., Landvoigt, T.T., Nieuwerburgh, S.V., 2022. Can the COVID bailouts save the economy? Econ. Policy 37 (110), 277–330.

Fluegge, R.B., 2022. Death, Destruction, and Growth in Cities: Entrepreneurial Capital and Economic Geography After the 1918 Influenza. mimeo. Harvard University.

Fordham, B.O., 2007. Revisionism reconsidered: exports and american intervention in world war I. Int. Organ. 61 (2), 277–310.

Fuchs-Schundeln, N., Krueger, D., Ludwig, A., Popova, I., 2022. The long term distributional and welfare effects of COVID-19 school closures. Econ. J. 132 (645), 1647–1683.

Garrett, T.A., 2009. War and pestilence as labor market shocks: U. S. manufacturing wage growth 1914–1919. Econ. Inq. 47 (4), 711–725.

Goolsbee, A., Syverson, C., 2021. Fear, lockdown, and diversion: comparing drivers of pandemic economic decline 2020'. J. Public Econ. 193. doi:10.1016/j.jpubeco.2020.104311.

James, S., Sargent, T., 2006. The economic impact of an influenza pandemic, Canada Department of Finances Working Paper 2007–04.

Karlsson, M., Nilsson, T., Pichlere, S., 2014. deltThe impact of the 1918 Spanish flu epidemic on economic performance in sweden: an investigation into the consequences of an extraordinary mortality shock. J. Health Econ. 36, 1–19.

Lilley, A., Lillley, M., Rinaldi, G., 2020. Public Health Interventions and Economic Growth: Revisiting the 1918 Flu Evidence. mimeo. Harvard University.

Lipsey, R.E., 1963. Price and Quantity Trends in the Foreign Trade of the United States. Princeton University Press, Princeton NJ.

Markel, H., Lipman, H.B., Navarro, J.A., Sloan, A., Michalsen, J.R., Stern, A.M., Cetron, M.S., 2007. Nonpharmaceutical interventions implemented by us cities during the 1918–1919 influenza pandemic. J. Am. Med. Assoc. 298 (6), 644–654.

Maurer, S., Rauch, F., 2022. Economic geography aspects of the Panama Canal. In: Oxford Economic Papers, pp. 1–21.

Moser, P., San, S., 2020. Immigration, Science, and Invention: Evidence from the Quota Acts.

Noy, I. Okubo, T., Strobl, E., 2020. The japanese textile sector during the pandemic influenza of 1918–1920. CESifo Working Paper no. 8651

Psacharopoulos, G., Collis, V., Patrinos, H. A., Vegas, E., 2021. The COVID-19 cost of school closures in earnings and income across the worldComp. Educ. Rev. 65(2), 271–287.

Rambachan, A., Roth, J., 2022. A More Credible Approach to Parallel Trends. mimeo. Brown University.

Rhode, P.W., 2001. The Evolution of California Manufacturing. Public Policy Institute of California, San Francisco.

Rockoff, H., 2020. Off to a Good Start: The NBER and the Measurement of National Income. NBER Working Paper No. 26895

Romer, C.D., 1988. World war I and the postwar depression: areinterpretation based on alternative estimates of GNP. J. Monet. Econ. 22 (1), 91–115.

Sprague, O., 1915. The crisis of 1914 in the United States. Am. Econ. Rev. 5 (3), 499–533.

Sundstrom, W.A., 1990. Was there a golden age of flexible wages? Evidence from ohio manufacturing, 1892. J. Econ. Hist. 50 (2), 309–320.

Velde, F. R., 2022What happened to the us economy during the 1918 influenza pandemic? A view through high-frequency data, J. Econ. Hist. 82(1), 284–326