Essays in Empirical Corporate Finance and Labor Economics

Omar Hossain Ahsan

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023

Abstract

Essays in Empirical Corporate Finance and Labor Economics

Omar Hossain Ahsan

In the first chapter of this dissertation, I exploit the Covid-19 pandemic and associated government restrictions as a natural experiment in order to study the resilience of businesses in the United States. I use a border-county identification strategy with data on government restrictions, employment and open small businesses, in order to assess the resilience of small businesses in the United States. In my main results, I find negative impacts of stay-at-home orders on the number of open small merchants. In particular, shutdowns of businesses accelerated 8 weeks after imposition of a stay-at-home order, suggesting that many businesses were only resilient enough to handle adverse conditions for 8 weeks. On average, a county with a stay-at-home order experienced an additional 1.51 percentage points loss in the number of open small businesses, relative to January 2020, 8 weeks later compared to a neighboring county that did not have a stay-at-home order. Firms were quicker to resort to layoffs. On average a county with an active stay at home order in a month experienced an additional 1.19 percentage point loss in employment, relative to January 2020, the following month compared to a neighbor that did not have a stay-at-home order the previous month. My results suggest that in future scenarios where governments consider enacting similar restrictions further aid is needed for businesses in order to help them stay afloat. In particular, more assistance should be delivered to businesses within two months from the enacting of the order.

In the second chapter of this dissertation, I study economic spillovers in the context of the Covid-19 associated government restrictions. I use a detailed geolocation dataset to construct data on the number of visitors per-capita between neighboring counties in the early stages of the pandemic, which I use as a proxy for economic spillovers. I employ a similar border-county identification strategy as in the first chapter to identify the causal effect of stay-at-home orders on inter-county movement. Additionally, I provide evidence for an assumption used in chapter one by examining if there are reduced spillovers in county-pairs that lie in the different commute zones. I find that stay-at-home orders caused reductions in inter-county visits in both directions in a county-pair. That is, I find a decrease in travel from the county without a stay-at-home order to the county with one, as well as a decrease in the opposite direction. On average, a county that does not have stay-at-home order will receive 408 fewer weekly visitors from their neighboring county that has a stay-at-home order. I also examine the effect of stay-at-home orders on the ratio of travel between the two directions in order to find evidence of a net spillover effect between the two counties and fail to find evidence of a net spillover effect. I also find that spillover effects are indeed reduced in neighbor county-pairs where the two counties are in different commute zones. The results of this paper imply that residents in counties with stay-at-home orders decreased travel to their neighboring counties even when those counties did not issue their own orders. In future situations where policy makers need to consider similar restrictions, they should focus on acting more quickly and not be concerned if neighboring counties are not cooperative.

In the third chapter of this dissertation, I test the predictions of career concerns models by studying Major League Baseball umpires. Major League Baseball games can be dramatically shaped by minor lapses in judgement from the umpires officiating the game. Due to the indefinite length a game may have, this can include having the game shaped in a way that ends it faster. I study whether evidence for the career concerns model can be found in baseball umpires. A career concerns model would suggest that older umpires, whose careers and reputations are much more established than younger ones, would be more inclined to improperly make judgements that favor the end of the game in extra innings. I use data on MLB umpires and extra-innings games from the 2010-2018 seasons to conduct my empirical analysis and use a linear probability model to isolate the impact of the umpires' tenure on the probability they make a "bad call." I find evidence supporting the career concerns hypothesis and that the probability that an umpire makes a bad call that shortens the length of the game and allows them to go home increases with their tenure. I show that these results are likely driven by career concerns, rather than carelessness, by showing their error rate does not increase with tenure in situations where it would not reduce their workload.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This dissertation would not be completed without the extremely valuable advice and guidance I received from many faculty members at both the Economics Department and Business School at Columbia University. I would like to express my sincere gratitude to my sponsors, Xavier Giroud and Glenn Hubbard, whose guidance throughout the process of working on this thesis was extremely helpful. I would also like to thank Sandra Black, Suresh Naidu and Noémie Pinardon-Touati for their invaluable advice at several points in working on this thesis and for serving on my defense committee. I am also grateful to the support and advice I received from the director of graduate studies, Navin Kartik, throughout my time at Columbia. None of my work at Columbia would have been possible without the help of my undergraduate faculty advisers at the University of Maryland-College Park. I would also like to thank Peter Cramton, Nuno Limão and Howard Elman for their help in my previous work.

This thesis would also not be complete without the strong support I received from my family and friends. I am forever indebted to my wife, Helen Guan, for supporting me throughout the entire PhD program. I am also grateful to my father, Ahmad Ahsan, my mother, Ishrat Hossain, my sister Ariana Nahar Ahsan, and my nanny Laila Begum for their lifelong support. I also would like to acknowledge support from my entire family, including all my aunts and uncles, in particular Tanjim Hossain and Migiwa Tanaka.

Finally, I would like to thank the rest of the PhD students in my cohort at Columbia University. Their presence throughout the program helped me overcome some of the most difficult challenges and also made my experience as a graduate student far more enjoyable. I would like to specifically thank Bookang Seol, Naman Garg and Krzystof Zaremba for their help in helping me navigate the PhD program and countless practice sessions before colloquium presentations.

Dedication

The process of working on the papers that compose this dissertation was made significantly easier with the fervent support of my grandmother, Gulnahar Hossain, who passed away in May of 2022. This dissertation is dedicated to the memory of all four of my grandparents, Shah Mohammad Ahsan, Shamsun Nahar Ahsan, Gulnahar Hossain and Tofazzal Hossain.

Chapter 1: Resilience in U.S. Firms: Evidence from the Covid-19 Pandemic

1.1 Introduction

Much attention has been paid to the resilience, or lack thereof, of households and governments to unexpected shocks. A survey by bankrate.com finds that just 39 percent of Americans can afford to pay an unexpected need for 1000 dollars. This inadequacy is despite the fact that such expenses are not uncommon, with researchers at the Federal Reserve finding that 17 percent of Americans incurred unexpected medical expenses between $1000 and $2000. There is a rich body of work examining the resilience of consumers and governments and factors impacting it. Papers such as Klapper and Lusardi (2020) and Hussain et. al (2019) have studied ways to improve consumer financial resilience. Resilience of local governments has likewise been studied in papers such as Ahrens and Ferry (2020) which examines the resilience of English local governments in the wake of Covid-19.

Less attention has been given to the resilience of businesses. Like households, businesses are also subjected to unexpected shocks on both the supply and demand side. This paper contributes to the growing body of work on the resilience of businesses to unexpected shocks. Other papers have been written about the resilience of financial markets and resilience of firms by studying cash balances. This paper more directly studies impacts on firm operations, such as remaining open and not resorting to layoffs, as a measure of resilience. How resilient are businesses in the United States to unexpected shocks? How long can firms withstand the expenses associated with unexpected shocks (such as a forced closure) before they are forced to start laying off employees? How long can they remain open at all? In this paper, I exploit variation in state responses to the Covid-19 pandemic to help answer these questions and more.

The Covid-19 pandemic wreaked havoc throughout the world in the early 2020s. At the time of writing, the disease had claimed just under 1 million deaths in the United States alone, and over 6 million deaths worldwide. The damage was not limited to health outcomes. Unemployment in the United States skyrocketed in the initial wave of the pandemic, rising from 3.5 percent in February to 14.7 percent in April. Without knowledge of the most effective treatment methods or vaccines, governments throughout the world largely resorted to non-pharmaceutical interventions (NPIs) in order to combat the virus. Some of the most restrictive NPIs used by many states in the U.S. were "Stay-at-Home Orders," which closed all non-essential businesses from operating and banned many types of gatherings.

The economic hardships faced by many individuals and businesses in this time made many of these NPIs quite controversial. Some argued that public health orders must take into account their impact on the economy and that the "cure" should not be made worse than the disease. Others attributed the economic impact as instead directly stemming from the disease itself and that the economy could not be saved without focusing on reducing the outbreak first. In this paper, I quantify the impacts of the government enacted NPIs versus the impacts of the underlying pandemic itself. To answer this question, I employ a border-county strategy in order to look at differences in outcomes caused by differences in government response. My identifying assumption here is that neighboring counties tend to be relatively similar. Specifically, that unobserved factors that may be relevant will also be similar between the two counties. One possible issue with this strategy is that it may be affected by county-to-county economic spillovers caused by stay-at-home orders. Strong spillovers may bias the results and exaggerate the effects of stay-at-home orders on businesses. In order to mitigate possible spillovers, I focus my results on a subset of border-counties that lie in separate commute zones. In the second chapter of this dissertation, I use a geolocation dataset to estimate spillover effects and find that there is no evidence of a net economic spillover from counties to their neighbors caused by stay-at-home orders.

In my main results, I find that Covid-19 related stay-at-home orders negatively affected business survival and employment. In particular, I find an acceleration of business shutdowns approximately 8 weeks after the implementation of a stay-at-home order in the affected county. This suggests that many businesses were only resilient enough to handle adverse conditions for about two months. On the employment side, losses of jobs peak approximately 1 month after the implementation of the stay-at-home order. This is consistent with the idea that layoffs are a less severe negative outcome than firm closure, which makes it natural that businesses would implement layoffs earlier than shutting down entirely. Additionally, there is evidence that the labor effects rebound faster than the firm shutdown effects.

There are two main strands of research which this paper contributes to. First, this paper contributes to the studies of the resilience of businesses. Piccolo and Pinto (2021) highlight the ance of businesses financial resilience and connect it to other issues such as labor negotiations. Papers such as Farrel and Wheat (2018) have examined the financial resilience of firms in the wake of disasters by using cash balances as a measure of resilience. This paper contributes to the business resilience literature by using the pandemic as a natural experiment to study a specific kind of resilience: the ability of firms to avoid layoffs and shutting down during the challenging conditions created by the pandemic.

Second, the paper also adds to the diverse body of work on the economic impacts of the Covid-19 pandemic. Papers such as Chetty et. al (2021), Crotes and Forsythe (2020), and Deryugina, Shruchkov and Stears (2021) all examine impacts of the pandemic on various kinds of workers. This paper contributes to this literature through studying the county-level employment effects of the pandemic. On the firm side, papers such as Bartik et. al (2019), Bloom, Fletcher and Yeh (2021), Bloom et. al (2021) all examine the effects of the pandemic on firms. This paper contributes to this literature through studying the impacts of the pandemic on firm closures.

A related body of work looks at the impact of government restrictions on both economic and health outcomes. Particularly relevant to this paper is Spiegel and Tookes (2021), which examines the impact of various Covid-19 government restrictions on deaths at a county level and contributes a novel data set which has detailed information on county level restrictions throughout the United States. Papers such as Amuedo-Dorantes (2020), Alexander and Karger (2021) and Caselli et. al (2020) examine other impacts of the stay-at-home orders and other NPIs. I contribute to this literature by conducting an event study on the impacts of a stay-at-home order on both employment and business closures as well as a more general difference-in-difference specification.

The rest of this paper is organized as follows. Section 2 goes through the data used in this project in detail. Section 3 examines the variation in state responses to the pandemic and explains the identification strategy used. Section 4 introduces the empirical specifications I estimate and results. Section 5 concludes.

1.2 Data

I use several different sources of data in this project which I outline below.

1.2.1 Main Outcome Variables

My primary research questions look at two different outcomes as proxies for resilience: employment and the number of open small businesses. Specifically, I studied the resilience of firms insofar as avoiding layoffs, and avoiding a shutdown.

Data on the number of businesses is from Womply and accessed via the Opportunity Insights data contributed by Chetty et. al. (2021). The Womply measure of open small businesses is weekly and collected at the county level. It is reported in each county as the percentage change in the number of open small businesses compared to January 2020.

Employment data comes from the Bureau of Labor Statistics' Quarterly Census of Employment and Wages which contains monthly data on employment at the county level. In order to interpret results in a similar manner to the merchants data, I convert the raw employment data to a percentage change as well using January 2020 as a baseline.

Both of these dependent variables are used to study the primary question of the effect of the pandemic and government restrictions on the resilience of firms in the United States. Two different outcomes are used in order to examine resilience in two different settings. Employment data is used to study the effect of the pandemic and government policies on resilience by interpreting resilience as the ability of a firm to withstand conditions without laying off workers. The number of open small businesses is used to study resilience as the ability of firms to remain operational entirely. These variables are chosen as both layoffs and firm failures are major negative outcomes for firms.

1.2.2 Government Restrictions Data

For this project, I use two measures of government restrictions. The first is the stringency index constructed by researchers at Oxford's Government Response Tracker. The second is data collected on the start and end dates of various NPIs from Spiegel and Tookes (2021).

The stringency index is an ordinal measure of the intensity of a state's collection of NPIs. This includes measures such as stay at home orders, school closures, restaurant capacity constraints, as well as others. It is generated daily at the country level as well as at the state level in the United States throughout the pandemic. For this project, I primarily use the stringency index as motivation for my identification strategy. Table 1.1 below displays some summary statistics about the stringency index during the time period of interest.

Table 1.1: Stringency Index summary statistics from March 2020 to December 2020. The left column shows variation of the stringency index over time across states. The right column shows variation across states in the average stringency index in this time period

	Stringency Index	Stringency Index State Average
N	17085	51
Mean	45.099	45.099
Std. Dev.	21.917	11.190
Min	0	0
25th percentile	35.19	39.993
Median	47.69	45.281
75th percentile	61.11	51.533
Max	87.96	66.452

I also use data on start and end dates of various county level measures from Spiegel and Tookes (2021). This data contains information on when a diverse set of NPIs were enacted and terminated throughout all the counties in the United States. In some cases, start and/or end dates are listed at the state-level instead of county when directives are enacted by the state. For this project, I focus on the usage of stay-at-home orders. Stay-at-home orders were the most stringent level of restrictions enacted by state and local governments in the United States. Unlike other countries, American governments largely did not prevent people from leaving their homes entirely. People were instead instructed to not leave their homes for non-essential reasons. The stay-at-home orders were paired with mandates forcing the closure of non-essential business, such as gyms and movie theaters. Stay-at-home orders were chosen to be studied for this project as they were the most extreme measure taken in the early stages of the pandemic and often thought to have been the one to cause the most economic hardship.

1.2.3 Neighbor Counties

In order to match counties with their neighbors, I use the County Adjacency File which contains a list of neighboring counties for each county in the United States. I then match counties based on this data to create a list of all county-pairs that share a border. I manually removed some county-pairs that were considered to be neighboring by the county adjacency file but only shared a maritime border in actuality. This included some county-pairs consisting of Eastern Wisconsin and Western Michigan counties, which were labeled as neighbors but are only connected by Lake Michigan. In order to avoid double counting county-pairs, I toss all county-pairs where the main county's FIPS code is larger than its neighbors. For example, in the case of Chemung County, New York (FIPS 36015) and Bradford County, Pennsylvania (42015), the pair (36015, 42015) is kept while (42015, 36015) is tossed. Note that counties may appear in more than one county-pair. Bradford County, PA appears in pairs with both Chemung County, NY and Tioga County, NY.

1.2.4 Other Controls

I take daily data on COVID-19 death rates are from the Opportunity Insights project that is also mentioned above. Deaths are used instead of cases as they are likely to have been less undercounted than cases, especially in the early stages of the pandemic where tests were limited to severe cases. The deaths are reported as 7-day averages in order to account for day-of-week trends that are present due to how most health agencies report data.

In order to control for the fact that the pandemic affected different industries in different ways, I use data from the County Business Pattern to control for the percentage of business in each county that are in various NAICS super-sectors. Some industries, such as food services and entertainment are thought to have been especially hindered by the pandemic. As such, counties that have an especially high concentration of these industries may suffer additional losses of businesses and employment independent of their government policies.

Political controls are taken from the Massachusetts Institute of Technology Election Lab as a proxy for the tendency for people in counties to obey restrictions and follow voluntary social-distancing measures. Gollwitzer et. al (2020) find that counties that had more Hillary Clinton voters exhibited more social distancing. Specifically, I use the average two-party vote share received by the Democratic candidate in the 2012 and 2016 presidential elections. Controlling for politics will also help account for differences in propensity for governments to enact restrictions when facing waves of the virus. While governments consider many factors when deciding their policy, one of the most significant concerns for them is their electoral concerns. Support for stringent restrictions became a highly politicized issue, and conservative and conservative-leaning people became far more likely to be opposed to restrictions as the pandemic unfolded. As such, heavily Republican governments were less likely to impose restrictions than their Democratic counterparts.

Access to financing was crucial for businesses during the pandemic. Many small businesses were only able to access the emergency loans by filing applications with their banks. As a proxy for access to financing, I compute the number of bank branches (per capita) using data from the Consumer Financial Protection Bureau. Small businesses in particular are more likely to file applications at a local bank branch.

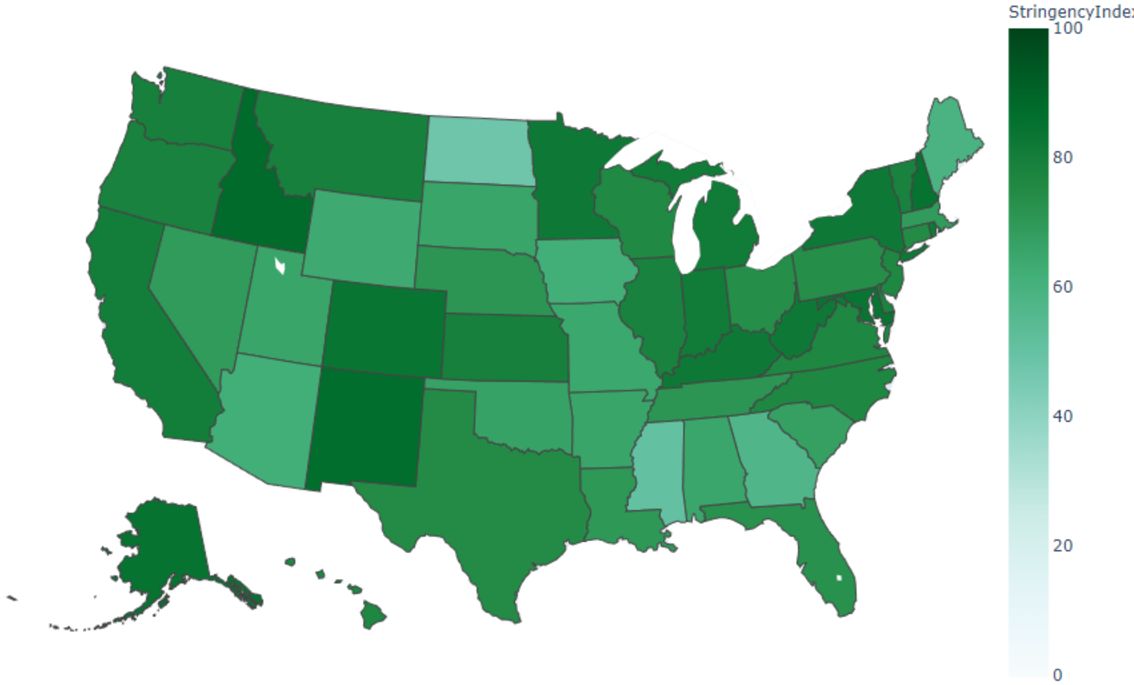
Lastly, in order to account for the possibility that spillovers drive my main results, I use data from Autor and Dorn (2013) to map counties to commute zones. I then estimate my main specifications on a set of border counties that do not lie in the same commute zone. I assume here that neighboring counties in different commute zones will have less movement between them, and thus, less spillovers. I discuss this issue in greater detail and show evidence for this assumption in the second chapter of this dissertation.

1.3 Identification Strategy

1.3.1 Variation in Government Actions

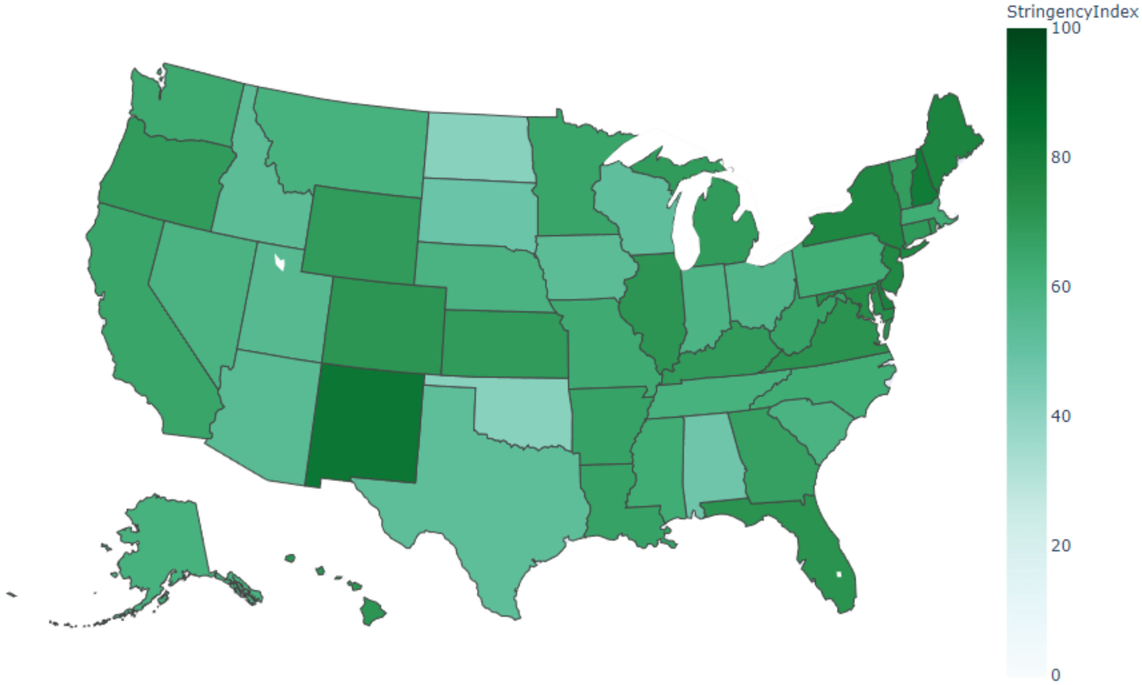
My specification relies on the existence of variation in government policies both across counties and over time. As very few restrictions were enacted federally in the first year of the pandemic, much of the decision making was left to state and local authorities. As such, due to differences in policy makers' preferences and politics, in many cases neighboring states imposed different levels of restrictions. To see some of the variation across states, consider Figure 1.1 below which shows the variation in state measures near the beginning of the first wave of the pandemic.

Figure 1.1: Stringency Index by state on April 1st, 2020. Darker is more stringent.



As seen in Figure 1.1, the intensity of government restrictions is quite varied across states. The first wave of the pandemic in the United States was largely more severe in the midwest and northeast, especially in the New York City area. As seen in the map, many of the darker states are indeed located in the northeast and midwestern regions of the country. Next, consider the situation 2 months later in the map below.

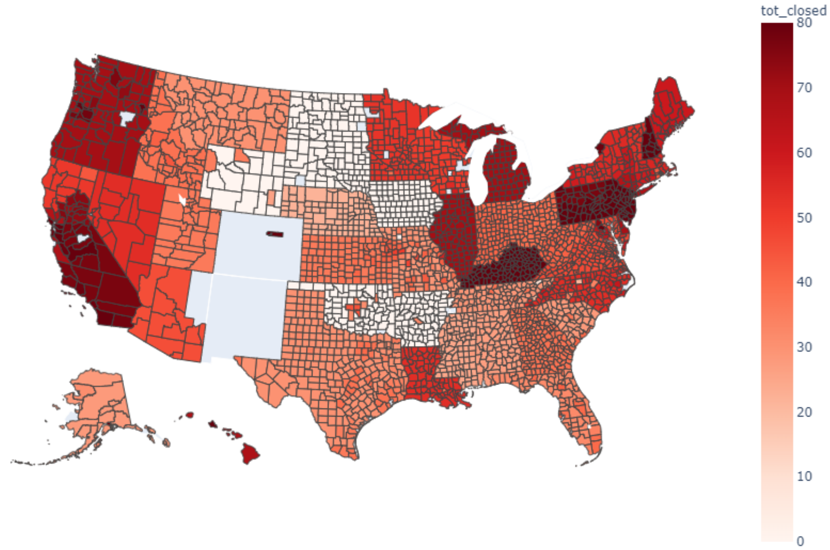
Figure 1.2: Stringency Index by state on June 1st, 2020.



As seen in Figure 1.2, the map is overall much lighter than in Figure 1 as the pandemic was overall much less severe in the country on June 1st compared to April 1st. Most parts of the country were experiencing fewer cases and deaths. However, some southern states were experiencing higher case counts at this time, labeled by some as the "Summer Surge." As seen in the figures, Georgia is one of the few states which is darker on June 1st than it was on April 1st. Overall, the two maps show considerable variation in stringency both across states and across time, which will allow studying the economic impacts.

Figure 1.3 below exhibits the variation across counties in government restrictions. As the stringency index is a state-level indicator, county level analysis instead uses the data from Spiegel and Tookes (2021) on the durations of stay-at-home orders.

Figure 1.3: Total number of days under active stay-at-home order by county. Darker counties had more stay-at-home order days. Grayed-out counties and states are grayed-out due to data issues.



As the figure above indicates, there is some county-to-county variation within states. For instance, very northern California counties largely had fewer days under a stay at home order than counties in the rest of the state. Metropolitan Oklahoma City counties enacted a stay-at-home order whereas the rest of the state did not.

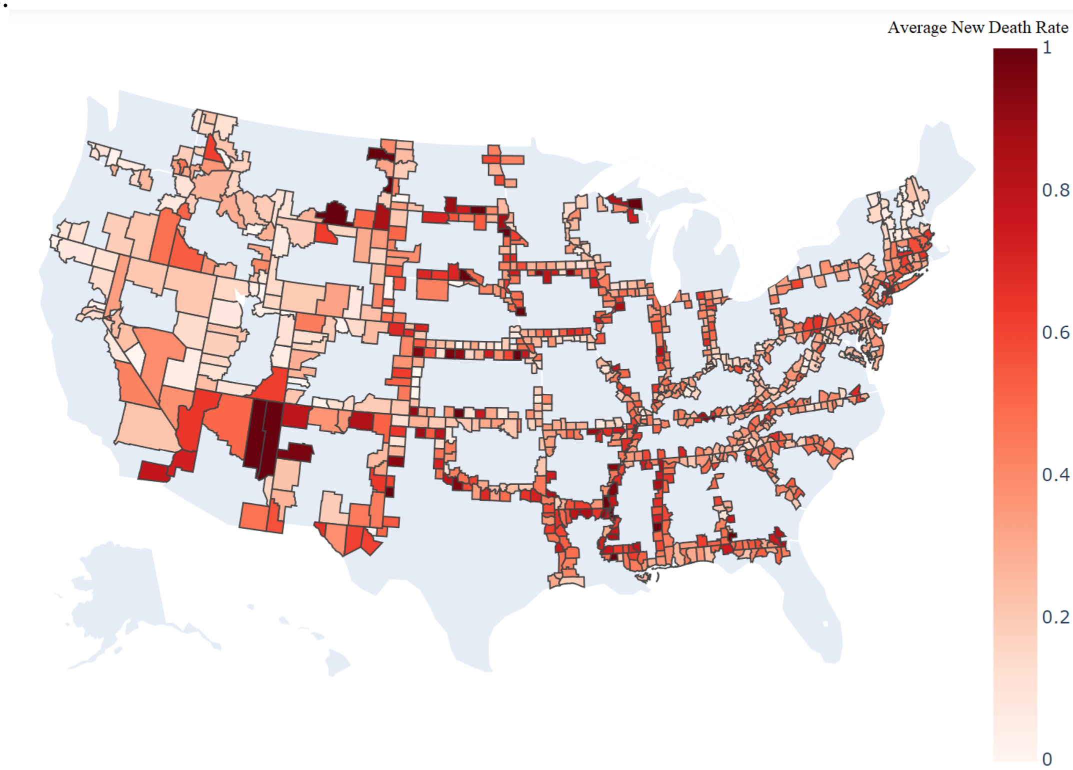
However, most of the variation between counties is found across state borders. While many counties did enact county-level orders, in many cases they were similarly timed to state-level orders. Some other counties enacted restrictions at the county level which were later rescinded as part of a statewide policy. For this reason, I conduct my main analysis on neighbor county-pairs that lie in different states.

1.3.2 Border-County Strategy

In order to study the pandemic and government restrictions on resilience, it is necessary to first address endogeneity concerns. Stay-at-home orders are enacted by politicians in conjunction with health departments in response to the pandemic. Unobserved factors may contribute to whether a government enacts an order or not and also impact businesses.

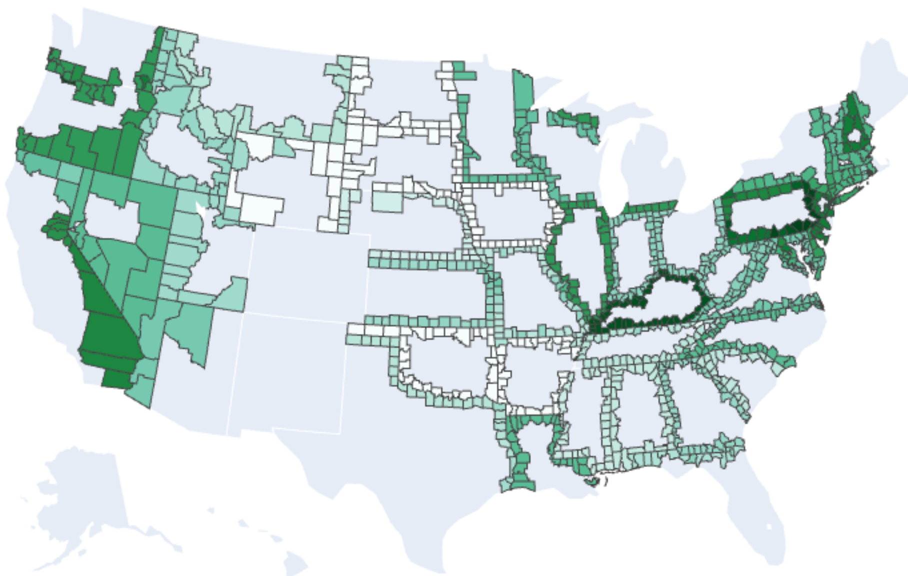
In order to mitigate these concerns, I use a border-county identification strategy. My identifying assumption is that neighboring counties tend to be relatively similar. In particular, the pandemic does not respect borders, as seen in Figure 1.4 below.

Figure 1.4: Average Covid-19 death rate by county in 2020. Darker counties experienced more deaths.



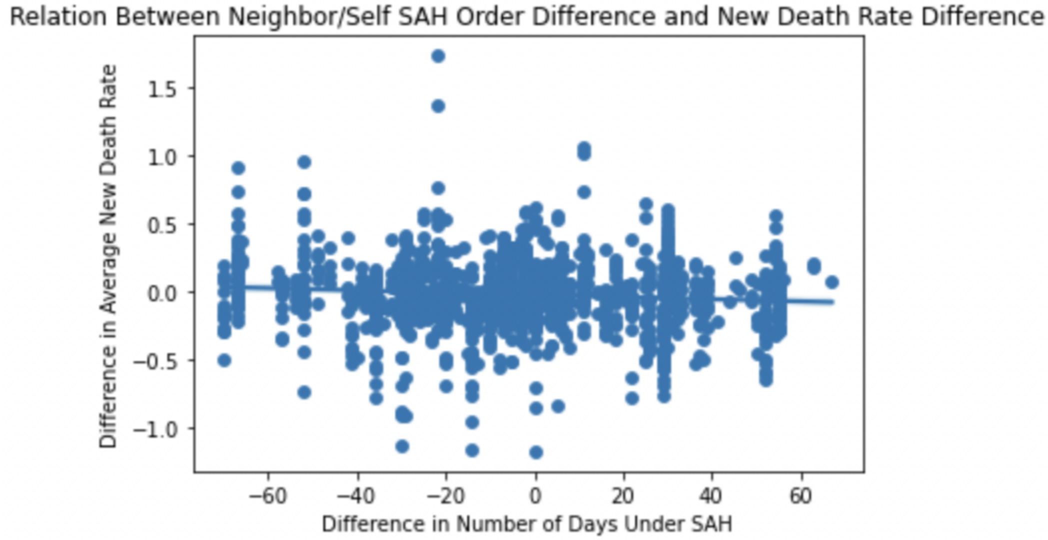
The above figure shows that Covid-19 deaths do not seem to be distributed with regard to state borders. Rather, deaths are geographically concentrated without respect for borders. A county next to another county with a high Covid-19 death rate is likely to have a high Covid-19 death rate itself, regardless of if its neighbor is in the same state or a different state. This is not the case with Covid-19 restrictions however, as seen in the next figure.

Figure 1.5: Number of days under stay-at-home order in the border-county sample.



Unlike the prior map, Figure 1.5 shows a clear discontinuity in stay-at-home order policies at state borders. Together the two maps reveal that the border-county sub-sample is a set of neighboring county-pairs where the two counties shared similar pandemic conditions, but faced differing stay-at-home order policies. This is shown using a scatter plot in Figure 1.6.

Figure 1.6: Difference in average 2020 Covid-19 death rate in neighboring counties versus difference in average 2020 days under a stay-at-home order.



The graph in Figure 1.6 reveals that there is not a strong relationship between the difference in average new death rate in neighboring counties and the difference in number of days under a stay-at-home order. The line is nearly flat, confirming that the set of county-pairs has significant variation in the two variables.

By running my empirical specifications using the difference between neighboring counties, I also lessen the impact of unobserved variables. This is because, on average, the difference in unobservable factors will also be relatively small in neighboring counties, thus reducing the effect they have on the results.

In order to further account for possible endogeneity concerns surrounding stay-at-home orders, I redo my main regressions using a sub-sample of county pairs where neither county is in the top five in their state when ranked by population, similar to Spiegel and Tookes (2021). This is because while stay-at-home orders were not random, they were usually determined at the state level and not controlled by individual counties. States likely considered the needs of their largest counties or population centers when deciding their policies. Smaller counties thus had to implement these orders whether or not their local conditions warranted them. As such, the imposition of stay-at-home orders on these counties can be more plausibly seen as exogenous.

1.3.3 Possible County-to-County Spillovers

Before discussing the methodology of this paper, I first acknowledge the possibility of county-to-county spillovers. Economic activity is not limited to intra-county activity. While people will likely primarily visit establishments in the county in which they reside, people also shop in other counties, especially those that are neighbors to their county of residence. If a county closes its businesses, there will likely be some people who will take their shopping to a neighboring county, rather than staying at home.

Consider two counties, A and B. Suppose that county A imposes a stay-at-home order and closes all nonessential businesses. In this project I study the impact of this on businesses in county A by looking at the change in difference in employment and the number of open businesses in the two counties. If people in county A increase shopping in county B as a response to the stay-at-home order, then economic indicators in county B will be boosted by these spillovers. This will increase the difference between the two counties and exaggerate the influence of the stay-at-home order on county A alone.

To minimize the impact of spillovers on the main results, I conduct the main analysis on a subset of county-pairs that lie in two different commute zones as defined in Autorn and Dorn (2013). My assumption here is that inter-county economic is reduced in when counties are not in the same commute zone, as they are likely to be less linked by roads and public transport. In the second chapter of this dissertation, I provide evidence for this assumption and study spillovers in greater detail.

After reducing the sample of counties to only those that neighbor a county in a different state that is not in the same commute zone, I end up with a reduced sample of 923 counties which I use to create pairs. Table 1.2 presents summary statistics.

Table 1.2: Mean values of main variables of interests. Left column is all 2989 counties in the original sample. Middle is reduced sample of only counties on a state border. Right column is further reduced to state border counties whose neighbor is in a different commute zone.

	Full Data	Border Counties	Border Counties with Diff. CZ Neighbor
N	2989	1105	923
Stringency Index	43.821	43.689	43.531
SAH	0.123	0.125	0.123
New Death Rate	0.333	0.328	0.333
Avg. DEM Vote Share	0.359	0.357	0.351
% NAICS 72	0.112	0.115	0.115
# Bank Branches (p.c.)	43.029	43.603	45.322
Population	105423.717	102557.760	94596.339

As seen in the table, the reduced sample of border counties with a different commute zone neighbor, is fairly representative of the 2989 counties in the whole sample. Importantly, the average stringency the government restrictions and Covid-19 levels are nearly identical in these counties to the full sample.

1.4 Empirical Specification and Main Results

My main specifications address the questions about the resilience of businesses in avoiding shutdown and layoffs. I use two different types of specifications to answer these questions. The first type of specification uses an event study that focuses only on neighboring county-pairs where one county implemented one stay-at-home order and its neighbor never issued one. This specification focuses on maximizing the identification by using a tight definition for treated vs. control counties, at the expense of sample size. The second type of specification is a more general difference-in-difference specification, which looks at a larger set of border counties and makes use of neighbor county-pairs where both may have implemented stay-at-home orders, but at different times or for different lengths.

1.4.1 Event Study Specification - Resilience to Shutdown

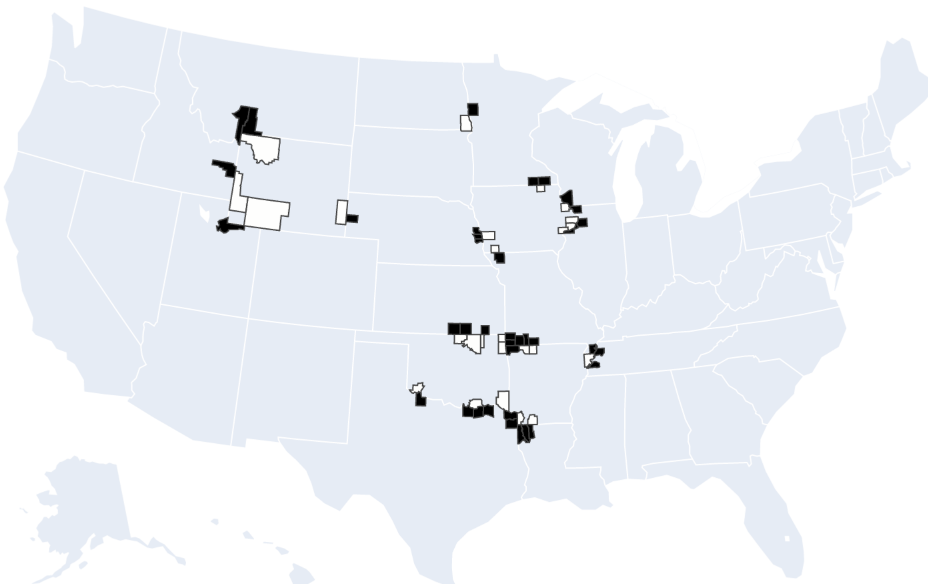
I first study business resilience by studying how Covid-19 and its associated restrictions affected the ability of business to stay open. For this specification I utilize the Womply variable which represents the change in number of open small businesses in a county each week relative to January 2020. As my observations are neighboring county-pairs, my dependent variable is the difference between the two counties in the percentage change in the number of open merchants from January 2020. More precisely, suppose the tuple (i, i_n) represents a county-pair where i is the "main" county and i_n is its neighbor. Define t to be the time index, which is weeks for this specification. The dependent variable is then defined as:

$$\Delta Merchants_{i,i_n,t} \equiv \frac{\#Merchants_{i,t}}{\#Merchants_{i,Jan2020}} - \frac{\#Merchants_{i_n,t}}{\#Merchants_{i_n,Jan2020}}$$

This definition means that coefficients from the regression defined below measure the percentage point additional change in the number of open merchants in the "main" county caused by an increase in the associated variable.

In all of the main results, I focus my attention to pairs neighboring counties that lie in different states and are also members of different commute zones. For this specification, I focus only on county-pairs where one county enacted a stay-at-home order only once, and its neighbor never issued a stay-at-home order. This approach is to clearly define the treatment as having a stay-at-home order issued, and not have confounding issues associated with further stay-at-home orders in some of the treated counties. As a consequence, the sample becomes 62 counties forming 43 county-pairs. In order to avoid double counting pairs as well as keep the "event" consistent for each observation, I keep only the counties where the "main" county is the one which implemented the stay-at-home order. As such, for each county-pair, the "event" being studied is the moment the main county in the pair implements its stay at home order. Figure 1.7 shows a map of this sample.

Figure 1.7: Counties included in event study regression on number of open merchants. Black counties issued a stay-at-home order in 2020 while white ones did not issue any.



I weigh the regression specification in order to not have results driven by smaller counties. For the merchants specification, I weigh county-pairs by their combined 2019 populations. The main weighted least squares event study specification for results on open small businesses is thus given by:

$$\Delta Merchants_{i,i_n,t} = \beta_{pre} Event_{pre,i,i_n,t} + \sum_{j=-5, j\neq -1}^{23} \left(\beta_j Event_{j,i,i_n,t}\right) + \beta_{post} Event_{post,i,i_n,t} + \gamma \mathbf{X}_{i,i_n,t} + \nu_{i,i_n} + \mu_t + \varepsilon_{i,i_n,t}$$

The variables $Event_{j,i,i_n,t}$ where $-5 \leq j \leq 23$ and $j \neq -1$ are indicators for the time period of the observation being j weeks after (or $-j$ weeks behind, when j is negative) the time period where county i implemented its stay-at-home order. For example, if county i implemented its stay-at-home order week 3, then $Event_{0,i,i_n,3} = 1$ and $Event_{-2,i,i_n,1} = 1$ as well. As is common in event studies, I drop the period $j = -1$ and treat it as the base period for comparison for each county-pair observation. $Event_{pre}$ and $Event_{post}$ are indicators for being more than 5 weeks before, or more than 23 weeks after, the week the stay-at-home order is implemented in county i.

X is a vector of controls as described in the data section earlier in the paper. One control worth highlighting is $\Delta NDR_{(i,i_n),t}$, which is defined as the difference in 7-day average daily death rates. ν_{i,i_n} and μ_t are county-pair and time fixed effects, respectively.

1.4.2 Event Study Specification - Resilience to Layoffs

I next study business resilience in terms of a businesses' ability to withstand economic conditions without resorting to layoffs. In this specification, I use monthly data on county level employment. In order to have a similar interpretation to the open merchants specification, I transform the employment data as a percentage change in employment from a baseline. Since the employment data is monthly, in order to have enough lags to verify the lack of a pre-trend, I include the last three months of 2019 data in the event study. I then use January 2020 as a baseline for employment. As such, the dependent variable for this specification is similarly defined as

$$\Delta Employment_{i,i_n,t} \equiv \frac{\#Employed_{i,t}}{\#Employed_{i,Jan2020}} - \frac{\#Employed_{i_n,t}}{\#Employed_{i_n,Jan2020}}$$

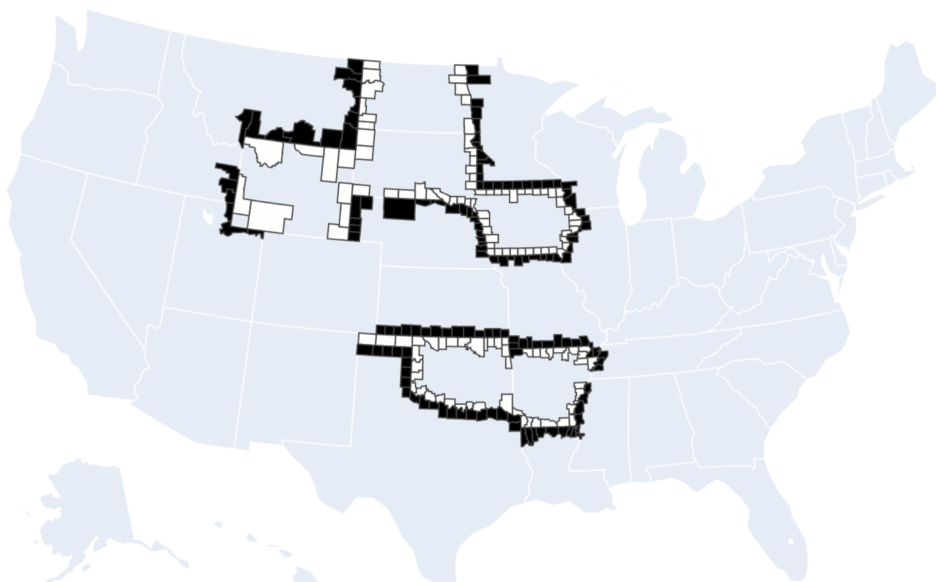
I weigh observations by combined January 2020 employment. The full weighted least squares specification is given by

$$\Delta Employment_{i,i_n,t} = \beta_{pre} Event_{pre,i,i_n,t} + \sum_{j=-4, j\neq -1}^{5} \left(\beta_j Event_{j,i,i_n,t}\right) + \beta_{post} Event_{post,i,i_n,t}$$
$$+ \gamma \mathbf{X}_{i,i_n,t} + \nu_{i,i_n} + \mu_t + \varepsilon_{i,i_n,t}$$

The variables $Event_{j,i,i_n,t}$ where $-4 \leq j \leq 5$ and $j \neq -1$ are similarly defined to the previous specification with the main difference being t denoting months rather than weeks. The period 1 month before the Event is dropped similar to the last specification as well. $Event_{pre}$ and $Event_{post}$ are indicators for being more than 4 months before, or more than 5 months after, the month the stay-at-home order is implemented in county i.

As in the previous specification, my sample consists of county-pairs where one county issued 1 stay-at-home order and its neighbor issued none. As the employment data are available for a larger set of counties than the Womply open merchants data, my resulting sample is a slightly larger 259 counties and 255 county-pairs. Figure 1.8 showcases this sample.

Figure 1.8: Counties included in event study regression on number of open merchants. Black counties issued a stay-at-home order in 2020 while white ones did not issue any.



1.4.3 Event Study Results - Resilience to Shutdown

Figure 1.9: Results from event-study regression on number of open merchants. The y-axis is the difference between county issuing the stay-at-home order and the neighboring county in the percentage change in number of small businesses in each county compared to January 2020. Time is indexed in weeks. Week 0 is the week the stay-at-home order imposing county issued the order. Weeks -6 to -2 are included to examine any pre-trend and Week -1 is dropped as the baseline. Bands show 90% and 95% confidence intervals.

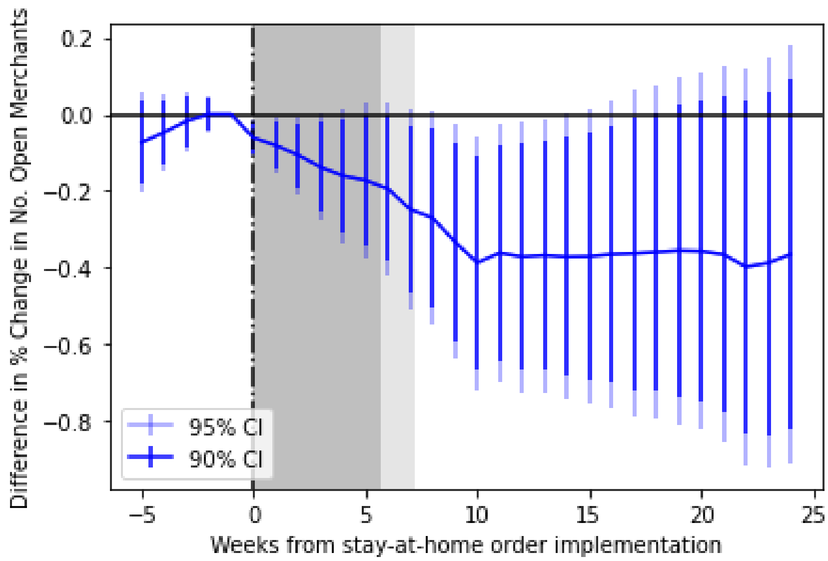
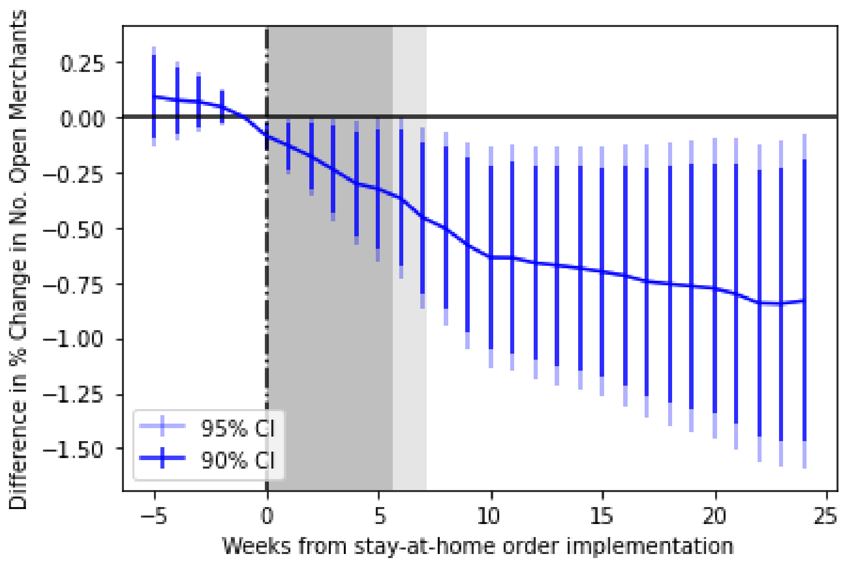


Figure 1.9 presents the results of the event study regression using the number of open merchants as the dependent variable. I present complete results that report all coefficients are available in tabular format in the appendix, table A.1. As standard in event studies, it is important to first examine if there are any pre-existing trends. All coefficients before week -1 are statistically insignificant from zero, confirming that there is not much evidence of a pre-existing trend.

Some impacts of the stay-at-order are felt immediately, as there is a statistically significant and negative coefficient beginning the week the stay-at-home order is issued. Recall that the dependent variable is the difference in percentage change of number of small businesses from January 2020 between the county issuing the stay-at-home order and its neighbor. As such, a negative effect is consistent with the idea that stay-at-home orders caused shutdowns of businesses. The presence of some immediate effects are unsurprising as the data include temporary closures that are to comply with the active stay-at-home order. Longer-term effects are more likely produced by economically-driven closures. Indeed, the estimated coefficient remains statistically significant and becomes increasingly negative as the weeks post the event go on. Weeks 8 through 10 appear to have the most significant negative effects, as the coefficient falls the fastest in this range and becomes increasingly significant. 10 weeks after the stay-at-home order was first enacted, the enacting county on average has additional 40-percentage-point loss from January 2020 in the number of open small businesses compared to its neighbor that did not enact a stay-at-home order compared to the week before the stay-at-home order took effect. After week 10, the effect seems to wane, as the point estimate remains remarkably flat after this point. Statistical significance wanes after week 10, and the impact of the stay-at-home order is no longer statistically significantly different from zero 24 weeks after the implementation.

Figure 1.10: Results from event-study regression on number of open merchants with the top 5 most populous counties in each state dropped. The y-axis is the difference between county issuing the stay-at-home order and the neighboring county in the percentage change in number of small businesses in each county compared to January 2020. Time is indexed in weeks. Week 0 is the week the stay-at-home order imposing county issued the order. Weeks -6 to -2 are included to examine any pre-trend and Week -1 is dropped as the baseline. Bands show 90% and 95% confidence intervals.



Results are quite similar when counties that are among the top 5 most populous in their state are decreased. I find a similar persistent negative effect of the impact of stay-at-home orders on the number of open small businesses in county. Results in this specification are also at a more negative point estimate and remain significant through the end of the horizon.

1.4.4 Event Study Results - Resilience to Layoffs

Figure 1.11: Results from event-study regression on employment. The y axis is the difference between county issuing the stay-at-home order and the neighboring county in the percentage change in employment in each county compared to January 2020. Time is indexed in weeks. Month 0 is the week the stay-at-home order imposing county issued the order. Months -6 to -2 are included to examine any pre-trend and Month -1 is dropped as the baseline. Bands show 90% and 95% confidence intervals.

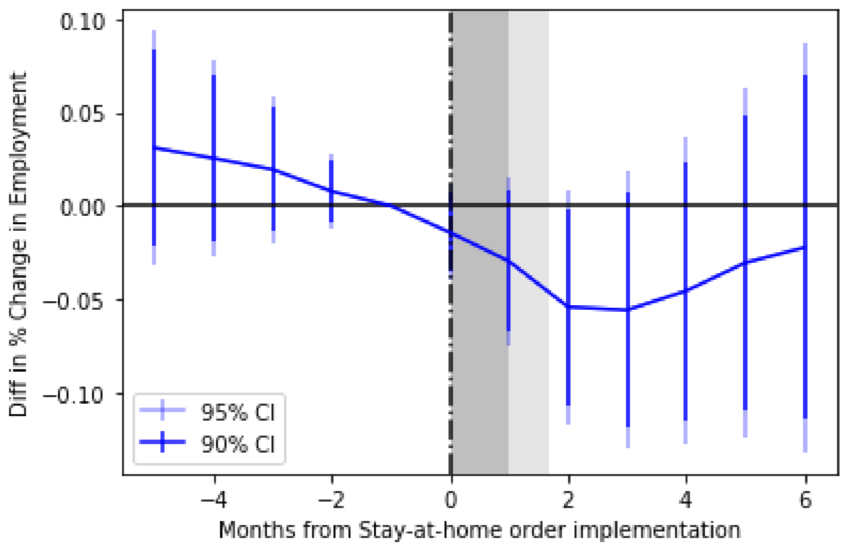


Figure 1.11 shows the results of the event study regression using difference in change in employment as the dependent variable. As in the merchants results, there is no evidence of a pre-existing trend before the implementation of the stay-at-home order as all coefficients for months -5 to -2 are not statistically significant.

Unlike the merchants result, the effects of the stay-at-home order are not immediately seen in the results on employment. This is not surprising as the employment data is not affected by temporary closures in the same way the data on open small businesses is. Both the coefficients on the month that the stay-at-home order is enacted and the first month after are negative but insignificant. The strongest effects from the stay-at-home order are seen in the second month after the event occurs, with the coefficient on month 2 negative and significant at the 90% level. On average, three months after the order is first enacted, counties that enacted a stay-at-home order see an additional 4.5 percentage point loss in employment since January 2020 compared to its neighbor with no stay-at-home order versus the month before the stay-at-home order went into effect. Effects seem to diminish after the three month mark. The point estimates remain negative, but start increasing after month three, and are no longer precisely estimated.

Figure 1.12: Results from event-study regression on employment with the top 5 most populous counties in each state dropped. The y axis is the difference between county issuing the stay-at-home order and the neighboring county in the percentage change in employment in each county compared to January 2020. Time is indexed in weeks. Month 0 is the week the stay-at-home order imposing county issued the order. Months -6 to -2 are included to examine any pre-existing trend and Month -1 is dropped as the baseline. Bands show 90% and 95% confidence intervals.

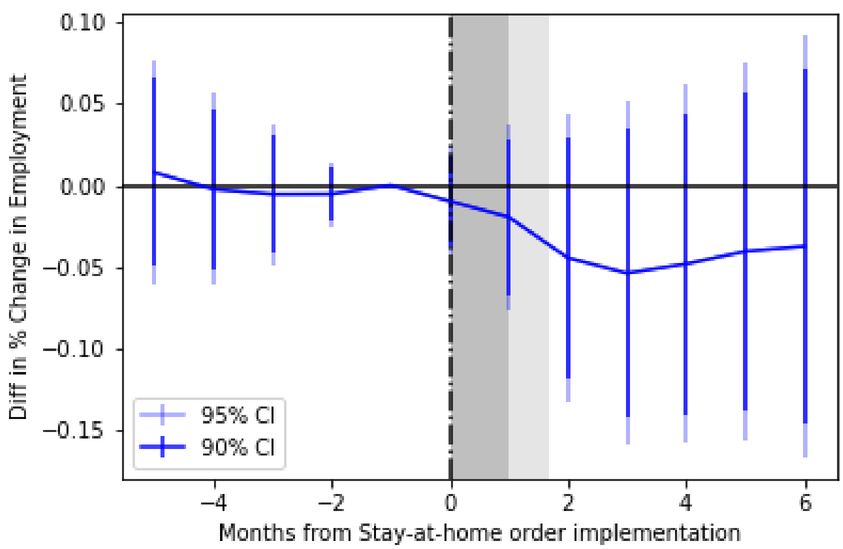


Figure 1.12 shows the results of a similar specification as in figure 11 but with the top 5 most populous counties in each state dropped. While statistical significance of the terms after the event is no longer present, we still see the same qualitative story as in figure 1.11. The evidence for a

lack of pre-existing trend is even stronger than in Figure 1.11 and, while statistical significance is not seen, there is a clear negative trend after the event which is similar to the previous results.

1.4.5 General Difference-in-Difference Specification - Resilience to Shutdown

I now introduce the specification used for my second set of main results. While the goal of the event study specifications was to define a very precise group of treated and non-treated counties, the goal of these difference-in-difference specifications is to take advantage of a larger group of counties and more types of variation in stay-at-home orders.

I weigh the regression specification in order to not have results driven by smaller counties. For the merchants specification, I weigh county-pairs by their combined 2019 populations. The main weighted least squares specification for results on open small businesses is thus given by:

$$\Delta Merchants_{i,i_n,t} = \sum_{k=-2}^{5} \beta_k \left(\Delta SAH_{i,i_n,t-2k}\right) + \gamma \mathbf{X}_{(i,i_n),t} + \nu_{i,i_n} + \mu_t + \varepsilon_{i,i_n,t}$$

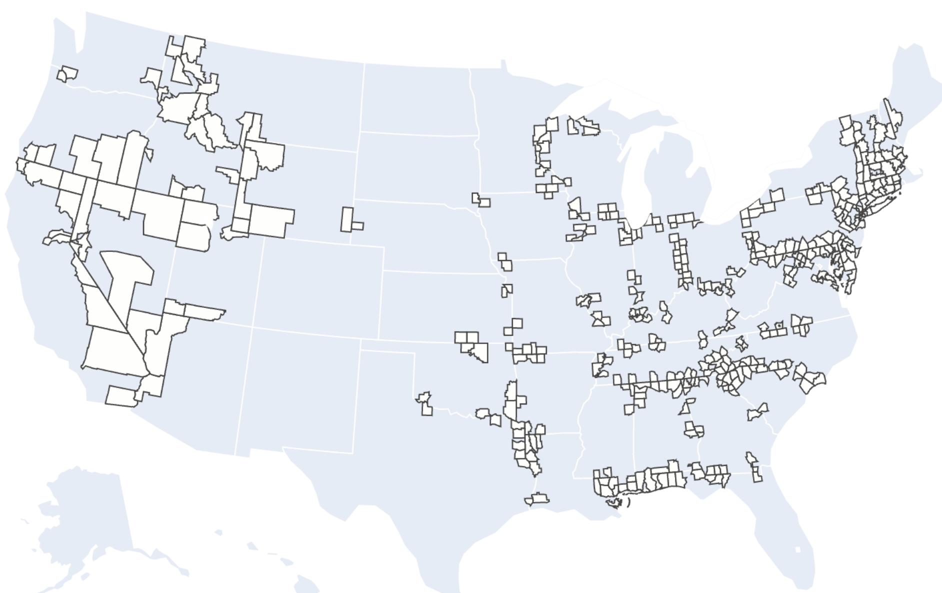
The variable $\Delta SAH_{i,i_n,t}$ is the difference in amount of time under a stay-at-home order between county i and i_n during week t. For example, if county i has a stay-at-home order during the entire week t and county i_n has no stay-at-home order at that time, then $\Delta SAH_{i,i_n,t} = 1$. ΔSAH may be non-integer values if there are weeks partially covered by stay-at-home orders.

As the impacts of the Covid-19 pandemic and its associated government restrictions are not likely to be contemporaneous, I include them in my regression as lags. Specifically, I include lags up to 10 weeks prior to the current date at 2 week intervals. As the data on stay-at-home orders and Covid-19 deaths are highly co-linear from week to week, I drop the odd numbered lags in order to keep the variation between lag terms. I also include two forward lags ($t+2$ and $t+4$) in order to verify that forward terms are not significant. Another reason for including several lags is to absorb the effect of stay-at-home orders causing temporary closures that are merely reflecting compliance with the stay-at-home order. As my dependent variable is simply the number of open small businesses, I am not able to directly distinguish shutdowns induced by economic conditions versus closures to comply with the government action. The nearer-term lags then importantly act to absorb the temporary short-term impact of the stay-at-home orders so that the longer-term lags are more likely to fully have significance driven by permanent, economic closures.

X is a vector of controls as described in the data section earlier in the paper. ν_{i,i_n} and μ_t represent county-pair and week fixed effects, respectively.

After excluding counties where data are missing, the final sample consists of 437 counties forming 361 unique county-pairs. Figure 1.13 below highlights the counties that make up this sample.

Figure 1.13: Counties included in the general difference-in-difference results on business resilience using number of open merchants as the dependent variable.



1.4.6 General Difference-in-Difference Specification - Resilience to Layoffs

I use lags of stay-at-home order and Covid-19 deaths similar to the previous specification. Since the employment data are monthly, I use 1 and 2 month lags of each. I use a weighted least squares specification using the total employment in the two counties in each pair as the weight for each observation. The specification is given by:

$$\Delta Employment_{i,i_n,t} = \sum_{k=-2}^{2} \beta_k \left(\Delta SAH_{i,i_n,t-k}\right) + \gamma \mathbf{X}_{(i,i_n),t} + \nu_{i,i_n} + \mu_t + \varepsilon_{i,i_n,t}$$

The time index t now represents months in this specification. The independent variables are defined similarly to the previous specification adjusted to the month level.

1.4.7 Difference-in-Difference Results - Resilience to Shutdown

Table 1.3: Results from difference-in-difference regression on number of open merchants. Variables with a Δ are differences between a county and its neighboring county. Merchants is the percentage change in number of open small businesses at time t in a county versus January 2020. Observations are weighted by combined county-pair 2019 population.

	$\Delta Merchants_t$	
	(1)	(2)
ΔSAH_t	-0.0395***	-0.0353***
	(0.0103)	(0.0089)
ΔSAH_{t-2}	-0.0183***	-0.0106*
	(0.0059)	(0.0058)
ΔSAH_{t-4}	-0.0112**	-0.0115**
	(0.0053)	(0.0053)
ΔSAH_{t-6}	-0.0009	0.0021
	(0.0046)	(0.0054)
ΔSAH_{t-8}	-0.0151***	-0.0113***
	(0.0043)	(0.0041)
ΔSAH_{t-10}	-0.0053	-0.0042
	(0.0038)	(0.0068)
$\Delta NewDeathRate_t$	-0.0067**	0.0016
	(0.0030)	(0.0035)
R-squared	0.7309	0.7095
R-squared Adj.	0.7229	0.7006
Observations	13461	9120
County-Pair FE	Yes	Yes
Week FE	Yes	Yes
Top 5 Dropped	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

It is important to first discuss the different interpretation of the results in this specification. Whereas the event-study shows the cumulative effect on the dependent variable over time, the coefficients in this regression show the independent effects of differences in explanatory variables at individual time periods. Additionally, as mentioned in the section discussing the specification, there will be a confounding effect caused by temporary shutdowns that are merely to comply with the mandate. This problem is apparent in this result on the coefficient on the 0-week lag of ΔSAH. While it is highly statistically significant and negative, it is likely heavily driven by temporary closures caused by the stay-at-home order, rather than actual business failures. As such, I focus my attention to longer run effects, specifically the coefficient on the 8-week lag of ΔSAH. The estimated coefficient on this term is statistically significant and negative, similar to findings in the event study specification. On average, a county that has an active stay-at-home order during a full week with a neighbor that does not impose one will have an additional 1.51 percentage points loss in the number of open small businesses relative to January 2020 compared to its neighbor after 8 weeks. Results are similar when reducing the sample to only counties that are not ranked in the top 5 of their state by population. I similarly find a highly significant and negative coefficient of -0.0113 on the 8-week lag of ΔSAH. The nearer term lags of ΔSAH remain significant as well in column 2, but at reduced significance.

Additionally, the contemporaneous coefficient of ΔNDR is significant and negative in column (1), though not column (2). The significance however suggests that even after accounting for impacts of the stay-at-home orders, I cannot fully reject the counter-factual that the pandemic has its own direct negative effects on the economy. This gives some credence to the idea that simply avoiding these non-pharmaceutical interventions would not be sufficient to avoid economic damages to small businesses. Indeed, if the interventions are significantly reducing deaths, undoing them may cause additional harm, rather than improve economic conditions.

1.4.8 Difference-in-Difference Results - Resilience to Layoffs

Table 1.4: Results from difference-in-difference regression on employment. Variables with a Δ are differences between a county and its neighboring county. Merchants is the percentage change in employment at time t in a county versus January 2020. Observations are weighted by combined county-pair January 2020 employment.

	$\Delta Employment_t$	
	(1)	(2)
ΔSAH_t	-0.0053	-0.0131**
	(0.0040)	(0.0057)
ΔSAH_{t-1}	-0.0119***	-0.0163***
	(0.0029)	(0.0035)
ΔSAH_{t-2}	0.0015	-0.0089**
	(0.0044)	(0.0041)
$\Delta NewDeathRate_t$	0.0008	0.0008
	(0.0016)	(0.0012)
$\Delta NewDeathRate_{t-1}$	-0.0015	-0.0016
	(0.0011)	(0.0012)
$\Delta NewDeathRate_{t-2}$	-0.0035**	-0.0040**
	(0.0014)	(0.0016)
R-squared	0.5986	0.5634
R-squared Adj.	0.5612	0.5225
Observations	10152	8388
County-Pair FE	Yes	Yes
Month FE	Yes	Yes
Top 5 Dropped	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 1.4 contains the results of the regression using the difference in difference specification and difference in change in employment as the dependent variable. As mentioned in the specification section, this specification varies slightly from the previous, with the time unit being months rather than weeks.

I estimate a statistically significant and negative impact of the stay-at-home orders on employment, similar to the result using the event study specification. On average, a county implementing a stay-at-home order for the duration of a month with a neighbor that does not have any stay-at-home order that month will see an additional 1.19 percentage point loss in employment relative to January 2020 in the following month. It is worth noting here that the significance in the employment specification is on the one month lag term of ΔSAH, whereas it was on the 8 week lag term in the merchants specification. This suggests that on average, businesses' resilience to layoffs was closer to 1 month. This result makes sense intuitively, as layoffs are a less severe outcome than going out of business. Additionally, the coefficient remains negative with a more negative value of -0.0163 in the specification without the most populous counties in each state.

I also estimate a statistically significant and negative coefficient on the two-month lag term of ΔNDR, showing that there are separate effects from the pandemic itself similar to the results on the number of open merchants. Unlike those, this negative effect on employment is present in both columns of table 1.4, further giving credibility to the hypothesis that the pandemic has its own direct effects.

1.5 Conclusion

The Covid-19 pandemic provided a stark reminder of the importance of business resilience. Though many people are already aware of the importance of personal resilience, such as in the way of emergency funds for things like rent, the concept is just as important to businesses. Many firms throughout the country found themselves unready for a prolonged shutdown brought on in March 2020.

In this paper, I exploit the Covid-19 pandemic and use a border-county identification strategy to study the resilience of firms in the United States. I find that, on average, most layoffs occurred about one month after the implementation of stay-at-home orders and that most business failures occurred 8-10 weeks after. Intuitively it makes sense that firms resort to laying off employees first before they decide to shut down as layoffs are a significantly less severe outcome. This analysis also answers questions about the degree to which stay-at-home orders, versus the pandemic itself, impacted firms and workers throughout the United States. I find that both the pandemic and stay-at-home orders individually contribute to the losses in employment and businesses seen in the country throughout 2020. Ultimately, the results presented here agree with the viewpoint that simply "reopening the economy" by revoking stay-at-home orders may not be sufficient to undo the economic damages as some stem from the pandemic itself, regardless of economic restrictions. It is important for policy makers to consider results here alongside other economic literature and public health studies when making considerations on how to implement non-pharmaceutical interventions.

There are many policy implications from these results, however it should be noted that the complete analysis indicating how the government should proceed in its implementation of stay-at-home orders and other NPIs will also require analysis on impacts on death rates such as in Spiegel and Tookes (2021). While I find that there is a significant harm to businesses caused by stay-at-home orders, careful consideration should be paid to their health benefits as well. Indeed, the results in the regressions in this paper alone imply that if health benefits to stay-at-home orders are large, then they may not have much of a negative economic impact. What is clear from the results here is that economic aid to businesses should be increased and targeted more, such as to places where there are more stringent policies or higher concentrations of restaurants and food services. As discussed in Hubbard and Strain (2020), more comprehensive revenue replacement programs may be appropriate to greater assist firms where non-payroll expenses are more significant. There is some indication in the results in this paper that the effects on business survival were larger than employment losses in magnitude, suggesting that greater assistance in non-payroll expenses could have been beneficial.

I chose to focus this paper on the study of stay-at-home orders as they were the most stringent and controversial type of non-pharmaceutical intervention seen in the United States. Future work in the literature should examine how results vary when studying the economic impacts of other types of orders. Mask mandates in particular are thought to be less economically intrusive by many and still effective in reducing the severity of an outbreak. I also study the impact of access to financing via the number of bank branches in a county variable. It would also be interesting if future work examined other measures of access to financing, such as the amount of economic impact emergency loans approved in a county. It would also be valuable to look at other measures of business health as measures of business resilience, such as cash holdings.

Chapter 2: Covid-19 Stay-At-Home Orders and Economic Spillovers

2.1 Introduction

Policies enacted by regional governments often cause spillovers onto their surrounding areas. These spillovers are especially likely when neighboring areas are highly connected. These spillover effects can have significant consequences for the effectiveness of the policy created. In some cases, co-operation between neighboring governments can help account for spillover effects caused by polices and lead to a more optimal policy prescription throughout the area. Recently, this thought process was prevalent among many neighboring regional governments during the Covid-19 pandemic.

The Covid-19 virus was first detected in China in early 2020. Effects were initially confined to the Chinese province of Wuhan, but soon proliferated throughout the rest of the world. By March of 2020, there were large numbers of Covid-19 cases in the United States. The pandemic lasted for over two years and, through multiple waves, caused serious negative health consequences throughout the United States and the rest of the world. At the time of writing, the disease had claimed just over 1 million deaths in the United States alone and over 6 million deaths worldwide. The damages caused by the Covid-19 pandemic were not limited to health outcomes, as economic consequences were far reaching as well. Unemployment in the United States skyrocketed in the initial wave of the pandemic, rising from 3.5 percent in February to 14.7 percent in April. Many businesses, especially those that were unable to transition effectively to remote work, were unable to survive the disruptions. Treatment options were extremely limited in the early stages of the pandemic with vaccines and anti-viral treatments not available until several months later. In order to combat the pandemic, regional governments were forced to rely on non-pharmaceutical interventions (NPIs). These NPIs often included restrictions on movement and business activity. Some of the most restrictive NPIs used by many states in the U.S. were "Stay-at-Home Orders," which closed all non-essential businesses from operating and banned many types of gatherings.

The implementation of NPIs was often associated with the economic hardships faced by people throughout the world in this time, and made the the implementation of them extremely controversial. Many people argued that the "cure" should not be worse than the disease. In the United States, views on stay-at-home orders often became aligned with political views. Republican-leaning individuals often tended to oppose further pandemic-related restrictions while democratic-leaning individuals tended to favor them. The political polarization in the country and economic hardships faced by many individuals and businesses in this time made many of these NPIs quite controversial. Some argued that public health orders must take into account their impact on the economy and that the "cure" should not be made worse than the disease. Others attributed the economic impact as instead directly stemming from the disease itself and that the economy could not be saved without focusing on reducing the outbreak first.

The differences in opinion between policy makers and the United State's system of Federalism made a unified national policy on NPIs impossible to implement. Government-enacted policies were mostly implemented at state and local levels, and co-operation only found in limited circumstances. One example of co-operation in this time period was between the governments of Connecticut, New Jersey and New York which implemented policies in-step with each other since all three governments were highly connected via the New York City metropolitan area. Many feared that the effectiveness of stay-at-home orders in reducing negative health outcomes caused by Covid-19 would be severely hampered without regional co-operation in health policy. If a county implements a stay-at-home order but neighboring counties do not, then there remains a possibility that residents of that county would disobey the stay-at-home order and travel to the open businesses in the neighboring counties. This would then greatly reduce the effectiveness of the stay-at-home orders and cause the number of lives saved to be reduced.

In this paper, I use a border-county strategy and difference-in-difference frameworks to estimate the causal impact of stay-at-home orders on inter-county movement. I use data on travel between neighboring counties as a proxy for spillovers of economic activity. I find that, unsurprisingly, stay-at-home orders caused a reduction of movement from counties without a stay-at-home order to their neighbors with a stay-at-home order. This is completely expected, as there is no incentive for people to travel to counties under a stay-at-home order. More surprisingly, I also find that stay-at-home orders caused a reduction of movement from the counties with a stay-at-home order to neighboring counties without one. This result goes against the hypothesis of stay-at-home order related spillovers. I also test the possibility that a net spillover effect remains by examining the effect of stay-at-home orders on the ratio of travel between the two counties. In this last result, I find no evidence of an impact of stay-at-home orders on the ratio of travel in both between counties with a stay-at-home order and their neighbors without one.

There are two strands of literature which this paper contributes to. First, this paper adds to the literature on the impact of government restrictions on both economic and health outcomes. Particularly relevant to this paper is Spiegel and Tookes (2021), which examines the impact of various Covid-19 government restrictions on deaths at a county level and contributes a novel data set which has detailed information on county level restrictions throughout the United States. Papers such as Amuedo-Dorantes (2020), Alexander and Karger (2021) and Caselli et. al (2020) examine other impacts of the stay-at-home orders and other NPIs. I contribute to this literature by establishing impacts of stay-at-home orders on movement between counties, which has strong implications for their efficacy.

Second, this paper contributes to the larger body of work on economic spillovers. Spillovers have been studied in many contexts, such as the impact of gun control legislation as in Bronars and Lott Jr. (1998). Economic spillovers are studied in Bernstein et. al. (2019), which looks at the negative economic impacts of bankruptcies onto neighboring establishments. Economic spillovers created by government actions are examined in papers like Chalermpong (2005) which examines negative spillovers created by a the construction of a highway. Holtz et. al. (2020) and Elenev et. al. (2021) in particular also examine economic spillovers caused by stay-at-home orders. I contribute to the literature examining this by further examining the role of commute zones in spillovers as well as studying directional effects on the ratio of travel in both directions.

The rest of this paper is organized as follows. Section 2 goes through the data used in this project in detail. Section 3 examines the variation in state responses to the pandemic and explains the identification strategy used. Section 4 introduces the specifications and the main results. Section 5 concludes.

2.2 Data

I use several different sources of data in this project which I outline below.

2.2.1 Movement and Commute Zone Data

In order to construct a proxy for county-to-county economic spillovers, I use a detailed geolocation dataset from Safegraph. The data contain information on movement in counties throughout the United States. The raw data have establishment level information on the total number of visits and detailed information on the visitors each week. In particular, it contains information on the home census block group of each visitor. I filter out any establishments that had under five visits in a week, as Safegraph adds noise to establishments with less than five visitors in order to protect privacy. I then transform the establishment-level data to convert data on the home census block of visitors into data on their home county.

I aggregate the establishment-level data by county. This results in a dataset that has the number of visitors across all establishments in each county, grouped by the home county of the visitors. I then match neighboring counties to create county-pairs. Next, I extract the specific number of visitors that come from the neighboring county into the "main" county and vice-versa in order to create "self-to-neighbor" and "neighbor-to-self" visitor fields. I then drop county pairs where the "neighbor" county has a higher FIPS code than the "main" county to drop redundant county pairs. To account for population differences, I convert these measure to per-capita measures for use as the proxy for economic spillovers in the main analysis

Lastly, I incorporate data from Autor and Dorn (2013) in order to do a check on an assumption used in chapter one of this dissertation. In chapter one, I account for the possibility that stay-at-home order-induced spillovers bias my results by restricting my sample to a set of neighboring county-pairs that lie in two different commute zones. I assume that spillover effects would be reduced in counties that are less connected to each other. I use this data on commute zones to map counties to commute zones and then, for each county-pair, I create an indicator variable which is equal to one if they are assigned classified as belonging to two different commute zones, and zero otherwise. This allows me to include this in my specification later in the paper in order to test the crucial assumption used in the first chapter.

2.2.2 Government Restrictions Data

As in chapter one of this dissertation, I use two distinct measures of non-pharmaceutical interventions issued by regional governments. First, I use the stringency index constructed by researchers at Oxford's Government Response Tracker. The stringency index is an ordinal measure of the strictness of a regional government's restrictions during each day of the pandemic. This includes measures such as stay at home orders, school closures, restaurant capacity constraints, as well as others. I primarily use the stringency index as motivation for my identification strategy, as it effectively showcases the variation in state policies throughout the United States. Table 2.1 exhibits summary statistics of this measure throughout the time period of study.

Table 2.1: Stringency Index summary statistics from March 2020 to December 2020. The left column shows variation of the stringency index over time across states. The right column shows variation across states in the average stringency index in this time period

	Stringency Index	Stringency Index State Average
N	17085	51
Mean	45.099	45.099
Std. Dev.	21.917	11.190
Min	0	0
25th percentile	35.19	39.993
Median	47.69	45.281
75th percentile	61.11	51.533
Max	87.96	66.452

The second is data collected on the start and end dates of various non-pharmaceutical interventions implemented in the United States. This incredibly detailed dataset is contributed by Spiegel and Tookes (2021). This data contains information on when various non-pharmaceutical interventions were in effect at the county level in the United States. As in chapter one, I focus on the usage of stay-at-home orders as they were the most severe intervention used by American regional governments and were the most controversial.

2.2.3 Neighbor Counties

As in chapter one, I use the government provided county adjacency file which maps each county in the United States to a list of neighboring counties. I then match counties based on this data to create a list of all neighboring county-pairs. As in the first chapter, I remove some county-pairs which are officially neighbors but do not share a land-border.

2.2.4 Other Controls

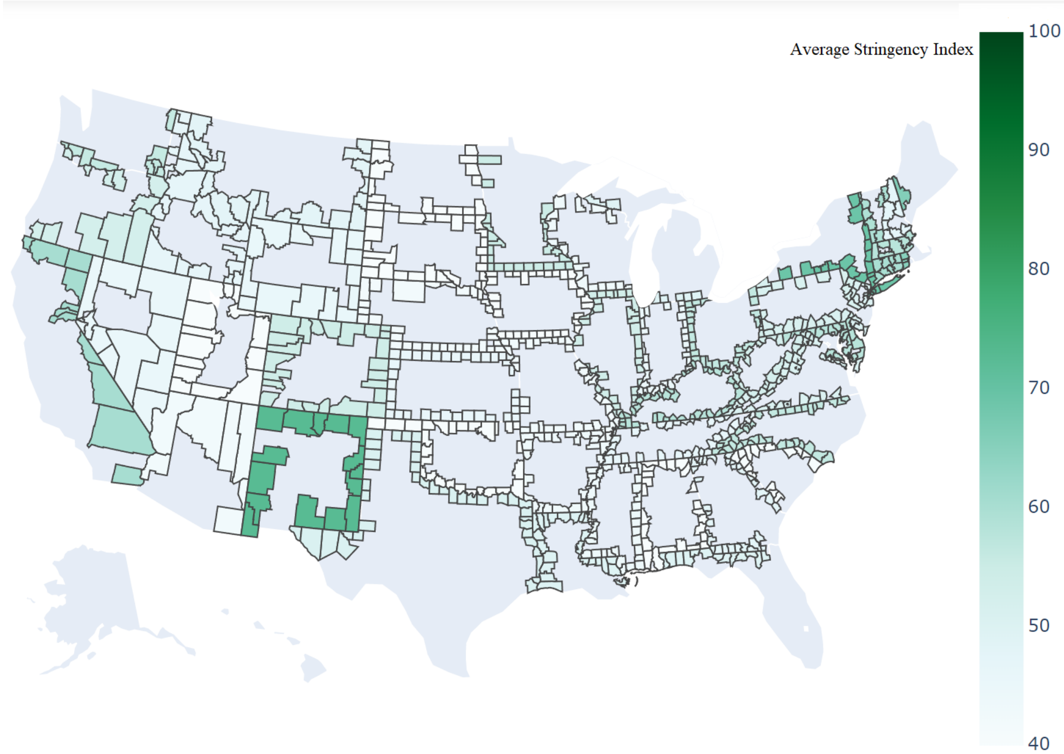
I use daily data on COVID-19 7-day-average death rates from the Opportunity Insights project that is also mentioned above. Deaths are used instead of cases as they were less under-reported than cases in the early stages of the pandemic and are thus a more accurate representation of the underlying pandemic. I include them in the specification to account for time varying differences in the level of the underlying pandemic between neighboring counties, which may also have an effect on inter-county movement and spillovers.

2.3 Identification Strategy

2.3.1 Variation in Government Actions

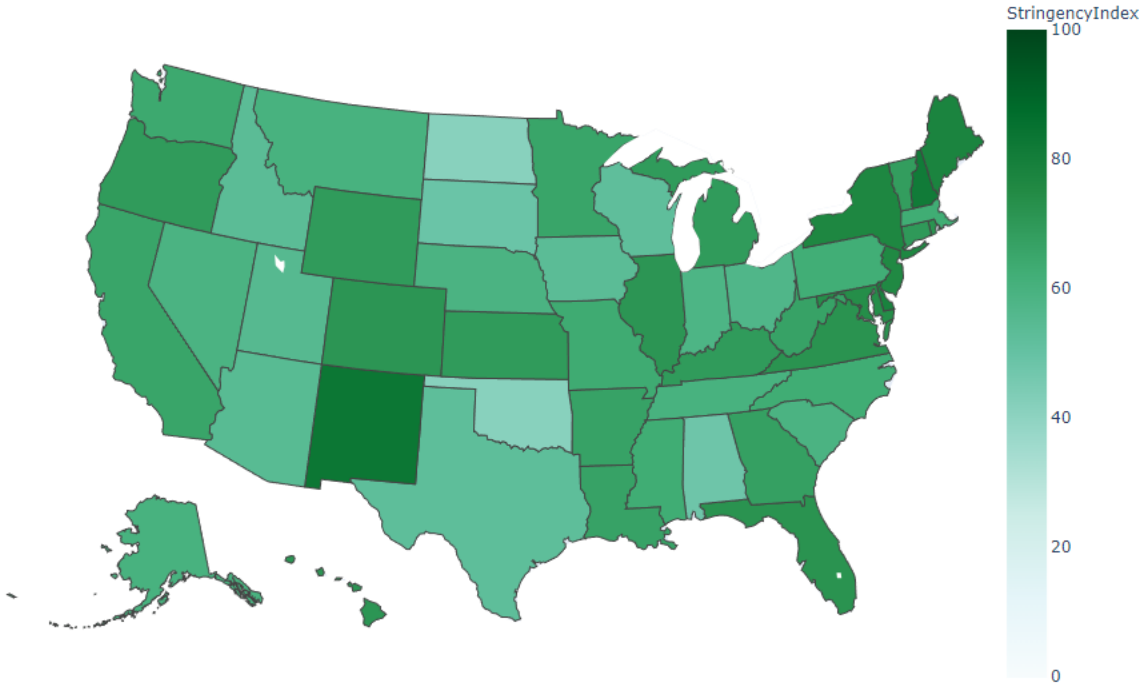
In order to identify the causal impact of stay-at-home orders on movement between neighboring counties, I rely on variation between government policies across neighboring counties over time. This strategy is based on the identification strategy used in the first chapter of this dissertation. During the first year of the pandemic, the federal government issued very few of its own restrictions, and those that it did issue were primarily focused on restricting foreign travel to the United States. Most restrictions enacted domestically were done by state and local governments. Figure 2.1 showcases the variation in government restrictions across states during the first wave of the pandemic in late March and early April of 2020.

Figure 2.1: Average Stringency Index by county in 2020. Darker is more stringent.



Deaths during the first wave of the pandemic were heavily concentrated in the Northeastern and Midwestern United States. As such, it is unsurprising to see more severe restrictions in those regions of the country in the month of April 2020. Next, consider how the situation evolved two months later in June 2020.

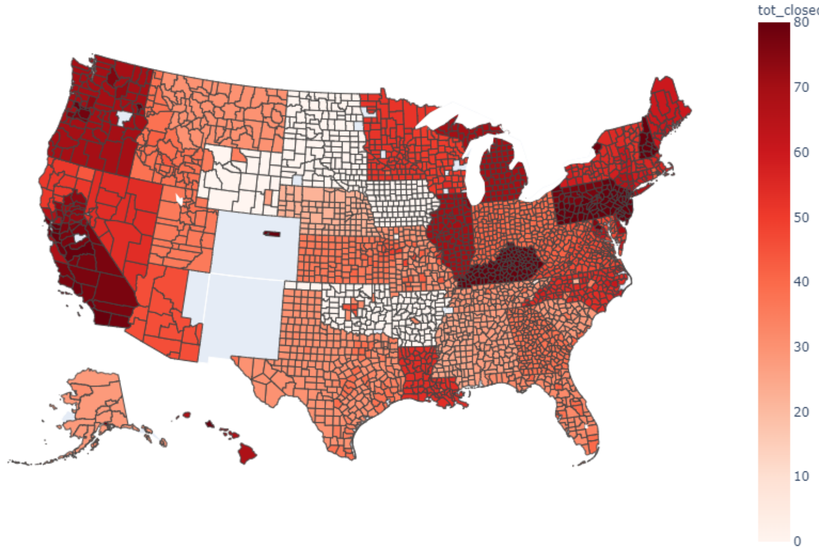
Figure 2.2: Stringency Index by state on June 1st, 2020.



For most of the country, the pandemic was less severe in June. As seen in Figure 2.2, the map is overall much lighter than in Figure X as state governments lifted some restrictions in conjunction with the improved situation. While most places, including the original epicenter of New York, were doing better in this time period, some southern states were experiencing higher case counts at this time. This was labeled by some as the "Summer Surge." As an example, Georgia is a case where the average stringency level in the state is actually higher in June of 2020 than it is in April.

Next, I provide similar evidence of the variation in government policies using the county-level data on stay-at-home order dates from Spiegel and Tookes (2021). Figure 2.3 shows the variation across counties in the total number of days under a stay-at-home order in 2020.

Figure 2.3: Total number of days under active stay-at-home order by county. Darker counties had more stay-at-home order days. Grayed-out counties and states are grayed-out due to data issues.



There is some within state-variation in stay-at-home order policy. One stark example is in Oklahoma, where most of the state had no stay-at-home order enacted for the full year. However, the counties in the metropolitan Oklahoma City area were under stay-at-home orders for part of 2020. This is one example of several cases where local governments enacted their own restrictions when state-level policies were considered insufficient. In other cases, restrictions were initially implemented statewide and then rolled back on a case-by-case basis. For example, New York implemented its "New York on pause" restrictions statewide in late March 2020. When the state government started rolling these restrictions back, it initially only did so in parts of upstate New York, with the New York City being the last to fully rescind its restrictions.

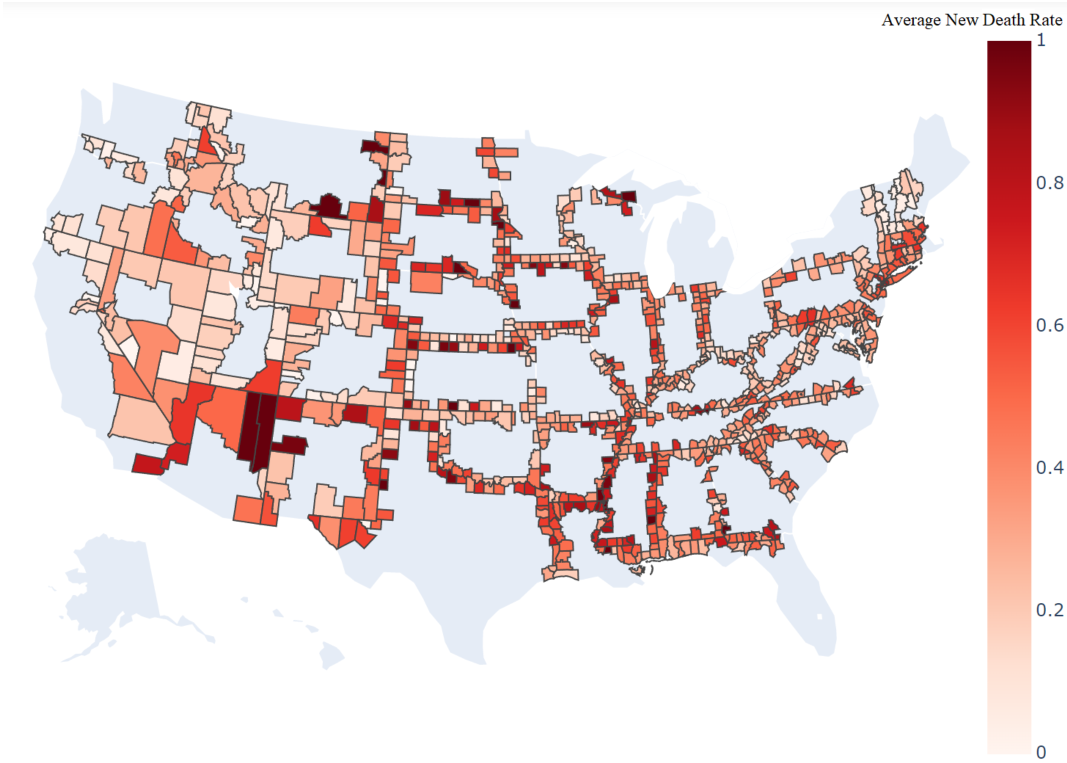
While the map shows the variation within states, it clearly demonstrates that the vast majority of the variation in stay-at-home orders is found across states. For this reason, I focus my main analysis on neighbor county-pairs which lie across a state border.

2.3.2 Border County Strategy

Identifying the causal impact of stay-at-home orders on county-to-county spillovers requires exogenous variation in stay-at-home order policies. However, stay-at-home orders were specifically enacted in order to combat the underlying Covid-19 pandemic, which would cause concerns about endogeneity. In order to create pseudo-random variation in stay-at-home orders and identify their causal effect, I use a border-county identification strategy.

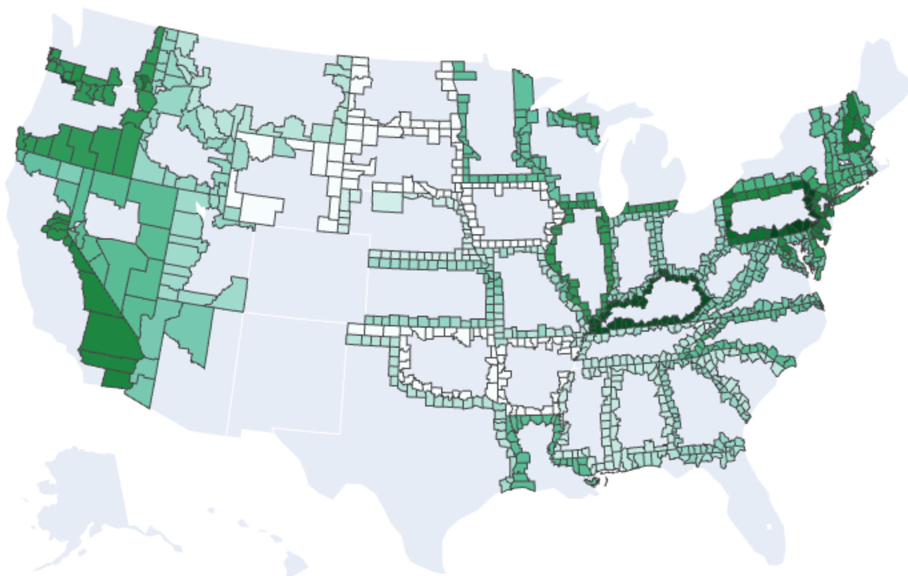
My identifying assumption is that neighboring counties tend to be relatively similar. With this assumption, differences in stay-at-home order policy across neighboring counties can be treated as pseudo-random. To show this similarity across neighboring counties, I focus on the primary potential source of endogeneity: the underlying pandemic. The Covid-19 pandemic does not respect borders. As such, a large number of cases in one county is likely to result in a large number of cases in its neighbors as well. This is true regardless of whether the neighbors are in the same state or not. Consider the following map:

Figure 2.4: Average Covid-19 death rate by county in 2020. Darker counties experienced more deaths.



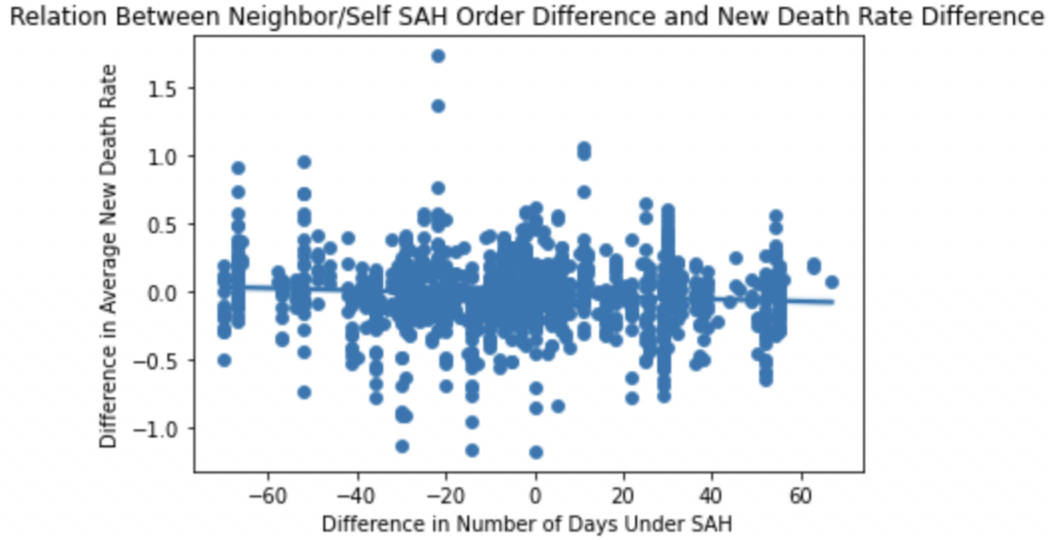
As seen in Figure 2.4, Covid-19 deaths levels in counties are similar to their neighboring counties. Counties with large numbers of deaths are likely to be located next to other counties with large numbers of deaths. Importantly, deaths are geographically concentrated without respect for state borders. Stay-at-home orders policies at the county level, however, are highly determined by the state the counties are in, as seen in the next figure.

Figure 2.5: Number of days under stay-at-home order in the border-county sample.



Unlike the previous map, Figure 2.5 shows a clear discontinuity in stay-at-home order policies at state borders. Taking the two maps in conjunction shows that the sample of neighboring counties in different states is a sample of counties that had experienced similar levels of the pandemic, but did not necessarily have identical responses to it. This is seen more precisely in Figure 2.6.

Figure 2.6: Difference in average 2020 Covid-19 death rate in neighboring counties versus difference in average 2020 days under a stay-at-home order.



As seen in the graph, there is not a strong relationship between the difference in average new death rate in neighboring counties and the difference in days under a stay-at-home order. This provides the argument to treat the variation in stay-at-home orders across neighbors in this sample as pseudo-random.

Additionally, by running my empirical specifications using the difference between neighboring counties, I also reduce the impact of other unobserved factors since, on average, the difference in other unobservables will also be relatively small in neighboring counties.

Lastly, in order to further strengthen the argument that the variation in stay-at-home orders may be treated as pseudo-random, I re-estimate my main regressions using a further reduced sample of county pairs where neither county is in the top five most populous in their states. This strategy follows that of Spiegel and Tookes (2021). This is done to exploit the fact that most stay-at-home order policies were created at the state level. States likely considered the needs of their largest counties or population centers when deciding their policies. Smaller counties thus had to implement these orders whether or not their local conditions warranted them. In this sample, it is even more likely that differences between neighbors in stay-at-home order policy is not driven by differences in local conditions. Rather, the difference is driven by one of the counties in the pair happening to share a state government with a larger county that has different pandemic conditions. As such, the imposition of stay-at-home orders on these counties can be more plausibly seen as exogenous and allows for identifying of their impact on spillovers.

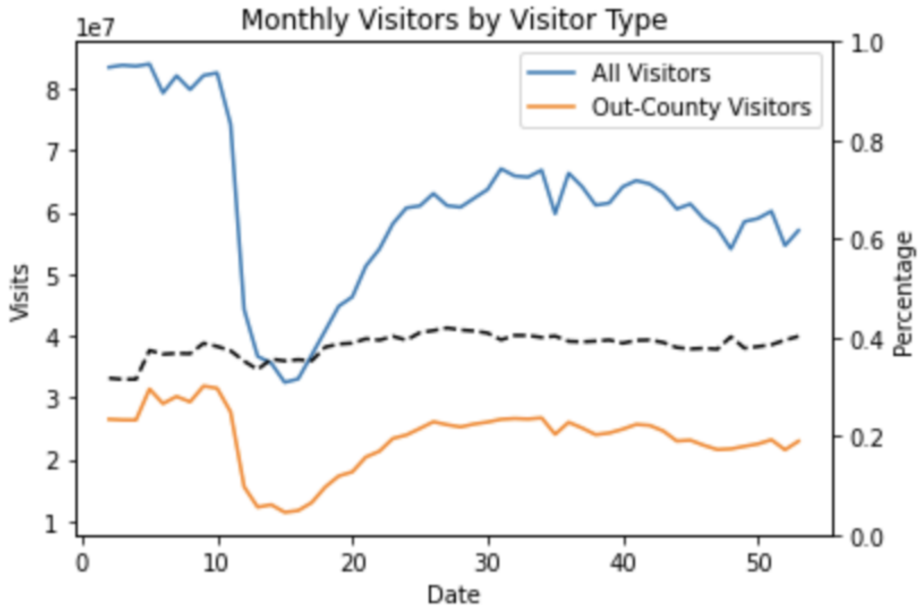
2.4 Main Specifications and Results

As discussed briefly earlier in the paper, one concern with the main results is that they may be driven by spillovers. Increased travel to neighboring cities caused by stay-at-home orders may cause overestimating of the negative effects of stay-at-home orders. In order to mitigate this issue, the main results discussed so far were all done on county-pairs in which the two counties were part of different commute zones. The assumption is that counties are less connected to neighboring counties if they are not in the same commute zone. This would imply that there is less travel between the two counties, which means less people shopping in neighboring counties and potentially less spillovers associated with stay-at-home orders. In this section, I provide further evidence for this claim and also study county-to-county spillovers due to the pandemic in more detail.

2.4.1 Movement Between Counties

In order to examine how spillovers may be impacting results, I first look at movement data between counties. I use Safegraph provided data that contains information on visitors to establishments in each county and information on the home census block groups of visitors. I first examine trends in movement and movement between counties by looking at data on weekly visits to establishments in 2020.

Figure 2.7: Number of visitors by visitor type in all border counties in the United States by week in 2020. The dotted line is the average percentage of visitors that come from outside the county.

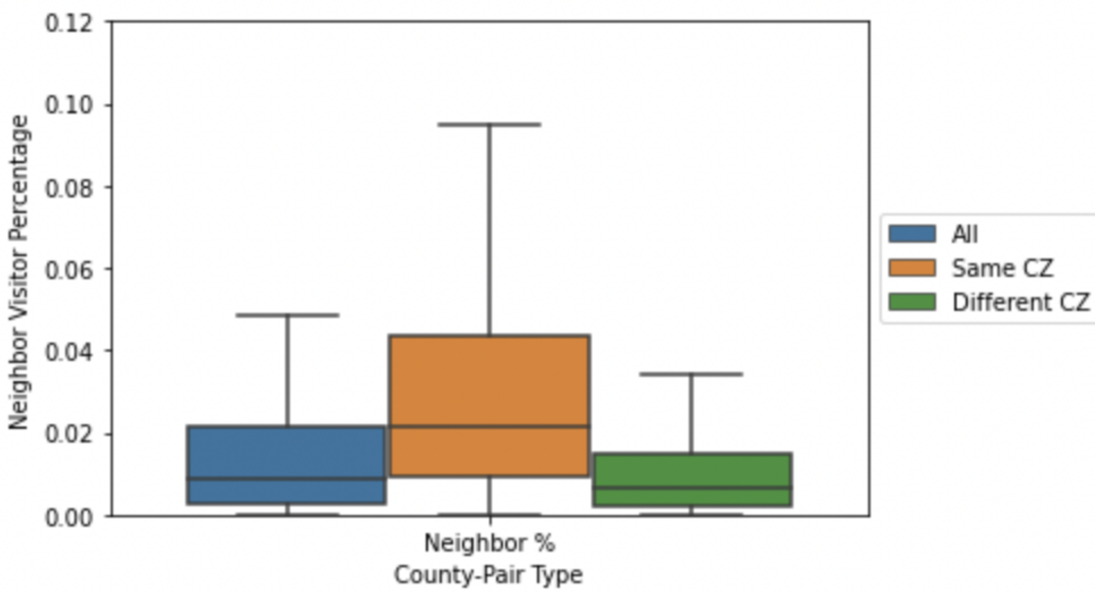


As Figure 2.7 indicates, the number of visitors to establishments sharply dropped around 10 weeks into 2020, coinciding with the first wave of the Covid-19 pandemic. This decrease can be seen in both intra- and inter-county visitors to establishments. The black dotted line which represents the percentage of visitors that come from outside the county drops slightly at this time as well, but recovers by week 20 to pre-pandemic levels. Weeks 10 through 20[1] of the year included the heaviest use of stay-at-home orders throughout the United States during the pandemic. Though future waves had higher measured case counts, governments often avoided re-enacting harsher NPIs due to their unpopularity. While mask mandates and other less stringent NPIs were reintroduced in many counties, the majority of the country did not implement additional stay-at-home orders. Despite the stringent measures in place during the early wave of the pandemic, figure 2.7 clearly shows that there was no uptick in the fraction of visits that are inter-county to establishments during this time.

[1] Approximately March 8th through May 16th

Next, I examine the differences in travel between counties in the same commute zone versus those in different commute zones. In this step, I focus on look at travel between pairs of counties rather than more general inter-county movement. As before, this analysis is on county-pairs that straddle a state border. Figure 2.8 shows the distribution of the percentage of visitors in establishments in a county that come from its neighbor county across all the county-pairs in the sample.

Figure 2.8: Distribution of percentage of visitors to establishments in the main county in a county-pair that come from the neighbor county. The blue box plot shows the distribution across all county-pairs in the sample. The orange and green box plots show the breakdown when decomposing the sample into county-pairs that lie in the same commute zone and those that lie in different commute zones, respectively.



As the figure shows, the percentage of visitors that come from the neighboring county in a county-pair varies considerably depending on whether or not the two counties share a commute zone. Across all county-pairs, the average percentage of visitors that come from the neighbor is roughly 2.1 percent. For county-pairs lying in the same commute zone, this number is around 4.0 percent, whereas for different commute zone pairs, it is 1.4 percent. The box-plots show that the distribution as a whole for same commute zone pairs is quite different, as the median county-pair here has a higher neighbor visitor percentage than 75th percentile county-pair in the different commute zone set. Figure A.1 in the appendix shows the breakdown in inter-county movement in more detail for the restaurant industry.

From this it is clear that there is strong evidence for my assumption that counties will experience fewer spillovers from neighbors that lie in different commute zones than they will from neighbors in the same commute zone. Next, I show that the spillovers caused by the stay-at-home orders in particular are diminished.

2.4.2 Estimating Spillovers Specification

I estimate a similar specification to the difference-in-difference specifications in the first chapter of this dissertation. First, I consider the impacts of stay-at-home orders on movement between counties. Unlike the specifications in chapter one, there are two distinct directions of movement per county-pair that could be impacted by policies and the pandemic: movement from the main county to its neighbor, and movement in the opposite direction. I define two dependent variables for the first set of regressions. As before, county i represents the "main" county and i_n its neighbor in the county-pair. The first variable is visitors traveling from the neighbor county to the main county, which is defined by:

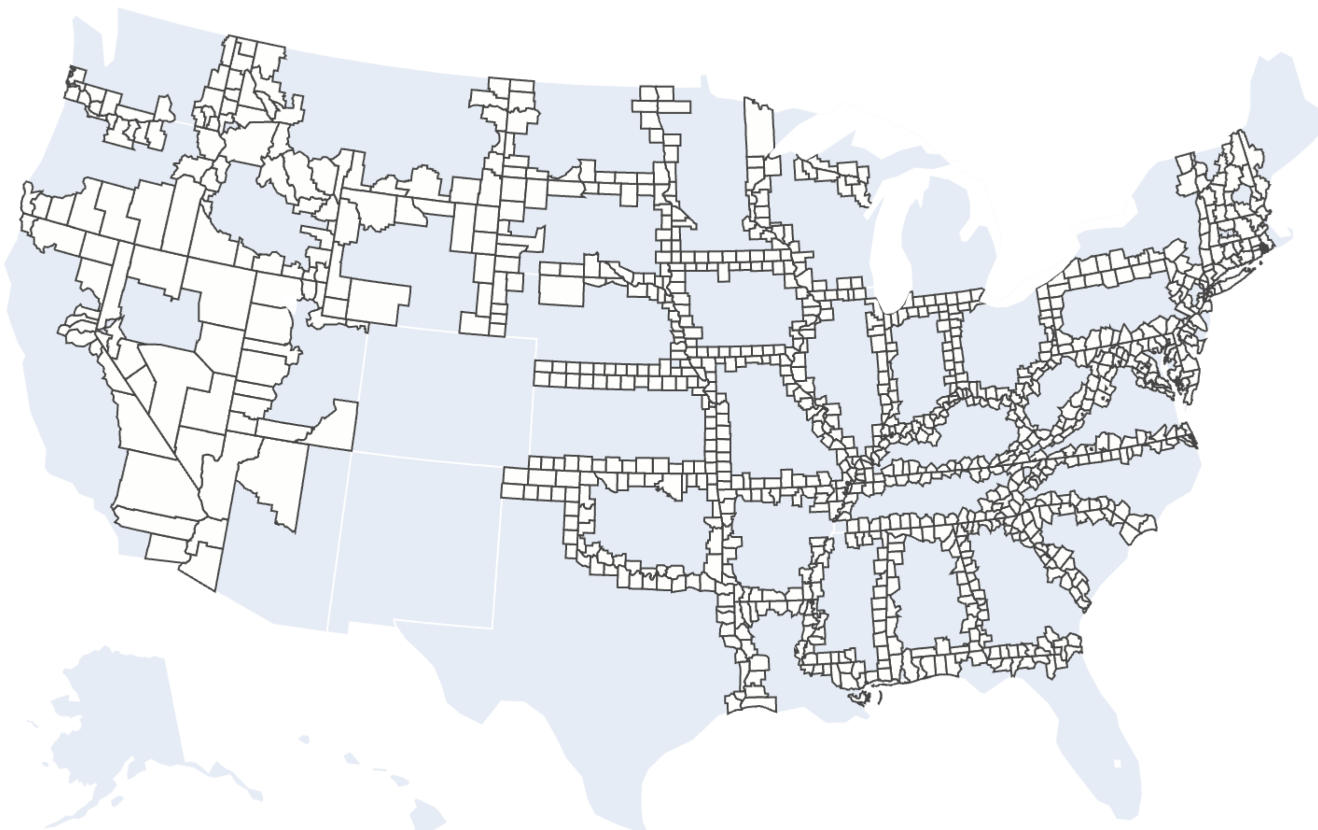
$$Visitors_{i_n \to i,t} \equiv \frac{\#Visitors_{i_n \to i,t}}{Population_{i_n}}$$

Similarly, I define per-capita visitors traveling in the opposite direction by the following equation:

$$Visitors_{i \to i_n,t} \equiv \frac{\#Visitors_{i \to i_n,t}}{Population_{i}}$$

Part of the purpose in studying spillover effects is to examine the differences between pairs that share a commute zone versus pairs that do not. As such, unlike in the first chapter of this dissertation, I do not remove county-pairs that share a commute zone in the specification in this chapter. As such, my sample of counties for this specification consists of 1071 counties and 1160 unique county-pairs. The sample is illustrated in figure 2.9.

Figure 2.9: Counties included in results on impact of Covid-19 stay-at-home orders on inter-county movement.



Stay-at-home orders may not affect movement between counties in strictly linear ways. Residents of a county may not adjust their travel in the same magnitude when moving from a county with no active policy to one with an active stay-at-home order as they would when moving in the opposite direction. To allow for this, I estimate following specification:

$$
\begin{aligned}
Visitors_{i_n \to i,t} = {} & \beta_1 Rel.Closed_{i,i_n,t} + \beta_2 Rel.Open_{i,i_n,t} \\
& + \beta_3 \left(Rel.Closed_{i,i_n,t} \times DCZ_{i,i_n}\right) + \beta_4 \left(Rel.Open_{i,i_n,t} \times DCZ_{i,i_n}\right) \\
& + \gamma \mathbf{X}_{(i,i_n),t} + \nu_{i,i_n} + \mu_t + \varepsilon_{i,i_n,t}
\end{aligned}
$$

where

$$RelClosed \equiv \mathbf{1}\left(\Delta SAH_{i,i_n,t} > 0\right) \text{ and } RelOpen \equiv \mathbf{1}\left(\Delta SAH_{i,i_n,t} < 0\right)$$

and

$$DCZ_{i,i_n} \equiv \mathbf{1}\left(i \text{ and } i_n \text{ have different commute zones}\right)$$

RelClosed is an indicator variable that is equal to 1 when the main county in a pair, i, is under a stay-at-home order for a larger portion of week t than its neighbor i_n. Hence, i is *relatively closed* compared to i_n. Likewise, if i_n is more stringent in this time period, ΔSAH will be negative and *RelOpen* will be equal to 1 since i is *relatively open* compared to i_n. *DCZ* is an indicator for if counties i and i_n are in the same commute zone. The specification using travel in the opposite direction, i.e. visitors from the main county to its neighbor, is identical on the right hand side.

In addition to directly estimating spillover effects caused by stay-at-home orders, the goal of this section of the paper is also to evaluate the hypothesis used in the main economic results. That is, establishing if spillover effects caused by stay-at-home orders are in fact reduced in county pairs where the two counties lie in separate commute zones. The coefficients of interest here are thereby those on the interactions between *RelClosed* and *DCZ* and between *RelOpen* and *DCZ*, which are β_3 and β_4 above. Where β_1 and β_2 measure direct spillovers, β_3 and β_4 measure the marginal impact of counties in the pair lying in two separate commute zones on these spillovers. As such, if the hypothesis that these spillovers are mitigated is correct, then β_3 should have the opposite sign of β_1 and β_4 should have the opposite sign of β_2.

2.4.3 Results on Spillovers

Table 2.2: Results from difference-in-difference regression on inter-county movement within county-pairs. Observations are weighted by combined county-pair 2019 population. The left two columns use movement from the neighbor county to main county as the dependent variable whereas the right two columns look at the opposite direction. Each dependent variable is used with both the linear specification using stay-at-home order and the non-linear one using indicators for relatively closed or open.

	Neighbor County to Main County Visitors$_t$	
	(1)	(2)
$Rel.Closed_t$	-626.3188***	-407.1668***
	(180.9939)	(131.4806)
$Rel.Open_t$	-408.0482***	-221.2564**
	(154.5871)	(98.5817)
$Rel.Closed_t \times DCZ$	592.5045**	478.5878***
	(236.9369)	(146.4631)
$Rel.Open_t \times DCZ$	438.7342**	230.5418**
	(170.4588)	(114.4471)
R-squared	0.9468	0.9553
R-squared Adj.	0.9457	0.9543
Observations	60318	48722
County-Pair FE	Yes	Yes
Week FE	Yes	Yes
Top 5 Dropped	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2.2 presents the results from the specification using per-capita visitors from the neighboring county to the main county in each pair. Unlike the results from the first chapter, effects of the stay-at-home order are decomposed into two distinct coefficients. Somewhat surprisingly, I find negative and highly statistically significant estimated coefficients on both $Rel.Closed$ and $Rel.Open$ in the regression. I find that, on average, a county that is relatively closed compared to its neighbor will receive 626 fewer weekly visitors per-capita from its neighbor. Additionally, on average, a county that is relatively open compared to its neighbor will receive 408 fewer visitors per-capita from its neighbor. In order to give some context to these numbers, note that the average the number of visitors from a neighbor county to the main county during a week in this sample is about 1600 visitors per-capita. This suggests that these results represents between a one fourth and one third reduction of travel in both directions on average. While the negative estimated coefficient on $Rel.Closed$ is unsurprising and consistent with spillovers from the closed county onto the open one, the estimated coefficient on $Rel.Open$ is not consistent with this theory of spillovers. The negative coefficient on $Rel.Open$ implies that there are negative spillovers from the relatively more closed county onto the less open county as well. If Covid-19 restrictions caused people to shop in neighboring open counties instead, we would expect the coefficient on $Rel.Open$ to be positive. Instead, I find spillovers are negative in both directions. Results are similar in column (2) which is the same specification but with the 5 most populous counties in each state removed from the sample.

Next, I evaluate the crucial assumption from the first chapter that these spillovers are mitigated in county pairs where the two counties lie in different commute zones. I find positive and significant estimated coefficients on both interaction terms between *DCZ* and the two indicators for the difference in stay-at-home order. As the spillovers in both directions are negative, positive coefficients on the interaction terms imply that spillovers caused by the stay-at-home orders are indeed mitigated by the two counties in a pair not being in the same commute zone.

Table 2.3: Results from difference-in-difference regression on inter-county movement within county-pairs. Observations are weighted by combined county-pair 2019 population. The left two columns use movement from the neighbor county to main county as the dependent variable whereas the right two columns look at the opposite direction. Each dependent variable is used with both the linear specification using stay-at-home order and the non-linear one using indicators for relatively closed or open.

	Main County to Neighbor County Visitors$_t$	
	(1)	(2)
$Rel.Closed_t$	-170.3127***	-53.8137
	(49.5822)	(43.4909)
$Rel.Open_t$	-268.9482***	-291.2850***
	(63.8029)	(101.7833)
$Rel.Closed_t \times DCZ$	254.0445***	137.5031***
	(56.7189)	(47.5937)
$Rel.Open_t \times DCZ$	283.2866***	223.7066**
	(68.3146)	(107.4454)
R-squared	0.9536	0.9632
R-squared Adj.	0.9526	0.9624
Observations	60318	48722
County-Pair FE	Yes	Yes
Week FE	Yes	Yes
Top 5 Dropped	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

In table 2.3, I show results for the same specification as the results in table 2.2, but changing the direction of movement in the dependent variable. Here the results paint a similar story. I again estimate that there are negative spillovers in both directions of travel. That is travel from the relatively closed county to the relatively open as well as the relatively open county to the relatively closed county falls. I again find highly significant positive estimated coefficients on the interactions between *DCZ* and the difference in stay-at-home order indicator variables in this specification. This provides further evidence that stay-at-home order-induced economic spillovers are mitigated when neighboring counties are in different commute zones.

The results from tables 2.2 and 2.3 clearly demonstrate that spillovers from stay-at-home orders in the United States are negative in both directions. Residents in both counties reduce their movement to their neighboring county regardless of if their county is the relatively closed one or the relatively open one. This runs against the idea that consumers in the county under a stay-at-home order would then increase their shopping in a neighboring county. The possibility remains, however, that even though travel is reduced in both directions, it reduces asymmetrically and that stay-at-home orders may increase the *relative* amount of travelers in one direction. To examine this, I introduce a new dependent variable, the ratio of travel in both directions, which is given by

$$VisitorRatio_{i,i_n,t} \equiv \frac{Visitors_{i_n \to i,t}}{Visitors_{i \to i_n,t}}$$

I then re-estimate the specifications in tables 2.2 and 2.3 using this dependent variable. Results from this regression are found in table 2.4.

Table 2.4: Results from difference-in-difference regression on ratio of inter-county movement within county-pairs. Observations are weighted by combined 2019 county-pair population. $N \rightarrow S$ and $S \rightarrow N$ represent per-capita visitors from the neighbor county to the main county and vice-versa, respectively.

	Visitor Ratio$_t$	
	(1)	(2)
$Rel.Closed_t$	1.9601	-0.7979
	(1.8989)	(0.8034)
$Rel.Open_t$	-2.1420	1.0629
	(1.7743)	(1.3435)
$Rel.Closed_t \times DCZ$	-3.3133*	0.0812
	(1.9739)	(1.0098)
$Rel.Open_t \times DCZ$	-0.2042	-0.9648
	(1.8784)	(1.6484)
R-squared	0.6345	0.6562
R-squared Adj.	0.6265	0.6486
Observations	56778	45370
County-Pair FE	Yes	Yes
Week FE	Yes	Yes
Top 5 Dropped	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

I do not find statistically significant estimated coefficients on either $Rel.Closed$ or $Rel.Open$. This provides evidence that a difference in stay-at-home order policy between the main county and its neighbor does not affect the composition of travel between the two counties. The previous results show that gross spillovers are negative in both directions as people decrease inter-county travel. The results in table 2.4 go further and find no evidence of a net or directional spillover between the two counties. The lack of significance on the coefficients on $Rel.Closed$ and $Rel.Open$ show that there is no evidence that travel in one direction dropped more sharply than the other. As there is no evidence of a directional spillover, it makes sense that there is also minimal evidence of the importance of commute zones in this specification. While there is a statistically significant estimated coefficient on $Rel.Closed \times DCZ$ in column (1), this is only significant at the 10% level and is not present in column (2) where the top five most populous counties are dropped.

2.5 Conclusion

The implications of government policies are often not confined to the area under governance by that government. Many state and locals government were aware of this fact during the early stages of the Covid-19 pandemic and attempted to cooperate with their neighbors on stay-at-home order policy. Many feared that the limited partnerships in this real would lead to failure to effectively limit spillovers and the spread of Covid-19. In this paper, I use a border-county strategy and detailed geo-location data to estimate the spillovers caused stay-at-home orders in the United States to help answer this question.

I show that Covid-19 stay-at-home order related spillovers in movement are negative in both directions of travel. That is, travel reduces both from the county with a stay-at-home order to its neighbor without one, and from the neighbor to the county with a stay-at-home order. The lack of a positive impact of stay-at-home orders on movement to neighboring counties goes against the theory of spillovers. Additionally, I further show that there is no evidence of a net or directional spillover when analyzing the effect of stay-at-home orders on the ratio of travel in the two directions. These results together provide strong evidence that predictions that stay-at-home orders may be ineffective without regional cooperation were not realized. The results indicate that even when neighbors do not issue their own stay-at-home orders, residents of a county that does issue them largely remain at home. As such, it is imperative that in future situations where such restrictions may be necessary, such extreme events caused by climate change or political unrest, that governments enact necessary restrictions quickly and do not delay them in order to wait for cooperation from neighbor governments first.

This paper focuses on stay-at-home order related spillovers as stay-at-home orders were the most comprehensive and controversial non-pharmaceutical intervention implemented within the United States. Future work in this area should also look at spillovers related to other policies. Additionally, it would be valuable if other papers were to examine spillovers in a more granular way by looking at movement within certain industries. As more data become available, it would also be interesting to see this question studied using data on consumption, rather than just movement.

Chapter 3: Let's Go Home: Evidence for Career Concerns in Major League Baseball Umpires

3.1 Introduction

A career concerns model can help explain why agents may reduce their effort as their tenure increases. This decrease in effort over time is predicted due to the decrease in the reputational incentives agents have as their true skill level becomes known with higher precision by those observing them. When the precision on an agent's true ability is low, the outcome of each action they take as part of their job has a large impact on the principal's beliefs of their true ability. As the agent performs more actions and gains tenure, the precision on their true skill level increases. When the precision on their true ability is high, the benefits of exhibiting effort for a single action decrease as it will have a diminished effect on the principal's beliefs on their skill, thus reducing the agent's incentive to perform their tasks with high effort.

In this paper, I apply these predictions to Major League Baseball (MLB) umpires and test the hypothesis that longer tenured umpires exhibit less effort than their less experienced counterparts. In baseball, the job of an umpire is far more involved than just classifying pitches as balls and strikes. However, for this paper, I focus on pitch classification as it is likely the most important aspect of an umpire's job when they are positioned at home plate. Additionally, it remains one aspect of their job that is not yet subject to video review and coaches' challenges. As such, for the purposes of this paper, I consider an umpire's true ability to be completely measured by their ability to properly discern balls and strikes thrown by the pitcher. As long as baseball relies on human umpires to make these calls, there will be errors made. All else equal, an umpire expending more effort on calling balls and strikes should make fewer mistakes in classifying pitches than an umpire shirking from their duties.

As a measure of an umpire's effort, I consider the umpire's ability to fight their internal biases when placing the call as their desire to shirk from the task. In particular, I look at the umpire's tendency to be biased towards the winning team in late inning games in order to end the game faster so that they (and everyone else) may go home. Baseball games that remain tied at the end of 9 innings of play go into extra innings, which do not have a cap as too how many can be played[1]. This can take quite a bit of time, and crowds tend to thin out as the game goes deep into extra innings. There is some evidence that this desire for the game to end applies to the umpires who are calling the game as well (Lopez and Mills 2018).

The main incentive the umpires have to call the game correctly is their reputation. Umpires want to be known as fair callers of the game so that they can be considered to umpire for more prestigious playoff games which further their careers. Indeed, as umpires work their way up through the minor leagues this reputation is crucial, as it will be a major determinant of whether they will be promoted to umpiring the major league games. While the ball-strike calls an umpire makes during the games are not review-able yet in the major leagues, audiences and league officials have access to detailed PITCHF/x data both during and after the game. Even if the decision on the field cannot be overturned, umpires face a lot of scrutiny in their decisions. If a young umpire makes a high profile blown call, fans and league officials alike will be able to observe it. Younger umpires have much more uncertainty over their true skill, as such each blown call can do much more damage to their reputation than it could to a more tenured official.

A career concerns model is thereby appropriate to study the tendency of umpires to be biased towards ending a game earlier in extra innings games. A well-observed incorrect ball-strike call will do more damage to a young umpire's reputation and hurt their career prospects more significantly than they would a more experienced umpire. As such, I predict that this bias will be larger in longer-tenured umpires, in line with what a career concerns model would predict. To test this hypothesis, I estimate a linear probability model on the probability of making an incorrect call that can end the game faster in which the main coefficient of interest is the one on the tenure of umpires

[1]Rule changes implemented in the 2022 MLB season have since reduced the expected length of an extra innings game. However the game can still continue indefinitely if teams do not score.

making the ball-strike decision.

There are two main strands of literature to which this paper contributes to. The first body of work this paper contributes to is to the broader literature on career concerns. Several papers have studied how reputational concerns can influence otherwise rational actors to make sub-optimal decisions. Papers such as Dasgupta and Prat (2006) have examined how career concerns can affect decision making in financial managers. Others like Chevalier and Ellison (1999) find empirical evidence for this hypothesis using financial data. I contribute to this general body of work on the career concerns hypothesis by finding further evidence for career concerns in the setting of Major League Baseball.

This paper also contributes to the broad economics literature on decision making by various agents in sports. Papers such as Parsons et. al (2011) have looked at discrimination by Major League Baseball umpires when evaluating pitches by of pitchers of different races. Ohkusa (2010) also examines career concerns by studying Japanese baseball players. This paper is perhaps most similar to Garciano, Palacios-Huerta and Prendergast (2005), which examines how soccer referees may be pressured to make biased decisions under the pressure of the home stadium crowd. I contribute to this literature in a similar fashion by showing how career concerns can cause Major League Baseball umpires to act biased in a similar way due to their desire to end the game early.

The rest of the paper is organized as follows. In the following section, I provide a quick overview of how baseball deals with tied games and how the umpire is supposed to classify pitches. In the second section of this paper, I give an overview of the data I use to do the main analysis of the paper and discuss the empirical specifications that I estimate. Section three discusses the main findings of the paper. Section four concludes.

3.1.1 Overview of Baseball

Baseball is unique among the major American sports in terms of the length of time for which the game is played. While football, basketball, hockey and soccer are all guaranteed to end in a certain amount of time during the regular season, baseball games are not timed[2] and are not allowed to end in a tie. The game only ends after at least nine innings are played and after the completion of an entire inning where the score is no longer tied. It may also end in the "bottom" half of the inning when the home team takes the lead, since the away team has already had equal opportunity to score and can no longer score. This rule can lead to games being played for extremely long periods of time. On May 1st, 1920, the Brooklyn Robins and the Boston Braves played 26 innings before the umpire called the game off. The New York Times even joked the next day that the umpire remembered he had an "appointment pretty soon with a succulent beefsteak." More recently, on April 17, 2008 the Colorado Rockies defeated the San Diego Padres 2-1 after 22 innings, more than double the length of the standard game. When combined with the fact that the umpires work games nearly every day, it is apparent that the job can be incredibly demanding, especially during these long, arduous extra-inning games. Assuming that effort is costly, umpires clearly have a strong incentive to desire a shortened game.

There are four umpires in each regular season game and six umpires in postseason games. One umpire is stationed behind home plate and is responsible for classifying a pitch as a "ball" or a "strike," which favor the hitting team and pitching team respectively. A This decision is not meant to be subjective, as the formal definition of a strike is a pitch which crosses the plate inside a rectangular area known as the strike-zone. Of course, umpires are human, and mistakes are often made. While Major League Baseball has introduced challenges to review some umpire decisions and possibly overturn them, this has not yet been extended to ball-strike calls made by the home plate umpire[3]. There is no way to review an umpire's ball-strike decision, and any mistakes are played through, often at the frustration of the team that was shortchanged by the incorrect call. These mistakes can have very large impacts on a game, a strike mistakenly called when there are already two strikes on a count will result in the batter being struck out erroneously. On the other hand, a ball mistakenly called when there are already three balls will result in the batter being

[2] Arguably no longer true in 2023 with the introduction of the pitch clock. However the full game still does not have a time limit.

[3] At the time of writing, video review for balls and strikes had began in some minor league games.

awarded first base and more likely to score.

These impacts the umpires can have on the game and the above facts about how the game can go on indefinitely provide motivation for the main research question of this paper. Do career concerns lead Major League Baseball umpires to bias their ball-strike calling in extra-innings games in order to end the game prematurely?

3.2 Data and Empirical Strategy

In order to study the effects of tenure on game-shortening biased calls, I take data from Retrosheet. Retrosheet is a nonprofit organization which compiles detailed information related to Major League Baseball. The first dataset I take from there contains information on all individuals involved with MLB games. These individuals include players, managers, and umpires. Importantly, this data also includes information on the debut date of each umpire, which I am able to use in order to calculate their length of time employed in the league, or "tenure."

Next, I use another dataset from Retrosheet which has detailed information on each game played during each MLB season. Most important to this paper is its detailed information of the umpire stationed at home plate in each game. This information, combined with the previous data described, allows me to observe the tenure of the home plate umpire. I also use some other information from this data such as data on the stadium the game was played at in order to construct home team fixed effects.

Lastly, I use data on each pitch thrown in a game from Major League Baseball's Baseball Savant search function. For this paper, I download data from all pitches thrown after the 9th inning in games which went to extra innings. The data have detailed information of the game setting, including the scores of both the away and home team, home plate umpire ID, and the IDs of teams playing the game. It has detailed pitch information, including the outcome of the pitch, and the coordinates on the x and z axis of the ball as it crosses home plate. As my paper focuses on situations where the home plate umpire determines the classification of the pitch, I only use data from pitches where the batter did not swing at the ball. That is, I drop all pitches where the outcome is not either “called strike” or “ball”.

These coordinate points are important, as the position of the ball as it crosses home plate is precisely what is supposed to determine whether it is classified as a “ball” or a “strike.” I use this location to create the variable “inbox” which represents if the pitch crosses home plate at a point which is inside the rectangle representing the strikezone. More precisely, whether a pitch i is located in the strikezone is determined by the following equation:

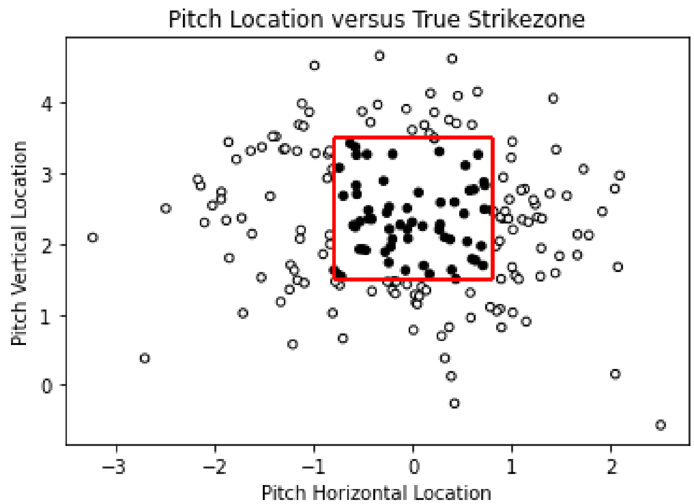
$$inbox_i \equiv \begin{cases} 1 & \text{if } x_{min} \leq x_i \leq x_{max} \text{ and } z_{min} \leq z_i \leq z_{max} \\ 0 & \text{else} \end{cases}$$

Where x_i and z_i are pitch i’s location on the x and z axis as it crosses home plate, respectively. x_{min} and x_{max} are the the left and right boundaries of the strikezone. Lastly, z_{min} and z_{max} are the bottom and top boundaries of the strikezone. I use the boundaries used by Brooks Baseball in their strikezone figures to determine the values of these cutoff points.

3.2.1 Pitch Position and Classification

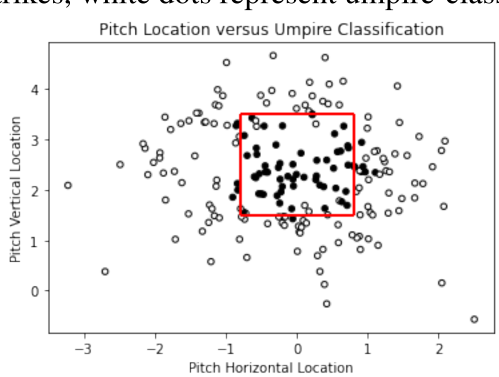
The position of a pitch as it crosses the plate in relation to the strikezone is sufficient to determine the correct classification of that pitch. Pitches located in the strikezone as they cross home plate should be labeled strikes and those that are not should be labeled balls. In Figure 3.1, I present a random sample of 200 pitches from the Baseball Savant data and what their correct classifications are.

Figure 3.1: Classification of 200 pitches using the defined strikezone using ranges on horizontal and vertical location of pitch as it crosses home plate. Black dots represent true strikes. White dots represent true balls.



In this figure, all pitches located inside of the red box are labeled strikes and all pitches located outside of it are labeled balls. A pitch's true classification, however, is determined by the home plate umpire. In Figure 3.2, I plot the location of the same 200 pitches with their classification as determined by the home plate umpire.

Figure 3.2: Actual classification of 200 sampled pitches based on umpire decision. Black dots represent umpire-classified strikes, white dots represent umpire-classified balls.



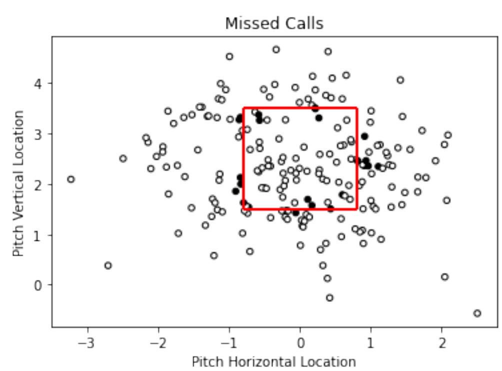
In the figure above, pitches colored white are classified as balls by the umpire and pitches colored black are classified as strikes by the umpire. As in Figure 3.1, the outline of the true strikezone is colored red. Note that some pitches inside the box are colored white and some outside the box are colored black. These are examples where the umpire fails to classify the pitch in accordance with the true strikezone.

I then use the data on pitch outcome and pitch location to construct one of my dependent variables, "bad_call" which represents if a pitch is incorrectly classified. Formally:

$$bad\_call_i \equiv \begin{cases} 1 & \text{if } (inbox_i == 0 \text{ and } outcome_i == \text{"called strike"}) \text{ or } (inbox_i == 1 \text{ and } outcome_i == \text{"ball"}) \\ 0 & \text{else} \end{cases}$$

The pitches that should be considered "bad calls" in Figure 3.2 are highlighted in Figure 3.3.

Figure 3.3: Umpire classification of the 200 pitches in figure 3.2 versus their proper classification. White dots represent pitches correctly classified by the umpire. Black dots represent pitches incorrectly classified by the umpire.



I merge the three data sources above to create a dataset where each observation is a pitch from an extra innings game from the 2010 through 2018 MLB seasons. Merging the two datasets required some brief data cleaning. During the data cleaning process, I rectify some inconsistencies in how teams are labeled. This included scenarios such as the New York Yankees being labeled either "NYA" or "NYY." The resulting data now have information on whether a pitch thrown landed in the strike zone, the call by the umpire, the score at the time of the pitch, and the umpire's tenure.

3.2.2 Outline of Studied Scenarios

Context is important in this setting, as I am most interested in extra inning situations in which the umpire makes a bad call that would move the needle towards the game ending faster, that is, the team that benefits is the team that is leading or in a better position to take the lead. I call these cases biased calls. For my main analysis, I look at three scenarios in extra innings games in which umpires may be able to shorten the game with their calls.

1. **Case One**: In the top half of an inning in an extra innings game, a game is likelier to end earlier if the team batting, which is the visiting team, scores. In these cases the away team must either be leading or the game is tied, since if the home team were leading the game would have ended the previous inning. As such, a biased call here would be an umpire erroneously labeling a pitch inside the strikezone a "ball" rather than a "strike," in order to assist the away team in scoring.

2. **Case Two**: In the bottom half on an extra innings game where the away team has taken the lead in the top half of that inning. In this case, the game will end unless the home team scores enough runs to tie the game at the end of the inning. Here the game will end most quickly if the home team fails to score at all. An umpire can make this likelier by having a very generous strikezone, in contrast with the last scenario in which they would have an incentive to have a strict one. In this situation, a biased call would be the umpire erroneously labeling a ball located outside of the strikezone a "strike" rather than a "ball."

3. **Case Three**: In the bottom half of an extra innings game where the score is tied. As in the second scenario, the home team is batting in this case, however as the score is tied, the quickest way for the game to end is for the home team to score as fast as possible. Therefore,

the umpire's incentives are similar to the first scenario described and they may be tempted to call a strict strikezone. As in the first scenario, a biased call here will be an umpire erroneously labeling a pitch inside the strikezone a "ball" rather than a "strike." However, in this case the biased call will benefit the home team.

It is important to note that what I call a "biased call" here is not that only kind of erroneous call an umpire can make. In the first setting for example, a pitch that crosses the plate outside of the strikezone called a "strike" is also an erroneous call. However, this is unlikely due to be explained by career concerns as such an error would actually be likely to increase the length of the game and, as a consequence, the umpire's required effort.

3.2.3 Empirical Specifications

I use a linear probability model in this section as it allows for the easiest interpretation of the coefficients of interest. Since my dependent variable is binary, I also include results using a logistic regression model in the appendix. Results using the logistic regression specification are similar to those discussed in this section using the linear probability model.

First, I test if umpires are more likely to make bad calls in general. For this specification, I use the full sample of pitches thrown in extra innings from the 2010 through 2018 MLB seasons. I estimate the following linear probability model:

$$bad\_call_i = \alpha + \beta_1 ump\_tenure_i + \beta_2 inning_i + \beta_3 margin_i + \nu_j + \varepsilon_i$$

Where ump_tenure_i is the number of years the umpire calling the pitch has worked since their debut at the major league level. $inning_i$ is the current inning of the game in which the pitch is thrown, and $margin_i$ is the number of runs the leading team is ahead by when the pitch is thrown and ν_j are stadium or home team fixed effects. bad_call_i is an indicator for if umpire's classification of the pitch is inconsistent with the proper classification determined by the strikezone.

Next, I focus my attention on the pitches where umpires are more directly able to influence the length of the game through their bias. In order to estimate the relationship between umpire tenure and the probability of a pitch call being biased, I estimate the following equation which is identical to the previous one except that it has a more specific sample and uses a different dependent variable:

$$bias\_call_i = \alpha + \beta_1 ump\_tenure_i + \beta_2 inning_i + \beta_3 margin_i + \nu_j + \varepsilon_i$$

The outcome $bias\_call_i$ is a dummy variable that is equal to 1 if the pitch is a "bad call" and if the pitch thrown falls into one of the three categories mentioned at the beginning of this section.

The main coefficient of interest in this specification is β_1. If β_1 is statistically significant and positive, it implies that longer tenured umpires are indeed more likely to miss calls that shorten the length of the game. This would match the prediction of a career concerns model. Since a longer tenured umpire would face fewer career ramifications if they were to make mistakes in order to reduce their effort.

I include the inning of the game and the margin of the game as additional controls. The motivation here is to include other factors that may influence an umpire's willingness to adjust the strikezone and affect the duration of the game. The impact of the inning of the game is somewhat ambiguous. On one hand, the longer the game goes on, the more tired the players, fans, and most relevantly, the umpire will be. As such, umpires may be more inclined to affect the result as the game goes on. On the other hand, games that go extremely long are likely to receive more attention. This could also increase the scrutiny on the umpire's behavior and drive them to be less bold. A game's margin is also likely to have an impact on the umpire's decision making, as the larger the margin is, the less the umpires calls make a difference. Lastly, I include home stadium fixed effects in order to account for any impacts the crowd has on umpire calling in each stadium. Other papers in the career concerns literature have found evidence that referees may be biased by the crowd as well, so I include these fixed effects to account for stadium-specific factors.

3.3 Results

I split the analysis into two sections. In the first, I study if longer tenured umpires are more likely to make errors when calling pitches in general. Specifically, I examine the coefficient on tenure when regressed on bad_call_i, which is an indicator for a pitch called a strike that was outside of the strike-zone or a pitch called a strike that was inside of the strike-zone. These erroneous calls may either reduce or lengthen the duration of the game. In the second section, I examine evidence for career concerns by specifically analyzing cases where errors reduce the expected length of the game and thereby the umpire's effort. I decompose this analysis into the three cases I discuss at the beginning of the previous section. For all results in this paper, standard errors are clustered at the home team or stadium level.

3.3.1 All Types of Bad Calls

In this section I use all of the PITCHF/x data from extra innings games to test if umpire's are more likely to miss ball-strike calls overall as their tenure increases. In Column 1 of table 3.1 below, I only include the main variable of interest, *ump_tenure*, in the specification. Column 2 includes additional controls, *margin* and *inning*. Column 3 is the full first specification given in the previous section of this paper and additionally includes home team fixed effects.

Table 3.1: Results from model testing for relationship between umpire tenure and general error rate on ball-strike calls in extra innings games.

	Bad Call	Bad Call	Bad Call
	(1)	(2)	(3)
Umpire Tenure	0.0003	0.0003*	0.0003*
	(0.0002)	(0.0002)	(0.0002)
Inning		-0.0001	0.0003
		(0.0010)	(0.0011)
Margin		-0.0020	-0.0020
		(0.0018)	(0.0018)
R-squared	0.0001	0.0001	0.0018
R-squared Adj.	0.0000	0.0000	0.0010
Observations	41977	41977	41977
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

The main prediction of the career concerns hypothesis is that umpire effort will decrease as their tenure increases and the precision on the signal of their true ability increases. In the second and third specifications in this set of results, I find mildly statistically significant estimated coefficients on umpire tenure. On average, a one year increase in the home plate umpire's tenure is associated with a 0.03 percentage point increase in the probability they classify a pitch incorrectly. While the likelihood of an error is small on an individual pitch basis, the effect is much more significant when put in the context of a full game, as the average Major League Baseball game sees hundreds of pitches thrown. Coefficients on the margin and inning controls are small and not statistically significant from zero. I find evidence that the umpire error rate is affected the margin of the game score or by the inning the game is in.

3.3.2 Calls to End the Game

Next, I turn my attention from all types of pitch classification errors to those that, in certain scenarios, can reduce the length of the game. I consider each of the scenarios I describe earlier in the paper in the order that I mention them.

Case 1: Top of Inning

In the first scenario, I consider all pitches in the top half of extra innings games. In the top half of an extra innings game, the score is always either tied or in favor of the away team, which is currently batting. The home team cannot be in the lead as the game would have concluded in the previous inning if so. In order to expedite the end of the match, an umpire would want to be more generous to the batter, and might be include to miss a few strikes that should have been called for the home team's pitcher. I call these pitches of this type "bias balls," since calling the pitch a ball is biased towards ending the game prematurely. I estimate the specification below, for all pitches in the top half of extra innings.

$$bias\_ball_i = \alpha + \beta_1 ump\_tenure_i + \beta_2 margin_i + \beta_3 inning_i + \nu_j + \varepsilon_i$$

Results are presented below in table 3.2.

Table 3.2: Results from model testing relationship between umpire tenure and erroneous ball-strike calls in the top half of extra innings games in a way that would be expected to shorten the game. In this case biased calls would be classifying a pitch inside the strike zone as a ball.

	Bias Ball	Bias Ball	Bias Ball
	(1)	(2)	(3)
Umpire Tenure	0.0005*	0.0005*	0.0005**
	(0.0002)	(0.0002)	(0.0002)
Inning		-0.0007	-0.0006
		(0.0010)	(0.0010)
Margin		0.0010	0.0012
		(0.0019)	(0.0019)
R-squared	0.0003	0.0004	0.0020
R-squared Adj.	0.0003	0.0002	0.0005
Observations	22165	22165	22165
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

The main estimated coefficient of interest on umpire tenure is statistically significant and positive in all three columns of this set of results. This suggests that longer tenured umpires are more likely to classify pitches inside the strikezone as balls in extra innings games. This result is consistent with the predictions of career concerns models. More experienced officials, on average, seem to make life harder for the pitcher in the top half of extra innings games in order to encourage more offense. These results may be driven by their desire to end the game earlier. In a later part of this section of this paper, I provide more evidence that these results are driven by longer tenured umpires' desire to end the game rather than a more general deterioration of their accuracy. As in the previous set of results, the estimated coefficients on margin and inning are not statistically significant.

Case 2: Away Team Leading in Bottom

The second regression I estimate is for the next second type of biased call I describe. In this scenario, I consider pitches thrown in the bottom half of innings in extra innings games where the away team has taken the lead in the top half of that inning. In this situation, the home team must score or the game will end with the away team victorious. As such, an umpire motivated to end the game earlier would want to be generous to the pitcher and has an incentive to call pitches outside of the strikezone as strikes. For this specification, I use an indicator labeled “bias strike” for these types of missed calls as the dependent variable. The specification is otherwise the same the previous case. Results from this specification are presented in table 3.3.

Table 3.3: Results from model testing relationship between umpire tenure and erroneous ball-strike calls in a way that would be expected to shorten the game. In this scenario, I focus on the bottom half of extra innings games where thee away team has taken the lead. In this case biased calls would be classifying a pitch outside of the strike zone as a strike.

	Bias Strike	Bias Strike	Bias Strike
	(1)	(2)	(3)
Umpire Tenure	0.0004	0.0004	0.0004
	(0.0005)	(0.0005)	(0.0005)
Inning		-0.0007	-0.0007
		(0.0019)	(0.0019)
Margin		0.0000	-0.0000
		(0.0035)	(0.0035)
R-squared	0.0002	0.0002	0.0069
R-squared Adj.	0.0000	-0.0003	0.0011
Observations	5489	5489	5489
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Unlike the previous scenario and in the general case, the estimated coefficients on umpire tenure for this scenario are not statistically significant. I fail to find evidence in support of the career concerns hypothesis which would predict that longer tenured umpires would be more likely to make a biased strike call in this scenario. It is important to note that a biased strike call is more significant in this scenario than a biased ball call was in the previous scenario. In the first scenario, if the away team scores, the home team still has a chance to respond in the bottom half of the inning. However, a biased strike call in this scenario has the potential to end the game immediately. As such, it is possible that umpires are more willing to make less significant biased calls than they are to make more significant ones.

Case 3: Bottom of Inning, Tied Game

The final case I consider is again in the bottom half of an inning, but when the score is tied. The fastest way for the game to end is for the home team to score, which would suggest the umpires have an incentive to be unfriendly towards the pitcher. The specification is similar to the first case, with bias_ball as the dependent variable, however the sample of pitches is taken from the bottom half of innings rather than the top. Additionally, for this specification I drop the margin control, as all pitches in this scenario are thrown with game tied.

Table 3.4: Results from model testing relationship between umpire tenure and erroneous ball-strike calls in a way that would be expected to shorten the game. In this scenario, I focus on the bottom half of extra innings games where the game remains tied. As was the case in the first scenario above, biased calls here would be classifying a pitch inside the strike zone as a ball.

	Bias Strike	Bias Strike	Bias Strike
	(1)	(2)	(3)
Umpire Tenure	0.0001	0.0001	0.0001
	(0.0002)	(0.0002)	(0.0002)
Inning		0.0004	0.0005
		(0.0013)	(0.0013)
R-squared	0.0000	0.0000	0.0022
R-squared Adj.	-0.0001	-0.0001	0.0001
Observations	14323	14323	14323
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Results in this table are again inconclusive throughout, failing to provide any evidence that more experienced umpires are favoring the home team in the bottom half of the inning. The estimated coefficients on umpire tenure are all statistically insignificant. This result is particularly surprising when considered in the results of Garciano, Palacios-Huerta and Prendergast's 2005 paper. In that paper the authors discuss the incentives soccer referees have to make stoppage time decisions favoring the home team. In this scenario, Major League Baseball umpires would be able to both favor the home team and also shorten the game by calling strikes poorly. Contrast this with the first case, where in order to end the game sooner, umpires would have to favor the visiting team.

Umpire Accuracy Deterioration Counterfactual

While I find strong evidence that longer-tenured umpires are more likely to make errors that shorten the game in the first scenario above, it is important to address the possibility that these errors stem from more general deterioration in ball-strike calling accuracy as tenure increases. Longer-tenured umpires may simply be more error-prone in general in the top half of extra innings games and are not necessarily acting in order to reduce the length of the game.

In order to test that longer tenured umpires are not simply more careless in the top half of innings, I estimate a counterfactual specification which tests if there is also a statistically significant and positive relationship between umpire tenure and probability of the umpire erroneously calling a pitch outside of the strikezone a "strike" in the top half of an inning in an extra innings game. This effectively tests this counterfactual, as an umpire that is making more errors in general would also more more likely to make these types of errors as well. However, this mistake will make it harder for the batting team to score in a situation where scoring would shorten the game. As such, there is no incentive for the umpire to make these types of errors if they want to reduce their effort. Results from this specification are presented in the table below.

Table 3.5: Results from model testing the relationship between umpire tenure and calls that would not be expected to shorten the length of the game. Here, I focus on the counterpart to the first scenario, that is, when umpires mistakenly classify pitches inside the strikezone as balls. This would be expected to lengthen the game. As such, there is no effort-based incentive for umpires to make these errors.

	Bias Strike	Bias Strike	Bias Strike
	(1)	(2)	(3)
Umpire Tenure	-0.0002	-0.0001	-0.0001
	(0.0002)	(0.0002)	(0.0002)
Inning		-0.0009	-0.0009
		(0.0011)	(0.0012)
Margin		-0.0025	-0.0027
		(0.0024)	(0.0024)
R-squared	0.0000	0.0001	0.0019
R-squared Adj.	-0.0000	-0.0000	0.0005
Observations	22165	22165	22165
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

In this specification I find a negative and statistically insignificant coefficient on ump_tenure. Therefore I fail to find any evidence that longer-tenured umpires are also more likely to make errors when calling pitches that are more likely to lengthen the game in the top half of the inning. This strengthens the argument that the statistically significant results from the first scenario analyzed are indeed due to career concerns and umpires trying to reduce the length of the game.

3.4 Conclusion

In this paper I investigate if predictions made by career concerns models are found in the behavior of Major League Baseball umpires. Major League Baseball collects extremely detailed information about every pitch thrown, allowing for thorough review of how umpires perform when calling pitches. This ability to review an umpire's performance fits well with the idea of reputational risks found within the career concerns model. The ability of umpires to dramatically change the outcome of an at-bat, and the nature of the game of baseball give them incentives to make some "mistakes" that can end the game earlier and send them home for the night. This allows me to test the career concerns model, and see if there is evidence that longer-tenured umpires shirk their duties since they have less incentives to improve their reputation.

I first investigate the effect of umpire tenure on their overall ability to properly call balls and strikes in extra innings games. I find a statistically significant and positive relationship between umpire tenure and the probability they make a erroneous ball-strike classification. This suggests the longer tenure umpires are indeed more likely to make an error when calling balls and strikes. I then examine three scenarios in which an umpire may be tempted to miss a call in a certain direction that would accelerate the completion of the game. I find a statistically significant and positive relationship between umpire tenure and the probability of failing to call a pitch inside the strike-zone a strike in the top half of extra innings games. This gives support to the career concerns prediction that a longer-tenured umpire is more likely to help the visiting team in order to end the game faster. I also show that the significant results in the top half of an extra innings game are likely to be related to effort and not just carelessness by checking the effect of umpire tenure on errors that would lengthen the game and find no significant effects. However, I fail to find evidence supporting career concerns in the bottom half of the inning when the score is tied or when the away team has taken a lead. This is despite the fact that when the game is tied in this scenario, making a biased call in this situation should be even more compelling to the umpire, as the umpire can also win the crowd over at the same time.

The results of the three cases are interesting, as the situation where the umpire tenure effect is most significant is actually the situation in which their bias may have the least impact. If the visiting team takes the lead in the top half of an inning, the home team can still strike back in the bottom half. On the other hand, the home team taking the lead in the bottom half or failing to tie the game when the away team is leading would end the game immediately. It is possible that umpires do want to end the game early, but are not willing to make more consequential errors in order to achieve this. They may view assisting the visiting team in the top half as a “small” mistake that can easily have no real impact. Future work should investigate these scenarios in more detail. Additionally, it would be interesting to revisit this analysis when video replay on ball-strike call reaches the major league levels. Umpires may further reduce their tendency to make biased calls when a video replay system can double check their work immediately.

Bibliography

[1] Ahrens, Thomas, and Laurence Ferry. "Financial resilience of English local government in the aftermath of COVID-19." *Journal of Public Budgeting, Accountin & Financial Management* (2020).

[2] Alexander, Diane, and Ezra Karger. "Do stay-at-home orders cause people to stay at home? Effects of stay-at-home orders on consumer behavior." *The Review of Economics and Statistics* (2020): 1-25.

[3] Autor, David H., and David Dorn. "The growth of low-skill service jobs and the polarization of the US labor market." American economic review 103, no. 5 (2013): 1553-1597.

[4] Bar-Isaac, Heski, and Joyee Deb. "What is a good reputation? Career concerns with heterogeneous audiences." International Journal of Industrial Organization 34 (2014): 44-50.

[5] Bartik, Alexander W., Marianne Bertrand, Zoe Cullen, Edward L. Glaeser, Michael Luca, and Christopher Stanton. "The impact of COVID-19 on small business outcomes and expectations." *Proceedings of the national academy of sciences* 117, no. 30 (2020): 17656-17666.

[6] Bernstein, Shai, Emanuele Colonnelli, Xavier Giroud, and Benjamin Iverson. "Bankruptcy spillovers." *Journal of Financial Economics* 133, no. 3 (2019): 608-633.

[7] Bloom, Nicholas, Robert S. Fletcher, and Ethan Yeh. "The impact of COVID-19 on US firms." No. w28314. *National Bureau of Economic Research* (2021).

[8] Caselli, Francesca, Francesco Grigoli, Weicheng Lian, and Damiano Sandri. "The great lockdown: dissecting the economic effects." *World Economic Outlook* (2020): 65-84.

[9] Chalermpong, Saksith. "Empirical study of economic spillovers of interstate 105 in Los Angeles county." *Transportation research record* 1864, no. 1 (2004): 94-102.

[10] Chetty, Raj, John N. Friedman, Nathaniel Hendren, and Michael Stepner. "The economic impacts of COVID-19: Evidence from a new public database built using private sector data." No. w27431. *National Bureau of Economic Research* (2020).

[11] Chevalier, Judith, and Glenn Ellison. "Career concerns of mutual fund managers." The Quarterly Journal of Economics 114, no. 2 (1999): 389-432.

[12] Cortes, Guido Matias, and Eliza Forsythe. "Heterogeneous Labor Market Impacts of the COVID-19 Pandemic." *ILR Review* (2020): 00197939221076856.

[13] Danisman, Gamze Ozturk, Ender Demir, and Adam Zaremba. "Financial resilience to the COVID-19 pandemic: The role of banking market structure." *Applied Economics* 53, no. 39 (2021): 4481-4504.

[14] Dasgupta, Amil, and Andrea Prat. “Financial equilibrium with career concerns.” Theoretical Economics 1, no. 1 (2006): 67-93.

[15] Davis, Noah, and Michael Lopez. “Umpires Are Less Blind than They Used to Be.” FiveThirtyEight, August 19, 2015. https://fivethirtyeight.com/features/umpires-are-less-blind-than-they-used-to-be/.

[16] Elenev, Vadim, Luis E. Quintero, Alessandro Rebucci, and Emilia Simeonova. “Direct and spillover effects from staggered adoption of health policies: Evidence from covid-19 stay-at-home orders.” No. w29088. *National Bureau of Economic Research, 2021.*

[17] Farrell, Diana, and Chris Wheat. “Bend, Don’t Break: Small Business Financial Resilience after Hurricanes Harvey and Irma.” *JPMorgan Chase Institute* (2018).

[18] Garicano, Luis, Ignacio Palacios-Huerta, and Canice Prendergast. “Favoritism under social pressure.” Review of Economics and Statistics 87, no. 2 (2005): 208-216.

[19] Green, Etan, and David P. Daniels. “What does it take to call a strike? Three biases in umpire decision making.” In 8th Annual Sloan-MIT Sports Analytics Conference Proceedings. 2014.

[20] Gollwitzer, Anton, Cameron Martel, William J. Brady, Philip Pärnamets, Isaac G. Freedman, Eric D. Knowles, and Jay J. Van Bavel. “Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic.” *Nature human behaviour* 4, no. 11 (2020): 1186-1197.

[21] Lopez, Michael, and Brian Mills. “Everyone Wants to Go Home during Extra Innings - Maybe Even the Umps.” FiveThirtyEight, April 18, 2018. https://fivethirtyeight.com/features/everyone-wants-to-go-home-during-extra-innings-maybe-even-the-umps/.

[22] Holtz, David, Michael Zhao, Seth G. Benzell, Cathy Y. Cao, Mohammad Amin Rahimian, Jeremy Yang, Jennifer Allen et al. “Interdependence and the cost of uncoordinated responses to COVID-19.” *Proceedings of the National Academy of Sciences* 117, no. 33 (2020): 19837-19843.

[23] Hubbard, R. Glenn, and Michael R. Strain. “Has the Paycheck Protection Program Succeeded?”. No. w28032. *National Bureau of Economic Research*, 2020.

[24] Belayeth Hussain, A. H. M., Noraida Endut, Sumonkanti Das, Mohammed Thanvir Ahmed Chowdhury, Nadia Haque, Sumena Sultana, and Khandaker Jafor Ahmed. “Does financial inclusion increase financial resilience? Evidence from Bangladesh.” *Development in Practice* 29, no. 6 (2019): 798-807.

[25] Klapper, Leora, and Annamaria Lusardi. “Financial literacy and financial resilience: Evidence from around the world.” *Financial Management* 49, no. 3 (2020): 589-614.

[26] Ohkusa, Yasushi. “Additional evidence for the career concern hypothesis with uncertainty of the retirement period-the case of professional baseball players in Japan.” Applied Economics 31, no. 11 (1999): 1481-1487.

[27] Ostrowski, Jeff. "Fewer than 4 in 10 Americans Could Pay a Surprise $1,000 Expense from Savings, Survey Finds." Bankrate, January 11, 2021. https://www.bankrate.com/banking/savings/financial-security-january-2021/.

[28] Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. "Strike three: Discrimination, incentives, and evaluation." American economic review 101, no. 4 (2011): 1410-1435.

[29] Piccolo, Alessio, and Roberto Pinto. "Financial Resilience in Labor Negotiations." *Kelley School of Business Research Paper* 2021-22 (2021).

[30] SafeGraph. The impact of coronavirus (COVID-19) on foot traffic, https://www.safegraph.com/dashboard/covid19-commerce-patterns (2020).

[31] Spiegel, Matthew, and Heather Tookes. "Business restrictions and COVID-19 Fatalities." *The Review of Financial Studies* 34, no. 11 (2021): 5266-5308.

Appendix A: Additional Figures and Tables

A.1 Chapter 1

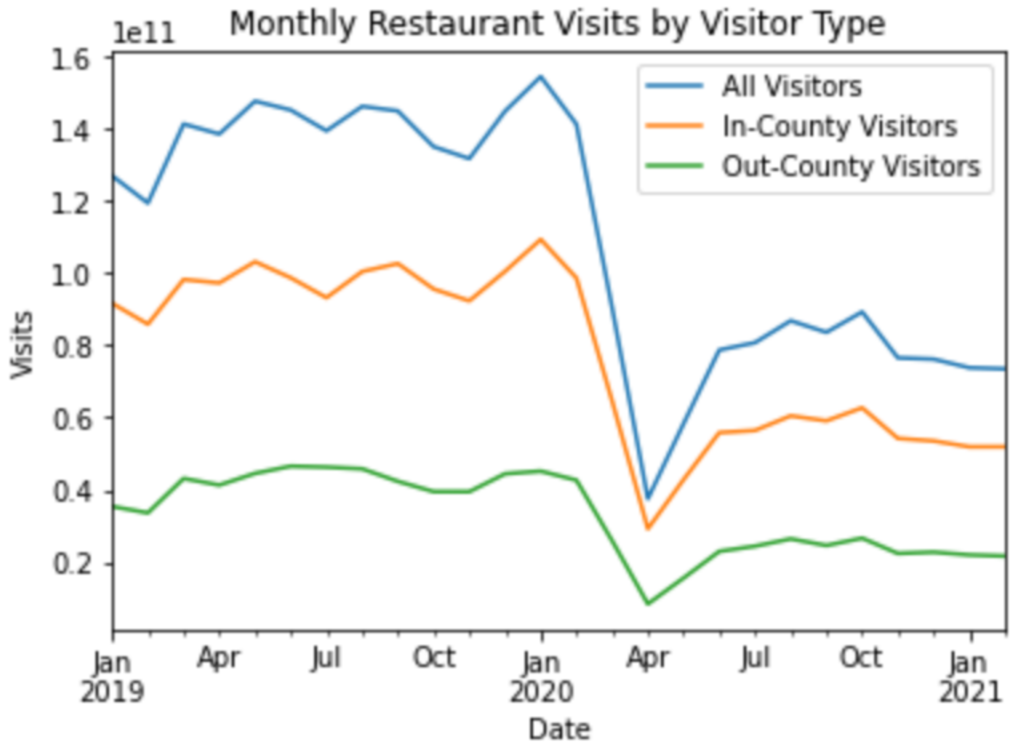
Table A.1: Event study regression on change in the number of open merchants. Event 0 is the indicator for the period the main county in a pair imposes a stay-at-home order. The time period before this (-1) is dropped.

	Δ Merchants$_{t,Jan2020}$		
Event$_{pre}$	-0.0733	Event$_{12}$	-0.3717*
	(0.0658)		(0.1791)
Event$_{-4}$	-0.0472	Event$_{13}$	-0.3679*
	(0.0504)		(0.1812)
Event$_{-3}$	-0.0181	Event$_{14}$	-0.3720*
	(0.0396)		(0.1886)
Event$_{-2}$	0.0000	Event$_{15}$	-0.3708*
	(0.0253)		(0.1935)
Event$_{0}$	-0.0612**	Event$_{16}$	-0.3651*
	(0.0250)		(0.2025)
Event$_{1}$	-0.0813**	Event$_{17}$	-0.3629
	(0.0376)		(0.2160)
Event$_{2}$	-0.1064**	Event$_{18}$	-0.3594
	(0.0503)		(0.2199)
Event$_{3}$	-0.1378*	Event$_{19}$	-0.3555
	(0.0706)		(0.2308)
Event$_{4}$	-0.1603*	Event$_{20}$	-0.3571
	(0.0892)		(0.2372)
Event$_{5}$	-0.1723	Event$_{21}$	-0.3656
	(0.1026)		(0.2489)
Event$_{6}$	-0.1938	Event$_{22}$	-0.3980
	(0.1143)		(0.2621)
Event$_{7}$	-0.2480*	Event$_{23}$	-0.3879
	(0.1323)		(0.2705)
Event$_{8}$	-0.2698*	Event$_{post}$	-0.3652
	(0.1419)		(0.2764)
Event$_{9}$	-0.3325**		
	(0.1558)		
Event$_{10}$	-0.3873**		
	(0.1677)		
Event$_{11}$	-0.3616**		
	(0.1711)		
Month-Year FE	Yes		
County-Pair FE	Yes		
R-squared	0.5552		
R-squared Adj.	0.5281		
N	2002		

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

A.2 Chapter 2

Figure A.1: Monthly restaurant visits to restaurants in the United States broken down by type of visitor.



A.3 Chapter 3

Table A.2: Results from logistic regression model testing for relationship between umpire tenure and general error rate on ball-strike calls in extra innings games.

	Bad Call	Bad Call	Bad Call
	(1)	(2)	(3)
Umpire Tenure	0.0028*	0.0028*	0.0031*
	(0.0017)	(0.0017)	(0.0017)
Inning		-0.0010	0.0027
		(0.0097)	(0.0098)
Margin		-0.0178	-0.0184
		(0.0157)	(0.0158)
Pseudo R-squared	0.0001	0.0001	0.0023
Observations	41977	41977	41977
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A.3: Results from logistic regression model testing relationship between umpire tenure and erroneous ball-strike calls in the top half of extra innings games in a way that would be expected to shorten the game. In this case biased calls would be classifying a pitch inside the strike zone as a ball.

	Bias Ball	Bias Ball	Bias Ball
	(1)	(2)	(3)
Umpire Tenure	0.0094**	0.0095**	0.0102**
	(0.0047)	(0.0048)	(0.0048)
Inning		-0.0137	-0.0113
		(0.0206)	(0.0205)
Margin		0.0209	0.0245
		(0.0401)	(0.0400)
Pseudo R-squared	0.0008	0.0009	0.0047
Observations	22165	22165	22165
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A.4: Results from logistic regression model testing relationship between umpire tenure and erroneous ball-strike calls in a way that would be expected to shorten the game. In this scenario, I focus on the bottom half of extra innings games where thee away team has taken the lead. In this case biased calls would be classifying a pitch outside of the strike zone as a strike.

	Bias Strike	Bias Strike	Bias Strike
	(1)	(2)	(3)
Umpire Tenure	0.0062	0.0064	0.0059
	(0.0068)	(0.0068)	(0.0070)
Inning		-0.0100	-0.0099
		(0.0279)	(0.0278)
Margin		0.0006	0.0002
		(0.0503)	(0.0498)
Pseudo R-squared	0.0004	0.0004	0.0146
Observations	5489	5489	5489
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A.5: Results from logistic regression model testing relationship between umpire tenure and erroneous ball-strike calls in a way that would be expected to shorten the game. In this scenario, I focus on the bottom half of extra innings games where the game remains tied. As was the case in the first scenario above, biased calls here would be classifying a pitch inside the strike zone as a ball.

	Bias Strike	Bias Strike	Bias Strike
	(1)	(2)	(3)
Umpire Tenure	0.0029	0.0026	0.0029
	(0.0044)	(0.0044)	(0.0045)
Inning		0.0407*	0.0426**
		(0.0208)	(0.0214)
Pseudo R-squared	0.0000	0.0006	0.0050
Observations	14323	14323	14323
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A.6: Results from logistic regression model testing the relationship between umpire tenure and calls that would not be expected to shorten the length of the game. Here, I focus on the counterpart to the first scenario, that is, when umpires mistakenly classify pitches inside the strikezone as balls. This would be expected to lengthen the game. As such, there is no effort-based incentive for umpires to make these errors.

	Bias Strike	Bias Strike	Bias Strike
	(1)	(2)	(3)
Umpire Tenure	-0.0024	-0.0021	-0.0021
	(0.0029)	(0.0029)	(0.0030)
Inning		-0.0142	-0.0128
		(0.0172)	(0.0177)
Margin		-0.0346	-0.0383
		(0.0318)	(0.0321)
Pseudo R-squared	0.0000	0.0002	0.0035
Observations	22165	22165	22165
Home Team FE	No	No	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$