I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15515

Measuring Small Business Dynamics and Employment with Private-Sector Real-Time Data

André Kurmann
Etienne Lalé
Lien Ta

AUGUST 2022

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15515

Measuring Small Business Dynamics and Employment with Private-Sector Real-Time Data

André Kurmann
Drexel University

Etienne Lalé
Université du Québec à Montréal and IZA

Lien Ta
Drexel University

AUGUST 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Measuring Small Business Dynamics and Employment with Private-Sector Real-Time Data*

The COVID-19 pandemic has led to an explosion of research using private-sector datasets to measure business dynamics and employment in real-time. Yet questions remain about the representativeness of these datasets and how to distinguish business openings and closings from sample churn – i.e., sample entry of already operating businesses and sample exits of businesses that continue operating. This paper proposes new methods to address these issues and applies them to the case of Homebase, a real-time dataset of mostly small service-sector sector businesses that has been used extensively in the literature to study the effects of the pandemic. We match the Homebase establishment records with information on business activity from Safegraph, Google, and Facebook to assess the representativeness of the data and to estimate the probability of business closings and openings among sample exits and entries. We then exploit the high frequency / geographic detail of the data to study whether small service-sector businesses have been hit harder by the pandemic than larger firms, and the extent to which the Paycheck Protection Program (PPP) helped small businesses keep their workforce employed. We find that our real-time estimates of small business dynamics and employment during the pandemic are remarkably representative and closely fit population counterparts from administrative data that have recently become available. Distinguishing business closings and openings from sample churn is critical for these results. We also find that while employment by small businesses contracted more severely in the beginning of the pandemic than employment of larger businesses, it also recovered more strongly thereafter. In turn, our estimates suggests that the rapid rollout of PPP loans significantly mitigated the negative employment effects of the pandemic. Business closings and openings are a key driver for both results, thus underlining the importance of properly correcting for sample churn.

JEL Classification: E01, E24, E32, E60

Keywords: small business activity, sample turnover versus business openings/closings, matching records, COVID-19, Paycheck Protection Program

Corresponding author:
Etienne Lalé
Université du Québec à Montréal
Department of Economics
C.P. 8888, Succ. centre ville
Montréal (QC), H3C 3P8
Canada
E-mail: lale.etienne@uqam.ca

* First version: April 28, 2020. This draft replaces earlier versions entitled "The Impact of COVID-19 on Small Business Dynamics and Employment: Real-time Estimates with Homebase Data". We thank Homebase, Safegraph and CrowdTangle for generously making their data available. We also thank John Coglianese, Ryan Decker, Cynthia Doniger, John Friedman, Chad Higdon-Topaz, Bart Hobjin, Erik Hurst and Jay Stewart as well as seminar participants at Arizona State University, the Deutsche Bundesbank, Drexel, McGill, Placekey, the St. Louis Fed, UQA`M, Wharton, the IZA workshop on labor market dynamics, the North American Meetings of the Econometric Society, the NBER Summer Institute, and FESAC for comments. All errors are our own.

1 Introduction

The COVID-19 pandemic unfolded with tremendous speed and affected labor markets in unprecedented ways. This led to an explosion of research with new private-sector establishment-level datasets to measure the state of the economy. The primary advantage of these datasets is that they provide information in real or near-real time and at considerably greater detail than what is typically available from official statistics.[1] At the same time, questions remain about the representativeness of these datasets and how to distinguish business openings and closings from sample churn – i.e., sample entries of already operating businesses and sample exits of businesses that continue to operate. These issues are of particular importance because, contrary to previous recessions, the disruptions brought about by COVID-19 were concentrated in service sectors where establishments with fewer than 50 employees account for about half of all jobs, and it is exactly among these small businesses where opening and closing rates are high even under the best of circumstances. Hence, the quality of real-time estimates of business dynamics and employment during the pandemic depends crucially on whether the dataset in question contains small businesses that are representative of the underlying population and on whether the estimates accurately incorporate the effects of small business openings and closings.

In this paper, we propose new methods to distinguish business openings and closings from sample churn, correct for sample selection issues, and assess the representativeness of the resulting estimates. We apply our methods to the case of Homebase, a scheduling and time clock software provider used by about 100,000 businesses in the U.S. that shares daily data on hours worked and hourly pay for each employee of their clients. The Homebase data is particularly well-suited for two reasons. First, the majority of the businesses in the dataset are small and operate in the service sectors most exposed to the pandemic. Second, Homebase has experienced strong growth and client turnover both before and during the pandemic. It therefore provides an excellent case to assess the efficacy of our methods.

The main methodological contribution of the paper relative to the many other studies using Homebase or similar establishment-level datasets is that we match the Homebase establishment records by name and address with information on business activity from Safegraph, Google, and Facebook. We use the match with Safegraph, a database of over 7 million places of interest including many small service-sector businesses, to assess the representativeness of the Homebase sample by comparing visit patterns from anonymized cell-phone records to businesses in Homebase with visit patterns to the universe of corre-

[1] As described in detail below, statistical agencies in the U.S. publish monthly survey estimates for the labor market, but these estimates become available only with a lag of several weeks and contain limited breakdowns by industry and establishment size. Administrative data on the population of establishments and employment are available only at quarterly or annual frequency and are released several months to several years later. The situation in other countries is similar.

sponding small businesses in Safegraph. The match with Google / Facebook, in turn, allows us to estimate the probability of business closings and openings among sample exits and entries. We then adjust for sample selection by benchmarking the quarterly birth and death rates implied by these estimates against *pre-pandemic* population counterparts from the Bureau of Labor Statistics' (BLS) Business Employment Dynamics (BED) and the U.S. Census Bureau's Business Dynamics Statistics (BDS). Finally, we use the adjusted estimates to construct weekly time series of small business employment that directly incorporate gains from business openings and losses from business closings.

To assess the quality of our methods, we compare our estimates *during the pandemic* with population counterparts from the BED and the Quarterly Census of Employment and Wages (QCEW) that have recently become available. We find that our estimates provide a close fit. In addition, given the sparse frequency of these official data sources, we show that average weekly visits to our Homebase establishments move more or less in lockstep with average weekly visits to all corresponding small businesses in Safegraph. Together, the results provide in our view compelling evidence that our Homebase estimates are representative of small business dynamics during the pandemic.

We show that properly distinguishing closings and openings from sample churn is crucial for these results. Counterfactuals that include either none of the entries and exits into Homebase or that treat all entries and exits as openings and closings would produce wildly different estimates. As such, the paper offers a cautionary tale about the increasing use of private-sector datasets to construct estimates of employment and business activity. These datasets are opportunity samples subject to client turnover that can result in large, time-varying gaps between sample entry and exit and true business openings and closings. While Homebase may represent an extreme example in this regard, the issue is also important for other establishment-level datasets, including the Current Employment Statistics (CES) that forms the basis of the BLS' monthly payroll estimate.[2]

In the second part of the paper, we exploit the high frequency / geographic detail of the Homebase data to study two important questions: have small service-sector businesses been hit harder by the pandemic than larger firms; and to what extent did the Paycheck Protection Program (PPP), a major provision of the 2020 CARES Act that provided small businesses with financial support, help small business keep their workforce employed.

The analysis yields four key insights. First, small business employment in Retail Trade, Education &

[2]As discussed in more detail below, the employment estimates from the CES historically did not take into account the effects of openings and closings directly and instead estimated employment changes from business birth/death with an econometric model applied to past data. Faced with extraordinary numbers of business closings in the beginning of the pandemic, the BLS modified this procedure in April 2020.

Health, Leisure & Hospitality and Other Services – the four service sectors hit hardest by the pandemic – contracted by an estimated 14 million between mid-February and mid-April 2020, a staggering 45% decline, and then regained about 9 million by mid-June 2020. Between mid-June 2020 and late-November 2021 (the end of the sample under consideration), small business employment gradually recovered most of the losses, and for Retail Trade and Leisure & Hospitality even exceeded their pre-pandemic levels. The decline and subsequent rebound in the beginning of the pandemic are both larger than the CES estimates for total employment in the four sectors. This implies, consistent with Cajner et al. (2020) and Dalton et al. (2020), that small businesses initially contracted more severely but then also recovered more strongly than larger businesses. Our estimates show that the faster recovery of small business employment continued all through 2021, which provides interesting new evidence for the ongoing debate on whether small businesses are more sensitive to economic shocks than larger businesses (e.g., Moscarini and Postel-Vinay, 2012; Fort et al., 2013; or Haltiwanger et al., 2018).

Second, we show that temporary closings account for 70% of the initial decline in small business employment, with total closings spiking to 40% in mid-April 2020. In the months thereafter, about two thirds of closed businesses reopened, resulting in a cumulative closing rate in the four sectors of 17% one year after the start of the pandemic. This is only about two percentage points higher than the cumulative closing rate over the same time period one year prior, which implies, perhaps surprisingly but consistent with concurrent analysis by Crane et al. (2022) based on alternative measures of business closures as well as Decker and Haltiwanger (2022) and Fairlie et al. (2022) based on more recently released administrative data, that the pandemic has not led to a substantially higher rate of permanent shutdowns.[3]

Third, new businesses openings in the four sectors considered have added more than 1.5 million new jobs since the pandemic started, constituting an important driver of the recovery from mid-June 2020 onward. Compared to 2019, the rate of new business openings remains lower, however. This differs from recent evidence by Fazio et al. (2021) and Haltiwanger (2021) who report record numbers of new business applications since the pandemic started. The difference is likely due to the fact that it takes several quarters from business application to employment of workers, and that new applications are disproportionally for businesses without a physical store location (e.g. nonstore retailers) which are excluded from our data. Our estimates may therefore represent a lower bound of the recovery in small business employment.

Fourth, we estimate that counties with a larger share of delayed PPP loans experience a persistently

[3]Other studies that document a large but temporary spike in small business closures in the beginning of the pandemic are Chetty et al. (2020) based on Womply revenue data, Yelp (2020) based on Yelp data, and Vaan et al. (2021) based on Safegraph visit data.

lower recovery of small business employment. The estimates are based on a county-by-week panel built from our Homebase sample that, following the research design proposed by Doniger and Kay (2021), allows us to exploit plausibly exogenous local differences in the delay of obtaining a PPP loan due to the temporary exhaustion of PPP funding in mid-April 2020. The negative employment effect, which is conditional on a rich set of controls including establishment fixed effects and is not driven by pretrends, is in large part due to higher rates of business closings.[4]

The result underlines the importance of properly taking into account closing and opening margins, and suggests that the temporary exhaustion of PPP in mid-April 2020 occurred at a critical moment when many small business owners, faced with an unprecedented negative shock, had to decide whether to continue operating and hope for government support or cut their losses and close shop. Quantitatively, our estimates imply that without delays in PPP loans, small business employment would have been about 10% higher in mid-July 2020 and about 6% higher in January 2021. Despite the many valid critiques about PPP in terms of inadequate targeting and potential for fraud, this suggests that the rapid and extensive deployment of economic support in response to the pandemic was an important reason for the relatively strong recovery in small business activity.

Related literature. The paper contributes to an extensive literature using non-traditional establishment- or household-level data sources to measure the economic impact of the COVID-19 pandemic; e.g., for the U.S., Bartik et al. (2020), Bick and Blandin (2020), Cajner et al. (2020), Chetty et al. (2020), Coibon et al. (2020), Crane et al. (2022), Kahn et al. (2020), Lewis et al. (2021), or the Household Pulse Survey by the U.S. Census Bureau, among many others. A related set of studies focuses explicitly on employment changes of smaller businesses; e.g. Bartik et al. (2020), Dalton et al. (2020), Fairlie (2020), Fairlie et al. (2022), or the Small Business Pulse Survey by the U.S. Census Bureau.

Aside from the high-frequency estimates of the effects of the pandemic on small business activity that we provide, the novelty of the paper is that we systematically match establishment records with alternative information on business activity to assess the representativeness of the data, estimate business closings and openings, and integrate the resulting effects into a consistent measure of small business employment. This is critical to accurately measure small business employment throughout the pandemic and to quantitatively evaluate the role played by business closings and openings. While the Homebase data we use has its limitations in terms of coverage, its focus on small businesses in service sectors hit hardest by the pandemic and the considerable growth and turnover of its client base provide us with an

[4]The regressions incorporate controls for changes in county-level COVID infection and death rates, non-pharmaceutical interventions (NPIs), school closings and weather, as well as differential week fixed effects by pre-pandemic county average household income.

excellent stress test of our methods. In this sense, the paper can also be seen as a proof-of-concept.

Our paper also contributes to the vigorous debate on the effects of PPP. Autor et al. (2020), Chetty et al. (2020), and Hubbard and Strain (2020) exploit the 500 employee threshold for PPP loan eligibility and find only limited effects. Concurrent studies by Bartik et al. (2021), Bartlett and Morse (2020), Doniger and Kay (2021), and Granja et al. (2020) based on alternative identification techniques generally find larger effects, especially for smaller businesses. Consistent with Chodorow-Reich et al. (2020), this suggests that small businesses are generally more financially constrained and were therefore more dependent on PPP loan support. Our estimates provide further support for this conclusion and are broadly in line with the main results of Doniger and Kay (2021), whose research design we adopt. Importantly, our analysis provides an explanation for why the temporary exhaustion of PPP loans had such long-lasting effects: it is primarily due to the adverse effects that these delays had on business closings. This finding is confirmed in more recent studies by Autor et al. (2022) and Dalton (2021) who compare employment and closing probabilities of businesses who received a PPP loan earlier with those of businesses who received a PPP loan later.

The rest of the paper proceeds as follows. Section 2 introduces our employment estimator; Section 3 describes the data; Section 4 discusses how we distinguish business openings and closings from sample churn and provides first evidence about the representativeness of the Homebase data; Section 5 analyzes small business dynamics during the pandemic and further assesses the representativeness of our results; Section 6 estimates the effects of PPP loan delays; and Section 7 concludes.

2 Estimating small business dynamics and employment

This section provides a general description of our approach to construct an estimate of small business employment from a sample of establishment-level data that directly incorporates the effects of establishment openings and closings. For each sector (e.g. Leisure & Hospitality), we start with reference employment level $\widehat{E}_0$, taken from the QCEW, and estimate employment in week t as

$$\widehat{E}_t = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}}\right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{C}_{i,t}}\right)}, \tag{1}$$

where ω_i denotes the sampling weight for cell i (in our case at the industry-size-region level), constructed as the ratio of establishment population counts to sample counts in that cell in the reference period; $\hat{e}_{i,t}^{\mathcal{A}_{i,t}}$ denotes employment of the set of establishments $\mathcal{A}_{i,t}$ that are active in the sample in both week t

and $t-1$; $\hat{e}_{i,t}^{\mathcal{O}_{i,t}}$ denotes employment gains from the set of establishments $\mathcal{O}_{i,t}$ that are newly opening or reopening in week t; and $\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}}$ denotes employment losses from the set of establishments $\mathcal{C}_{i,t}$ that are closing temporarily or permanently in week t.

The key challenge in constructing this estimate is to distinguish business openings and closings from sample churn; i.e. entry of businesses that already operated previously, and exit of businesses that continue to operate thereafter. Including already operating entrants as part of $\mathcal{O}_{i,t}$ would mean that $\hat{e}_{i,t}^{\mathcal{O}_{i,t}}$ overestimates employment gains from openings. Vice versa, including exiters that continue to operate in $\mathcal{C}_{i,t}$ would mean that $\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}}$ overestimates employment losses from closings.

Sample churn is important for many private-sector establishment-level datasets since the underlying data providers commonly experience client turnover and/or growth in their customer base over time. As shown in Sections 4 and 5, this means that simple strategies such as treating all entries as openings, respectively treating all exits as closings will result in spurious estimates of establishment openings, closings, and employment growth. Similarly, ignoring all entries and exits (i.e. setting $\hat{e}_{i,t}^{\mathcal{O}_{i,t}} = 0$ and $\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}} = 0$) and effectively estimating total employment growth from the set of continuing establishments $\mathcal{A}_{i,t}$ will be misleading during large business cycle swings such as the pandemic when there are important changes in the relative rate of establishment openings and closings.

We address this challenge by constructing a direct estimate of $e_{i,t}^{\mathcal{O}_{i,t}}$ and $e_{i,t-1}^{\mathcal{C}_{i,t}}$ that exploits information on individual business activity from alternative sources. Conceptually, we estimate employment gains from establishment openings as

$$\hat{e}_{i,t}^{\mathcal{O}_{i,t}} = \sum_{\ell \in i} \hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t}) \times e_{\ell,t} \times \theta_{\ell,t}^{\mathcal{O}}, \tag{2}$$

where $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t})$ denotes the estimated probability of establishment ℓ opening conditional on entering the sample in week t; $e_{\ell,t}$ measures employment of establishment ℓ at entry; and $\theta_{\ell,t}^{\mathcal{O}}$ is a possibly establishment- and time-specific adjustment factor such that the resulting birth rate is consistent with pre-pandemic population counterparts from official statistics. Similarly, we estimate employment losses from establishment closings as

$$\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}} = \sum_{\ell \in i} \hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t}) \times e_{\ell,t-1} \times \theta_{\ell,t}^{\mathcal{C}}, \tag{3}$$

where $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t})$ denotes the estimated probability of establishment ℓ closing conditional on exiting the sample in week t; $e_{\ell,t-1}$ measures employment of establishment ℓ prior to exit; and $\theta_{\ell,t}^{\mathcal{C}}$ is an adjustment factor such that the resulting death rate is consistent with pre-pandemic population counterparts from official statistics. As explained in detail in Section 4, these adjustment factors correct for potential sample selection issues.

It is instructive to compare our approach with the CES employment estimator that the BLS uses for its monthly Employment Situation.[5] While our estimator in (1) is conceptually similar to the "weighted link-relative technique" of the CES estimator, our estimator directly incorporates employment gains from openings and employment losses from closings. The CES estimator, in contrast, includes only a portion of the establishments that report zero employment in month t and establishments that return to positive employment in month t, respectively, and then adjusts separately for employment effects from new openings and other closings with a econometric "net birth/death" model based on current and past data.[6] As such, the CES estimator may not reflect the net effect of changes in business openings and closings that occur in response to large business cycle shocks such as the pandemic whereas in our approach, $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t})$ and $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t})$ and therefore $e_{i,t}^{\mathcal{O}_{i,t}}$ and $e_{i,t-1}^{\mathcal{C}_{i,t}}$ adjust with real-time information about business activity. In addition, since our approach is formulated explicitly in terms of opening and closing probabilities, it is straightforward to benchmark the resulting birth and death rates to population counterparts from official statistics.

3 Data

The data we use to apply our estimation approach comes from Homebase (HB), a scheduling and payroll administration provider used primarily by small, independently owned businesses employing fewer than 50 workers. The majority of the businesses operate in service sectors with a large propensity for in-person interaction that were most exposed to the disruptions and stay-at-home orders in the beginning of the pandemic. As such, the HB data is particularly well-suited to assess our approach of estimating small businesses dynamics in response to large shocks.

The HB data consists of anonymized daily records of individual hours worked and wages of employees, linked longitudinally to the establishment where they work and the firm that owns the establishment (almost all businesses are single-establishment firms). The data is recorded in real-time through HB's

[5]See `https://www.bls.gov/web/empsit/cestn.htm` for details on the CES and estimation.

[6]Historically, the CES estimator only included establishments that reported positive employment in both t and $t-1$ and adjusted for the net effect of births versus deaths with an econometric model based on past data. By doing so, employment growth of non-respondents was effectively assumed to equal employment growth of respondents. In light of the large labor market disruptions caused by the COVID-19 pandemic, the BLS changed its methodology from April 2020 onward by directly including a portion of reported zeros in the estimator and by adding current period employment growth in the econometric model to predict the net birth/death effect. See https://www.bls.gov/web/empsit/cesbd.htm for details.

proprietary software and is used by many of the businesses for payroll processing. HB provides free data access to researchers and updates the data frequently with the latest observations.

In addition to the publicly available data, HB shares with us counts of owners and managers that use the HB software, scheduled hours of employees when available, and name and address information for each establishment. The information on owners and managers and scheduled hours allows us to construct a broader measure of employment. In turn, we use the name and address information to match HB establishments to data from Safegraph, Google, and Facebook. This is a key advance over other studies using the HB data that allows us to attribute a consistent industry classifier to each establishment; assess the representativeness of the HB sample; and, most importantly, distinguish business openings and closings from sample churn. The procedure involves extensive data cleaning and standardization before matching the records sequentially by exact merges and then fuzzy name match and substring match algorithms. The Appendix provides details on the different steps as well as match statistics. We only retain HB establishment records that match exactly or with a high match rate.[7]

3.1 Employment and business activity

For each establishment, we construct weekly employment as the sum of individuals with tracked hours (actual or scheduled) during that week plus owners and managers that show activity in the HB software but do not have tracked hours.[8] Including owners and managers broadens the employment concept beyond hourly paid workers, which attenuates somewhat the decline in estimated employment at the onset of the pandemic when owners and managers with untracked hours had a higher propensity to remain active than employees with tracked hours. But otherwise, the results are not affected materially by their inclusion.

For an establishment to be retained in our sample it must show up at least once for a spell of three consecutive weeks with at least 40 weekly tracked hours across its employees. We thus exclude establishments that use HB only for a short trial period. For an establishment in the sample to be active in a given week, it must have employees with tracked hours in that week. Establishment activity is therefore independent of owners and managers logging in to the HB software (e.g. for reporting purposes).

[7]We compare our match algorithm to Safegraph's Placekey matching tool and find that our algorithm results in higher match rates, primarily thanks to extensive pre-cleaning of establishment names and deduplication of establishment records. Details are available upon request.

[8]For establishments that report both scheduled and actual hours, we compare the two measures and find them to be very close to each other. We are therefore confident that scheduled hours are an accurate measure of actual hours worked.

3.2 Industry classification

The historical HB data comes with an industry category for each establishment, but the available categories do not line up with standard industry classification, and for about one third of the records, industry category is missing altogether. This is an important limitation for the purpose of constructing estimates that can be compared to official statistics. We address this issue by using the above-described match of HB establishments with Safegraph's Core Places data, which contains consistent NAICS-6 industry coding for over 7 million Places of Interest (POIs) in the U.S., including a large fraction of all private-sector establishments.[9]

3.3 Sample characteristics and representativeness

The sample we consider covers January 2019 through November 2021. The beginning of the sample is imposed by the availability of sizable HB data.[10] The end of the sample is due to upload limitations to relate HB establishments to Facebook postings.[11] The raw HB data contains about 300,000 distinct establishments; however, many of them do not use HB regularly and therefore do not satisfy our retention criterion. The sectors with the largest coverage are Leisure & Hospitality (NAICS 71 and 72), Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), and Other Services (NAICS 81).[12] To ensure good coverage, we focus our analysis on these four sectors and exclude all establishments with 50 employees or more.[13] According to official statistics, establishments with fewer than 50 employees accounted for about half of all jobs across the four sectors and for about 23% of all private-sector jobs prior to the pandemic. Hence, the segment covered by our analysis represents a sizable share of total employment and, as highlighted above, consists primarily of businesses with a significant in-person component that were among the hardest hit by the pandemic.

[9]See the Appendix for details on the Safegraph data and summary statistics about the NAICS industry codes for the matched HB establishments. In December 2020, HB independently started attributing NAICS industry codes for each establishment in their dataset. This classification is available only for establishments active from that month onward. Since many establishments that were active in 2019 and 2020 exited the HB sample before December 2020, the HB NAICS codes are not directly useful for our analysis. However, for establishments active in December 2020 and beyond, we find a high level of overlap between our NAICS codes and the HB NAICS codes.

[10]HB data is available starting in January 2018, but the sample size is relatively small until 2019.

[11]CrowdTangle, Facebook's research database, imposes limits on how many records (establishments) can be uploaded and matched. We have periodically renegotiated additional upload credits and plan to do so again in the future.

[12]See the Appendix for details. Other Services includes "Repair & Maintenance" (NAICS 811) and "Personal & Laundry Services" (NAICS 812), which contains many of the HB establishments categorized under "home and repair", "beauty and personal care", and "health care and fitness". Aside from these four sectors, the HB data also contains several hundred establishments each in "Utilities" (NAICS 22), "Construction" (NAICS 23), "Food, Textile & Apparel Manufacturing" (NAICS 31) and "Real Estate, Rental & Leasing" (NAICS 53).

[13]We also exclude "Non-store Retail" (NAICS 454) and "Private Households" (NAICS 814) because HB contains only very few establishments in these industries, and the QCEW that we use to benchmark does not contain these industries.

Table 1 reports the number of establishments that satisfy our retention criterion and that we can match with a high degree of confidence to Safegraph. Across the entire January 2019 to November 2021 period, the sample contains about 100,000 distinct establishments. The mid-February 2019 base sample consists of 38,193 establishments that show activity between the beginning of January 2019 and the second week of February 2019 of which 34,757 are active in the second week of February.[14] For the mid-February 2020 base sample, the corresponding establishment counts are 49,268 and 45,454. From mid-February 2019 to mid-February 2020, there are 13,289 exits without return and 25,149 new entrants, and from mid-February 2020 to late-November 2021, there are 28,256 exits and 37,777 new entrants. Foreshadowing the discussion below, these entry and exit rates are much larger than birth and death rates in the official statistics, implying that the HB data is subject to considerable sample churn.

Table 1: Establishment counts of retained Homebase sample

	Feb. 2019 - Feb. 20		Feb. 2020 - Nov. 21	
Mid-February base sample	38,193	(100%)	49,268	(100%)
- active in mid-February	34,757	(91.0%)	45,454	(92.3%)
- temporarily inactive in mid-February	3,436	(9.0%)	3,814	(7.7%)
Exits without return	13,289	(34.8%)	28,256	(57.4%)
New entrants	25,149	(65.8%)	37,777	(76.6%)

Notes: The first column shows counts of HB establishments from mid-February 2019 to mid-February 2020, and the second column shows counts of HB establishments from mid-February 2020 to the end of November 2021 that (i) were successfully matched to Safegraph; (ii) belong to either Retail Trade, Education & Health Services, Leisure & Hospitality, or Other Services; and (iii) have fewer than 50 workers when active in mid-February or when entering Homebase.

It is instructive to compare the size of our HB sample to the number of small businesses in the four sectors sampled by the CES. While the BLS does not publish breakdowns by industry and size, we know that in 2021 the CES sampled about 45,000 businesses in the four sectors considered, and the share of businesses with fewer than 50 employees sampled across all sectors was 60%.[15] Assuming that this share is similar across industries, the CES therefore sampled about 27,000 small businesses in the four sectors considered. By this calculation, our HB sample is considerably larger.

A potentially important issue with opportunity samples such as HB concerns the extent to which the establishments covered are representative of the population, which in our case consists of the universe of small businesses in the four sectors considered. First, note that the estimator in (1) corrects for distributional differences across industry-size-region cells i by weighing each cell by the ratio of QCEW

[14] The remaining 3,436 establishments are temporarily inactive; i.e active prior to mid-February and then active again at some point after mid-February. This is consistent with administrative data from the BED that also reports a substantial rate of temporary closings prior to the pandemic.

[15] The CES survey includes about 130,000 businesses (UI accounts), which cover approximately 670,000 individual work-sites or establishments. Hence, many of the businesses sampled are larger, multi-establishments firms.

to HB establishment counts in the mid-February 2020 base sample so that the relative importance of each cell in the estimator is the same as in the QCEW.[16] Hence, for representativeness to be an issue, it would need to be the case that HB establishments are systematically different from their population counterparts within the different cells.

To provide a first assessment of this concern, we compare the pre-pandemic average number of employees per establishment in each industry-size cell with the QCEW counterparts. As shown in the Appendix, for all the cells, the HB and QCEW numbers are very close to each other, implying that at least prior to the pandemic, the HB establishments look highly representative within industry-size cells.

In Sections 4 and 5, we will further assess the representativeness of the HB data by comparing annual growth rates of small business employment implied by our estimates with population counterparts from the QCEW. In addition, given the sparse frequency of the QCEW data, we will pursue a novel approach that exploits our match of HB establishments with Safegraph POIs, which allows us to compare weekly visit patterns to the establishments in our sample with visit patterns to the entire Safegraph sample of small businesses in the four sectors considered. For both of these comparisons, we find a very close fit.

4 Distinguishing business closings and openings from sample churn

This section describes the main methodological contribution of the paper, which consists of (i) estimating the probability $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t})$ that an establishment ℓ exiting the HB sample in week t is a closing, respectively the probability $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t})$ that an establishment ℓ entering the HB sample in week t is an opening; and (ii) adjusting these estimates by benchmarking them against pre-pandemic birth and death rates from official statistics. We then assess the quality of our methodology by comparing the resulting small business employment estimates for the pre-pandemic period with population counterparts.

4.1 Identifying business closings and openings

To identify business closings in week t, we take all exiters defined as establishments that are active in HB in week $t-1$ and become inactive in week t and proceed in four steps. First, we check whether an exiter returns to activity in HB by the end of the sample. If so, we classify the exiter as a temporary closing and assign probability $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t}) = 1$. Second, for exiters that do not return to activity by the

[16] The Appendix provides details on distributional differences. While the HB sample has generally good coverage across all industry-size class cells, it under-represents the smallest establishment size class (1-4 employees) and over-represents the other size classes (5-9, 10-19, and 20-49 employees). Furthermore, the HB sample over-represents NAICS 72 (Accommodation & Food Services) at the expense of NAICS 62 (Health Care & Social Assistance) and NAICS 81 (Other Services), and the geographic distribution skews slightly towards Florida and Texas at the expense of California.

end of the sample, we match them to Google Places using Google's API service and assign probability $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t}) = 1$ if Google tags them as "temporarily closed" or "permanently closed".[17] These tags are reported by business owners and customers but cover only a subset of all closed establishments. Hence, as a third step, we match the remaining exiters to Facebook using CrowdTangle, Facebook's research database, and check whether establishments with regular posting histories while being active in HB stop posting regularly after exiting HB. If so, we assign probability $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t}) = 1$. Fourth and finally, for exiters that we cannot match to either Google or Facebook or that do not post regularly on Facebook while being active in HB, we assign probability $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t})$ equal to the industry-size cell average probability obtained in step three.

To identify business openings, we adopt a similar method except that Google Places does not contain a "new opening" tag. We take all entrants defined as establishments that are inactive in week $t-1$ and become active in week t and proceed in three steps. First, we check whether an entrant was active in HB at some point before week $t-1$. If so, we classify the entrant as a reopening and assign probability $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t}) = 1$. Second, for entrants that become active in HB for the first time in week t, we match them to Facebook and check whether they start posting regularly only after entering HB. If so, we assign probability $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t}) = 1$. Third, for the entrants that we cannot match to Facebook or that do not post regularly after becoming active in HB, we assign probability $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t})$ equal to the industry-size cell average probability obtained in the previous step.

As a basic quality check of our identification approach, we exploit the match with Safegraph and compare visit patterns for the different types of establishments in our HB sample. First, we verify that exits identified as business closings exhibit a large drop off in average weekly visits relative to establishments that remain active in HB. Second, we verify that entries identified as new openings appear in Safegraph only after entering HB. Third, we verify that exits and entries identified as sample churners show average visit patterns similar to those of establishments that remain active in HB. These results, which are available in the Appendix, provide independent support for our approach to identify business closings and new openings.

It is worth mentioning that in principle, the Safegraph visits data could be used to identify *individual* establishment openings and closings from their visit patterns. For instance, Crane et al. (2022) define a Safegraph POI as closed if year-over-year visits decline by more than a certain threshold. However, we found after extensive analysis that the Safegraph visits data can be very noisy at the individual estab-

[17] We include both Google tags "temporarily closed" and "permanently closed" because "temporarily closed" is sometimes updated to "permanently closed", and because our first step only identifies temporary closings that have returned by the end of the sample.

lishment level, especially for POIs in buildings with other occupants (e.g. malls, multi-story buildings) or POIs that conduct a lot of their business by delivery. This makes reliable identification of business closings extremely challenging. Furthermore, Safegraph regularly backfills historical visit data back to 2017, independent of whether a POI operated during the entire period or newly opened at some point after 2017.[18] This means that the visits data cannot be used to identify business openings. We therefore use information from Google and Facebook to identify openings and closings and exploit the Safegraph visits data only to assess the representativeness of aggregates for which measurement issues are likely to average out.

4.2 Benchmarking against administrative data

The second step of our methodological contribution consists of calculating adjustment factors $\theta^{\mathcal{C}}_{\ell,t}$ and $\theta^{\mathcal{O}}_{\ell,t}$ such that the birth and death rates implied by our estimates are consistent with *pre-pandemic* population counterparts from official statistics. These adjustment factors correct for possible selection issues that arise if the establishments exiting and entering the HB sample have a different propensity to be closings and openings than in the population.[19] Note that this adjustment procedure applies only to new openings and permanent closings, but not to temporary closings and reopenings. The reason for this is two-fold. First and as already mentioned above, Safegraph visits data indicate that the HB sample of both continuously open and temporary closed and reopening establishments is highly representative of the larger population of small businesses (see above). Second, there is no good publicly available official data of temporary closings and reopenings at the industry-size cell level.

We use two administrative data sources covering the quasi-population of establishments to calculate the adjustment factors. The first is the BED, which consists of all longitudinally linked establishments from the QCEW and reports birth and death rates by sector but *not* by establishment size class at quarterly frequency with a delay of about six months. The second source is the BDS, which consists of all longitudinally linked employer establishments from U.S. tax records and reports entry and exit rates defined similarly as BED birth and death rates by industry *and* size class, but only at an annual frequency and with the latest available data for 2019.[20] Since very small establishments have substantially

[18] Further information on these issues with Safegraph visits data is available upon request.

[19] The adjustment factors also correct for measurement error that could arise if, for example, Google closed tags are systematically inaccurate or if businesses posts regularly on Facebook even though they are closed. We assess the potential importance of such measurement issues by looking at Google closed tags and Facebook posting histories for businesses identified as temporarily closed in step one and find few discrepancies. Nevertheless, measurement error remains a possibility.

[20] See the Appendix for details on the definitions of birth and death rates in the BED, respectively entry and exit rates in the BDS, and a comparison of these rates. We note that while in Retail Trade and Leisure & Hospitality, pre-pandemic BED birth and death rates line up closely with BDS entry and exit rates, in Education & Health and Other Services, BED

higher birth and death rates even in normal times, and since we want to compare our estimates to official counterparts *during* the pandemic, we combine the BED with the more detailed cross-sectional information from the BDS to obtain birth and death rates by sector and size class at a quarterly frequency that extends through the first year of the pandemic. Specifically, for each sector, we compute ratios of entry and exit rates by size class relative to the sectoral average from the BDS and multiply these ratios with the sectoral birth and death rates from the BED. We label the resulting BED/BDS birth and death rates for each sector-size class i and quarter q as $p(\text{birth}_{i,q})$ and $p(\text{death}_{i,q})$.

For closings, we calculate an adjustment factor $\theta^{\mathcal{C}}_{\ell,t} = \theta^{\mathcal{C}}_{i,q}$ for each establishment $\ell \in i$ that permanently exits the sample in week $t \in q$ of 2019 such that the death rate for pre-pandemic quarter q in sector-size cell i implied by the thus adjusted HB closing probabilities equals the BED/BDS counterpart $p(\text{death}_{i,q})$. In practice, since BED/BDS death rates vary little over 2019 and are similar for all but the smallest size class (1-4 employees), we average over all quarters of 2019 and across sector-size cells for which the HB sample contains relatively few permanently exiting establishments. We use these adjustment factors when estimating the employment losses from closings during the pandemic. Hence, closing probabilities and therefore death rates are allowed to vary in real-time during the pandemic, but the adjustment factors that correct for selection remain fixed.

For openings, we proceed similarly by calculating adjustment factors for each establishment $\ell \in i$ that newly enters the sample in week $t \in q$ of 2019 such that the average birth rate for 2019 in sector-size cell i implied by the thus adjusted HB opening probabilities equals the BED/BDS counterpart $p(\text{birth}_{i,q})$. Different from above, however, we allow these adjustment factors to be time-varying as a function of the number of new entrants relative to total sample size; i.e., $\theta^{\mathcal{O}}_{\ell,t} = \theta^{\mathcal{O}}_{i} \times (\frac{n^{\text{entry}}_{i,t}/n^{\mathcal{A}}_{i,t}}{n^{\text{entry}}_{i,0}/n^{\mathcal{A}}_{i,0}})^{-1}$, which is akin to inverse probability weighting. The reason for this difference is that in contrast to exits, which are naturally bounded by total sample size, entry can vary independently of sample size.[21] Hence, opening probabilities and adjustment factors are both allowed to vary in real-time during the pandemic.

The Appendix provides further details on the methodology and reports the different industry-size adjustment factors. Generally, adjustment factors for the smallest size class (1-4 employees) are larger than one, while adjustment factors for the larger size classes (5-9, 10-19, 20-49 employees) are smaller

birth and death rates are notably higher than BDS entry and exit rates. The reason for this difference is unclear and can only be investigated with access to the underlying micro-data. This illustrates that even in official data, measuring birth and death rates is far from trivial and depends crucially on the definition of (employer) establishments.

[21] As a simple example, suppose that there are no selection issues; i.e., $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t}) = p(\text{birth}_{i,q})$ for all $\ell \in i$ and $t \in q$. Now consider a large increase in sample entry (e.g. as a result of increased customer acquisition efforts by the data provider) such that the number of new openings doubles. Then with constant adjustment factors, employment gains from openings would spuriously double, thereby overstating estimated employment growth. By making the adjustment factor time-varying as a function of $n^{\text{entry}}_{i,t}/n^{\mathcal{A}}_{i,t}$, we control for this issue.

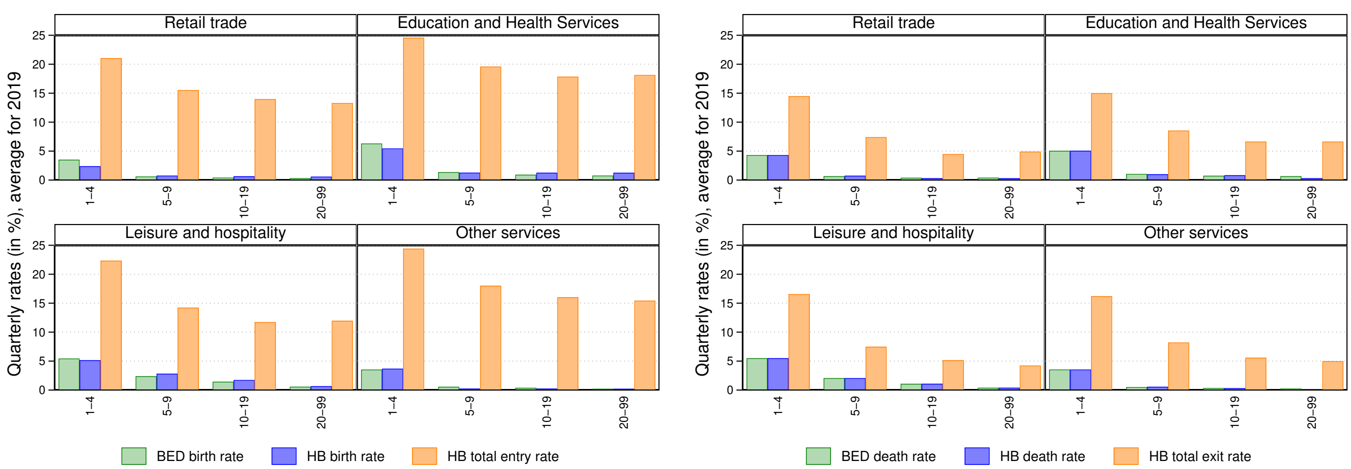
than one. This implies that that the HB sample of entries and exits, respectively the subset of these establishments that we observe in Google / Facebook to estimate closing and opening probabilities, are indeed subject to selection. The adjustment factors correct for this issue. In Section 5, we evaluate the quality of these adjustments when we compare the birth and death rates implied by our methodology *during* the pandemic with BED/BDS counterparts that have recently become available.

4.3 Pre-pandemic business dynamics and employment estimates

Figure 1 reports average quarterly rates of all new entries and permanent exits for 2019 in our HB sample, average quarterly birth and death rates implied by our adjusted estimates of new openings and closings, and the corresponding average quarterly birth and death rates from the BED/BDS benchmark.

Figure 1: Benchmarking with BED/BDS birth and death rates

(a) Birth rates (b) Death rates

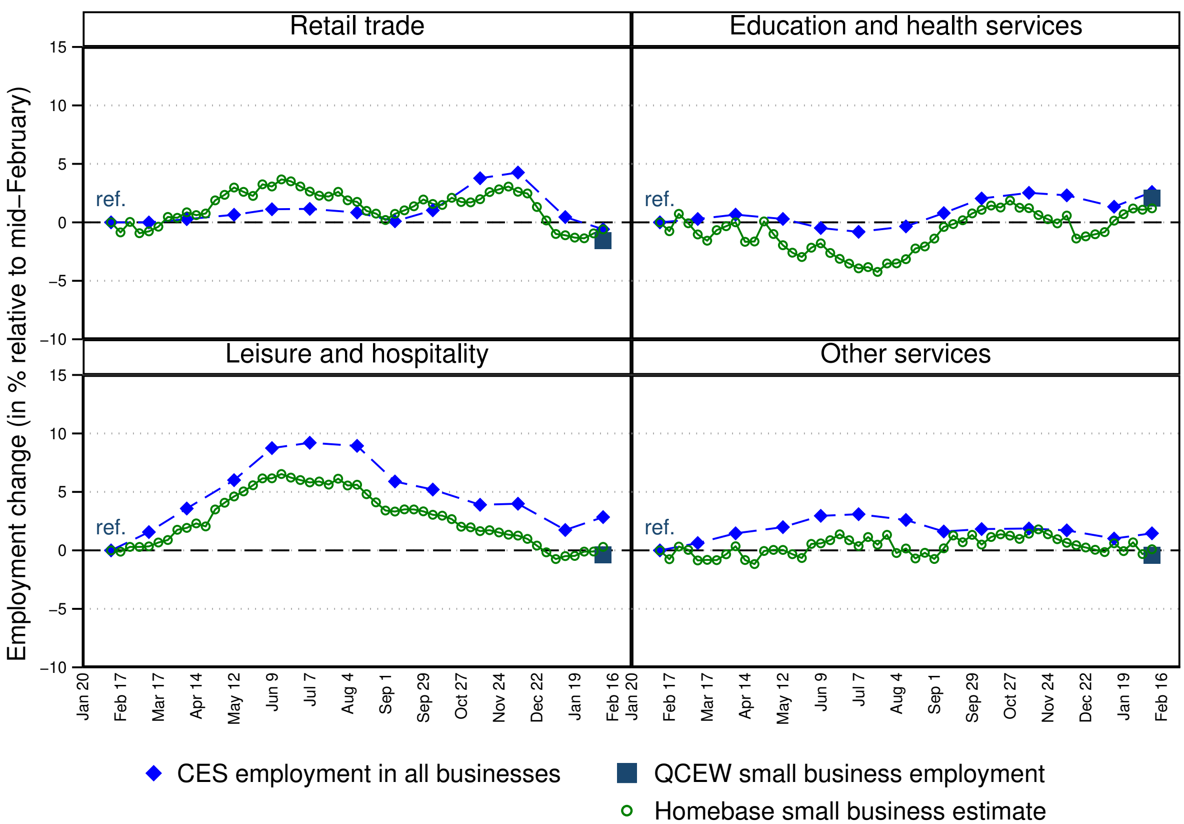


Notes: Quarterly birth and death rates by sector and establishment size class from BED industry data, combined with annual BDS industry-size ratios; corresponding quarterly birth and death rates from HB; and quarterly entry and exit rates from HB. See text for details on the computation.

By construction, the birth and death rates from our HB sample fit the BED/BDS counterparts very closely (the fit is not perfect because, as described above, we pool over some of the sector-size classes). Total new entry and permanent exit rates, in comparison, are much larger, as high as 25% per quarter in the Education and Health sector. This confirms that the HB data is subject to important sample churn: many establishments already operated prior to entry into HB, and many establishments continue to operate after exiting HB. Finally, the figure illustrates the large differences in birth and death rates between the smallest size class and the other size classes included in our sample. Taking into account these differences turns out to be important for the estimation of small business dynamics and employment during the pandemic.

Next, as a quality check of our methodology and to further assess the representativeness of our HB sample, we take the adjusted opening and closing probabilities to estimate employment losses from establishment closings and employment gains from establishment openings, equations (2) and (3), and plug these estimates into equation (1) to compute weekly small business employment for each of the four sectors.

Figure 2: Small business employment change compared to CES all business estimates for 2019



Notes: Employment change by small businesses with less than 50 employees and all businesses in percent of respective employment level during the week of Feb 10 - Feb 16, 2019 for Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81). None of the estimates are seasonally adjusted. The estimates for the weeks of Thanksgiving, Christmas, and New Year are smoothed by using the estimates of adjacent weeks.

Figure 2 reports the resulting time series from mid-February 2019 to mid-February 2020 together with the corresponding year-on-year growth rate for small business employment from the QCEW.[22]

[22]We use mid-February 2019 as the reference week because, as described above, the QCEW publishes employment and establishment counts by industry and establishment size class only for the first quarter of each year, with the numbers pertaining to the month of February. We use these numbers to construct the jump-off point E_0 and the cell weights ω_i for our employment estimator in (1).

For comparison, the figure also shows the monthly CES estimates of total sectoral employment (i.e. employment by businesses *of all size classes* as opposed to just small businesses).[23] Here and below, we refrain from seasonally adjusting employment estimates since usual adjustment procedures would not be appropriate for the type of large changes that the four sectors experienced during the pandemic. See Rinz (2020) for a discussion.

Our HB estimates provide an excellent fit of the year-on-year employment growth rate from the QCEW benchmark. Aside from the selection issues with regards to openings and closings for which we correct with the adjustment factors, the HB sample is therefore representative of small business employment growth during the pre-pandemic period.

The weekly HB estimates also comove with the monthly CES estimates, as should be expected, although there are some seasonal differences. These differences should not come as a surprise nor do they invalidate our approach since our HB estimates pertain to establishments with fewer than 50 employees whereas the CES estimates pertain to total employment by establishments of all size classes. Instead, the relevant comparison is with the small business employment benchmark from the QCEW and in this respect, the fit is close.

5 Small business dynamics and employment during the pandemic

As a first application, we study whether small service-sector businesses have been hit harder by the pandemic than larger businesses. The exercise also allows us to further evaluate the representativeness of our estimates of small business dynamics and employment during the pandemic. Finally, we exploit the richness of the HB data to compute results on average hours worked as well as gross job flows.

5.1 Employment

We start in Table 2 by reporting estimates of the change in small business employment in each of the four sectors considered from mid-February 2020 to mid-April 2020, mid-April 2020 to mid-June 2020, and mid-June 2020 to mid-February 2021. Across the four sectors, small business employment declined dramatically between mid-February and mid-April 2020 as states imposed business closures and stay-at-home orders, combining for a job loss of 14 million or 45% of the 30 million jobs prior to the pandemic. Between mid-April and mid-June, small business employment in all four sectors then regained about 9

[23]The employment estimates from the CES are publicly available only by sector and selected industries, but not by size class.

million or more than two thirds of the initial job loss. Between mid-June 2020 and mid-February 2021, small business employment further recovered although at a considerably lower pace.

Table 2: Small business employment loss and recovery during the pandemic

	Retail Trade	Education & Health	Leisure & Hospitality	Other Services	Total
Employment in mid-February 2020	7,205	8,539	9,714	4,394	29,852
Mid-Feb to mid-April 2020	-3,019	-3,097	-5,235	-2,209	-13,558
in % relative to mid-Feb 2020	*-42%*	*-36%*	*-54%*	*-50%*	*-45%*
Mid-April to end-June 2020	2,365	1,752	3,560	1,401	9,078
in % relative to mid-Feb 2020	*+33%*	*+21%*	*+37%*	*+32%*	*+30%*
Mid-June 2020 to mid-Feb 2021	362	1,059	201	524	2,146
in % relative to mid-Feb 2020	*+5%*	*+12%*	*+2%*	*+12%*	*+7%*

Notes: Employment is expressed in 1,000s of jobs and pertains to establishments with fewer than 50 employees in Retail Trade (NAICS 44-45), Education & Health (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81). None of the estimates are seasonally adjusted. Employment in mid-February 2020 is constructed as the employment estimate for all businesses from the CES times by the ratio of employment in businesses with fewer than 50 workers to employment in all businesses from the QCEW. The other estimates are computed with HB data using the estimator in equation (1).

Figure 3 traces out the week-by-week evolution of these estimates from the mid-February 2020 reference week to late-November 2021 (the end of the current sample) and compares them with the corresponding monthly CES total employment estimates (as described above, the CES does not publish estimates for different size classes). For reference, the figure also shows the mid-February 2020 to mid-February 2021 year-on-year growth rate of small business employment from the QCEW.

The figure visualizes the swift decline in small business employment from mid-March to mid-April 2020 as well as the ensuing rebound by mid-June 2020. Thereafter, the recovery continues at a more gradual pace, with a temporary interruption during Winter 2020-21. By late Spring of 2021, small business employment had recovered almost all of the losses in Education & Health and Other Services, and exceeded its pre-pandemic level in Retail Trade and Leisure & Hospitality, although part of this increase appears to be driven by seasonal factors (compare with Figure 2 above).[24]

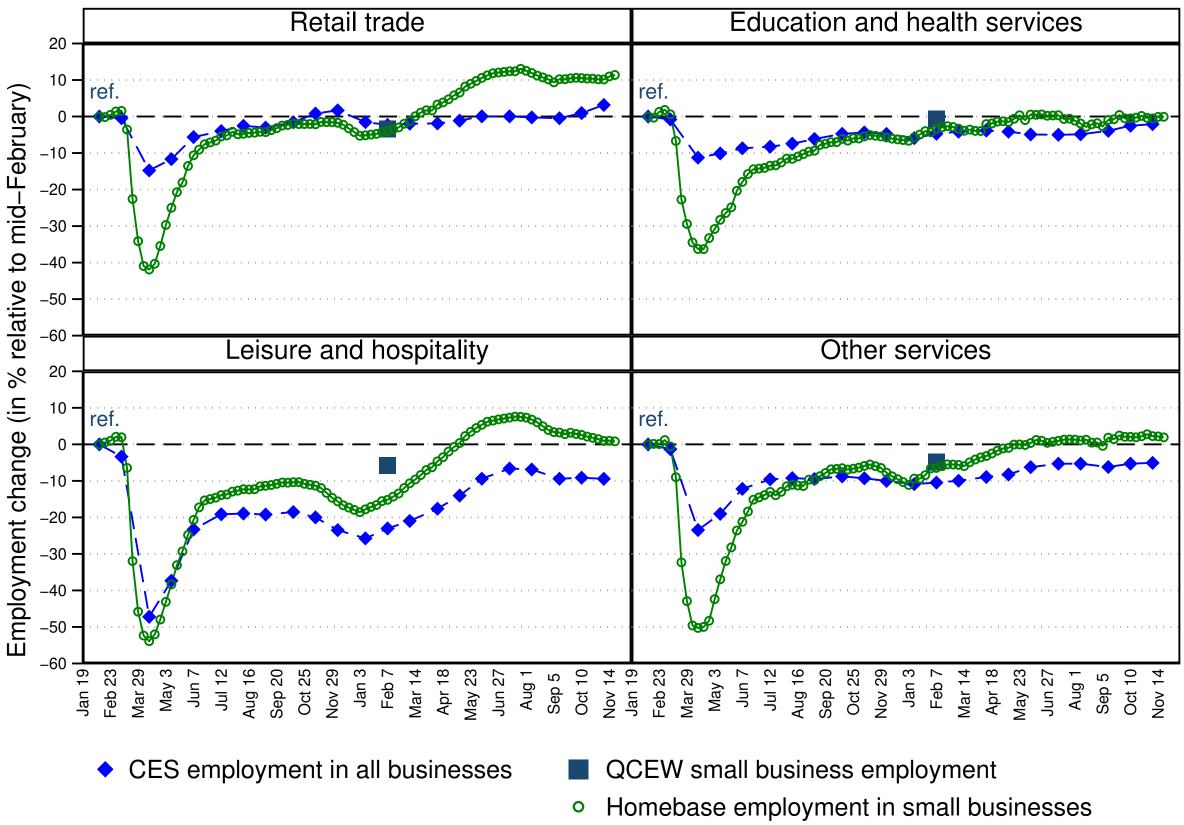
In comparison, the initial decline and subsequent rebound in total employment from the CES is two to three times smaller than for small business employment, with the exception of Leisure & Hospitality where the relative magnitudes are almost as large.[25] From mid-June 2020 onward, small business employment

[24] Interestingly, this rise in small business employment in Retail Trade and Leisure & Hospitality accords well with the increase in (inflation-adjusted) output above their pre-pandemic levels observed in national income accounts data. See https://www.census.gov/econ/currentdata/.

[25] Digging deeper, we find that even in retail subsectors considered essential such as Building Material Dealers (NAICS 444), Food and Beverage Stores (NAICS 445), Gasoline Stations (NAICS 447), or General Merchandise Stores (NAICS 452) where the CES estimates show almost no job loss across all businesses, our HB estimates show large declines in small business employment between mid-February and mid-April, followed by a large rebound. See the online Appendix for details.

is estimated to have regained somewhat more of its pre-pandemic level than total employment, especially in Retail Trade and Leisure & Hospitality where the small business estimate is consistently above the CES counterpart, ending up about 10% higher by the end of the sample in mid-November 2021.

Figure 3: Small business employment change compared to CES all business estimates



Notes: Employment change by small businesses with less than 50 employees and all businesses in percent of respective employment level during the week of Feb 9 - Feb 15, 2020 for Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81). None of the estimates are seasonally adjusted. The estimates for the weeks of Thanksgiving, Christmas, and New Year are smoothed by using the estimates of adjacent weeks.

We highlight that these weekly estimates of small business employment were available in near real-time.[26] In the official data, there are no comparable numbers that are as timely and at such a high frequency. In fact, the only publicly available official product to which we can compare our estimates are the annual small business employment numbers from the QCEW, which are released with a lag of about nine months. As the figure shows, our estimates align closely with the year-on-year changes implied

[26]The estimation of $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t})$ and $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t})$ to calculate job loss from closings and job gains from openings requires four weeks of Facebook data, although preliminary estimates based on extrapolated values of $\hat{p}(\mathcal{C}_{\ell,t}|\text{exit}_{\ell,t})$ and $\hat{p}(\mathcal{O}_{\ell,t}|\text{entry}_{\ell,t})$ could be obtained within only a few days.

by these QCEW numbers, except in Leisure & Hospitality where the QCEW year-on-year change is almost 10% *above* our estimate. Closer inspection of the QCEW suggests that this discrepancy is due to the fact that in each year the QCEW reports a new cross-section of employment by establishment size class. Since many establishments in Leisure & Hospitality were still substantially below their pre-pandemic employment level by mid-February 2021, this means that the count of establishments in the QCEW classified as having fewer than 50 employees increased substantially from mid-February 2020 to mid-February 2021, thereby inflating the year-on-year growth rate of small business employment in that sector. Our HB estimates, in contrast, follows establishments of a given size class over time and therefore does not face this compositional issue.

The results in Figure 3 imply that aside from the initial phase, small business employment in the four service sectors has not been hit harder by the pandemic than employment of larger businesses. The results offer a potentially interesting new perspective for the ongoing debate on whether small businesses are more sensitive to economic shocks than larger businesses. While Moscarini and Postel-Vinay (2012) and Haltiwanger et al. (2018) find that this is generally not the case, Fort et al. (2013) report that small/young business contracted more during the 2008-09 Great Recession. Our high-frequency estimates for the pandemic suggest that this difference in results may in part be about the timing of when in a recession employment is measured. Initially, small businesses may be affected more severely than larger businesses, perhaps due to more severe financial constraints, but smaller businesses may also recover more quickly, perhaps due to their ability of adapting better to a new economic environment.

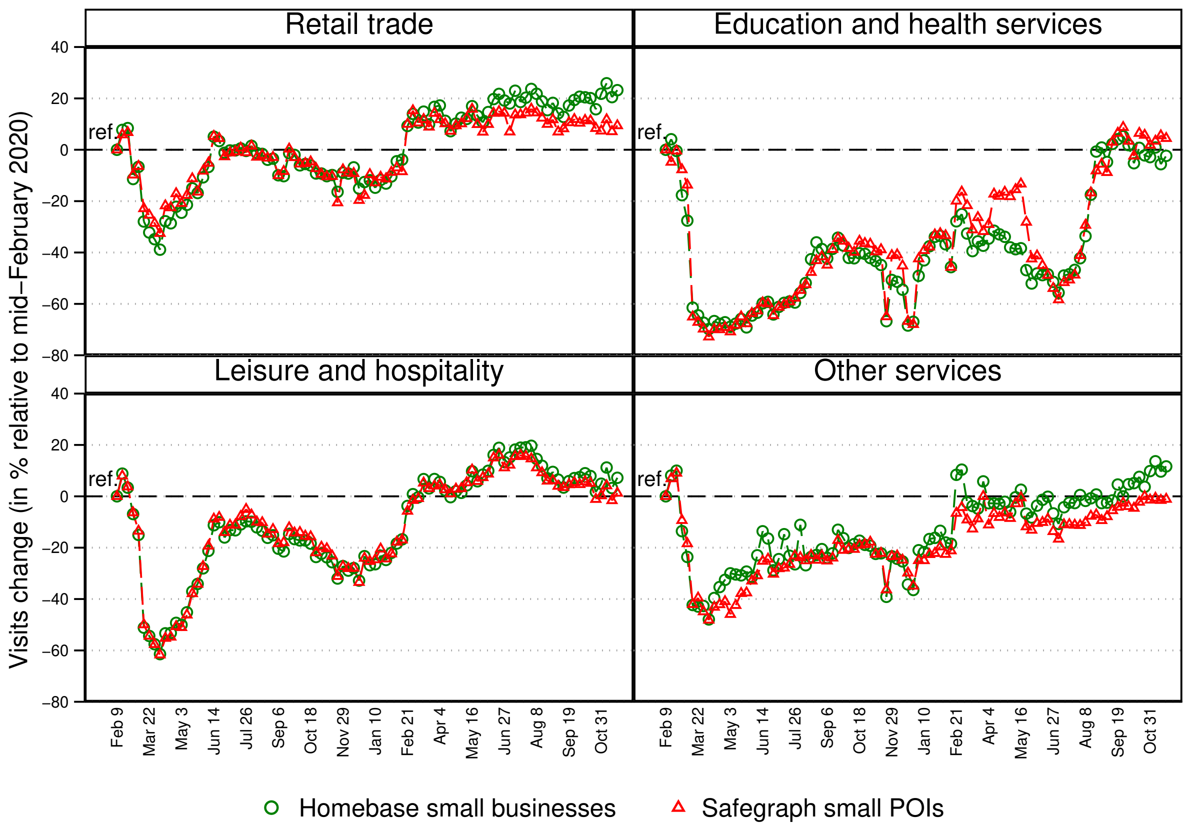
At the same time, we note that the total 14 million loss in small business employment between mid-March and mid-April 2020 and the rebound of about 10 million between mid-April and mid-June 2020 that we estimate across the four sectors are *larger* than the corresponding CES all business estimates (13.5 million and about 6 million, respectively).[27] Unless employment in businesses with 50 employees or more increased from mid-March to mid-April and then declined from mid-April to mid-June – an implausible scenario – this means that either our HB estimates or the CES estimates did not adequately capture small business employment at the onset of the pandemic.

As we have already documented, our HB estimates closely align with year-on-year sectoral growth rates for small business employment from the QCEW. Below, we further show that our methodology to distinguish closings and openings from sample churn implies birth and death rates during the pandemic that closely fit the BED/BDS population counterparts. While these are encouraging indicators of the

[27] For comparison, the headline CES employment estimate for all private sectors declined by 19 million from mid-February to mid-April on a seasonally unadjusted basis.

representativeness of our estimates, it could nevertheless be the case that the HB sample skews towards small businesses that were disproportionally affected during the first few months of the pandemic. Given that there is no publicly available administrative data at sufficiently high frequency to assess this possibility, we pursue a novel approach that exploits the match of our HB establishments with Safegraph POIs to show that this is unlikely to be the case.

Figure 4: Homebase weekly visits compared to Safegraph POIs with less than 50 employees



Notes: Weekly visits change by small businesses with less than 50 employees in Homebase (green circles) vs. Safegraph POIs with less than 50 employees (red triangles) in percent of respective employment level during the week of Feb 9 - Feb 15, 2020 for Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81). None of the estimates are seasonally adjusted. Information on the number of employees at Safegraph POIs come from NetWise employment data.

For many of its POIs, Safegraph reports information on the number of weekly visits and dwell time per visit. In addition, the Safegraph data can be combined with information from NetWise, a company collecting and analyzing vast amounts of data on U.S. businesses, to obtain an annual estimate of employees for each Safegraph POI. We use this information first to cross-check the average number of employees for each matched HB establishment to the corresponding NetWise estimate and find a close correspondence. Second and more importantly, we compare weekly visits of the matched HB establishments to all Safegraph POIs in the four sectors with fewer than 50 employees.[28] Figure 4 shows the results.

The fit is remarkable. Throughout the entire pandemic, weekly visits to matched HB establishments evolve, with a few exceptions, in lockstep with weekly visits to all corresponding Safegraph POIs. As shown in the Appendix, a similarly close fit obtains for median dwell time, share of visits lasting longer than four hours, and weekly visits per visitor. The results indicate that the establishments in our HB sample have on average very similar visits characteristics than the much larger Safegraph sample.[29] Together with the close fit relative to QCEW year-on-year employment changes documented above, this provides in our view compelling evidence that our HB sample is representative of small businesses in the four service sectors considered.

This leaves two likely explanations for the larger drop in small business employment in the beginning of the pandemic according to our HB estimates compared to total employment according to the CES estimates. First, differences in how employment is measured; and second, differences in how business closings and openings are identified. Employment in the CES is measured by the number of workers receiving pay for any part of the pay period that includes the 12th of the month, independent of whether they actually worked, while employment in HB is measured by the number of workers with positive tracked hours plus all untracked workers who used the HB software otherwise in a given week (e.g. owners, managers). So, if some workers who were temporarily furloughed in mid-April still received pay even though they were no longer working, then they were counted in the CES but not in the HB data, which would imply that the CES overestimated employment in mid-April.[30] Second and as discussed in Section 2, the CES estimator includes only a portion of the employment changes from establishments reporting zero employment and does not directly incorporate non-responding establishments. Our HB estimator, in contrast, directly includes the employment effects of business closings and openings. Since, as shown below, almost 40% of all small businesses in our sample are estimated to be closed in mid-April 2020, with about two thirds returning to activity by mid-June 2020, it is conceivable that the CES estimator missed part of the employment effects from this large increase in temporary closings in the beginning of the pandemic.

[28]The resulting Safegraph sample of POIs in the four sectors with fewer than 50 employees and visits data contains almost 1,000,000 unique observations, or about about 22% of the corresponding universe of establishments in the QCEW.

[29]In Figure 4, the Safegraph data refers to POIs that are not associated with a chains of commercial POIs (McDonald's, Starbucks, etc.). Results are almost identical when we include POIs that are associated with a brand.

[30]One could be concerned that our HB employment measure leaves out many non-hourly workers and that these workers were less likely to be furloughed in the beginning of the pandemic. However, when we check in the Current Population Survey (CPS) whether employment of salaried workers declined by more than employment of hourly-paid workers, we find only small differences. So, even if we do not capture all non-hourly workers, it seems unlikely that this would explain the large difference with the CES employment estimates.

Finally, we note that our estimates of small business employment for the first few months of the pandemic are broadly consistent with estimates in Cajner et al. (2020) based on data from ADP, the largest payroll processing company in the U.S., and Dalton et al. (2020) based on CES microdata. As in the CES, ADP's employment concept is pay-based, but different from the CES, Cajner et al. (2020) count all sample exits as closings and abstract from new openings. Similarly, Dalton et al. (2020) count all establishments reporting zero employment as closings and imputes employment for non-reporting establishment from employment changes in observed establishments. Since sample churn and new entry is quantitatively small in the beginning of the pandemic, this shortcut provided a good initial approximation although it would likely be much less accurate later in the pandemic (see below). Focusing on the results by Cajner et al. (2020), their estimates imply that employment of *all* businesses in the four sectors that we consider declined by 20.2 million between mid-February and late April. Furthermore, they report that employment in establishments with less than 50 employees across *all* private sectors of the U.S. economy declined by about twice as much between March and April 2020 as employment for larger establishments, but by the end of June 2020 had recovered as much as larger establishments. Considering that businesses with fewer than 50 employees accounted for almost half of employment in the four sectors prior to the pandemic and that ADP employment is pay-based, these numbers appear quite close to our estimated decline in small business employment of 14 million and subsequent rebound of about 10 million during the same time period.

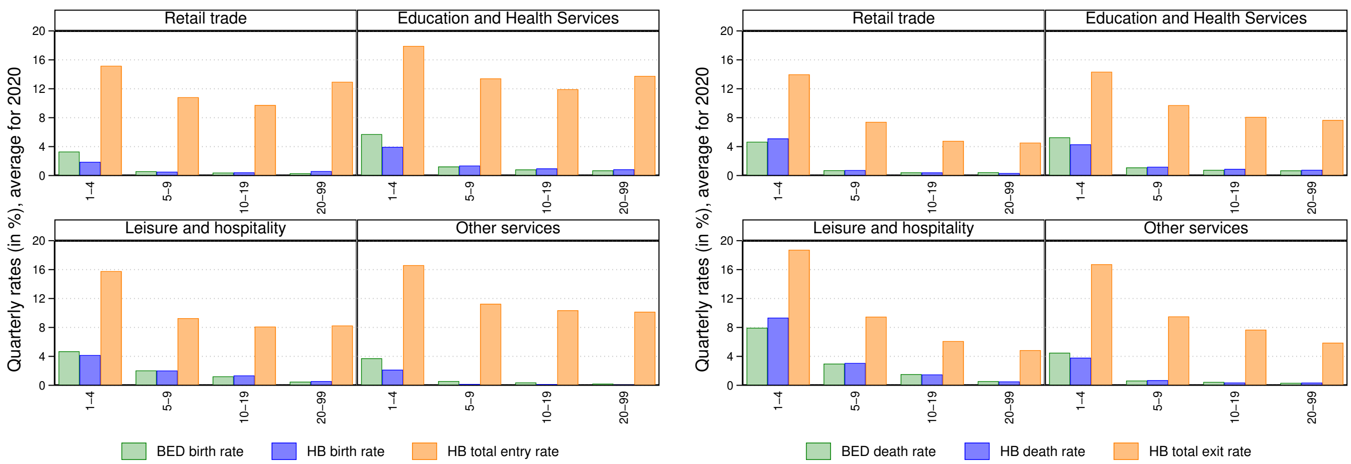
5.2 The crucial role of small business closings and openings

The preceding discussion highlights the importance of accurately taking into account small business openings and closings. To assess the quality of our methodology *during the pandemic*, Figure 5 compares average quarterly birth and death rates by sector and size class for 2020 as implied by our adjusted estimates of new openings and closings to the corresponding BED/BDS birth rates that have recently become available.[31] To illustrate the importance of sample churn, the figure also shows average quarterly rates of all new entries and permanent exits.

[31] For the first quarter of 2020, the BED shows a large increase in births in Education & Health that is likely spurious, due to a technical issue with the annual revision. Also see Decker and Haltiwanger (2022). For this sector, we therefore do not include the first quarter birth rate for the average BED/BDS birth rate calculation.

Figure 5: Comparison to BED/BDS birth and death rates

(a) Birth rates (b) Death rates



Notes: Quarterly birth and death rates by sector and establishment size class from BED industry data, combined with annual BDS industry-size ratios; corresponding quarterly birth and death rates from HB; and quarterly entry and exit rates from HB. See text for details on the computation.

The figure shows a close match between HB birth and death rates and the BED/BDS counterparts, indicating that our methodology of using information from Google / Facebook to identify new openings and permanent closings in real-time and adjusting for selection with *pre-pandemic* information works well.[32] Given the large unexpected shock that the pandemic represents, we think this result is quite remarkable. The figure also shows that, similar to 2019, total entry and exit rates are several times larger than new openings and permanent closings. Correcting for sample churn is therefore as important during the pandemic as prior to the pandemic.

To further study small business dynamics during the pandemic, we compute weekly rates of new closings, reopenings, total closings, and new openings since the beginning of the pandemic and compare them to the same time period one year earlier. Specifically, we compute

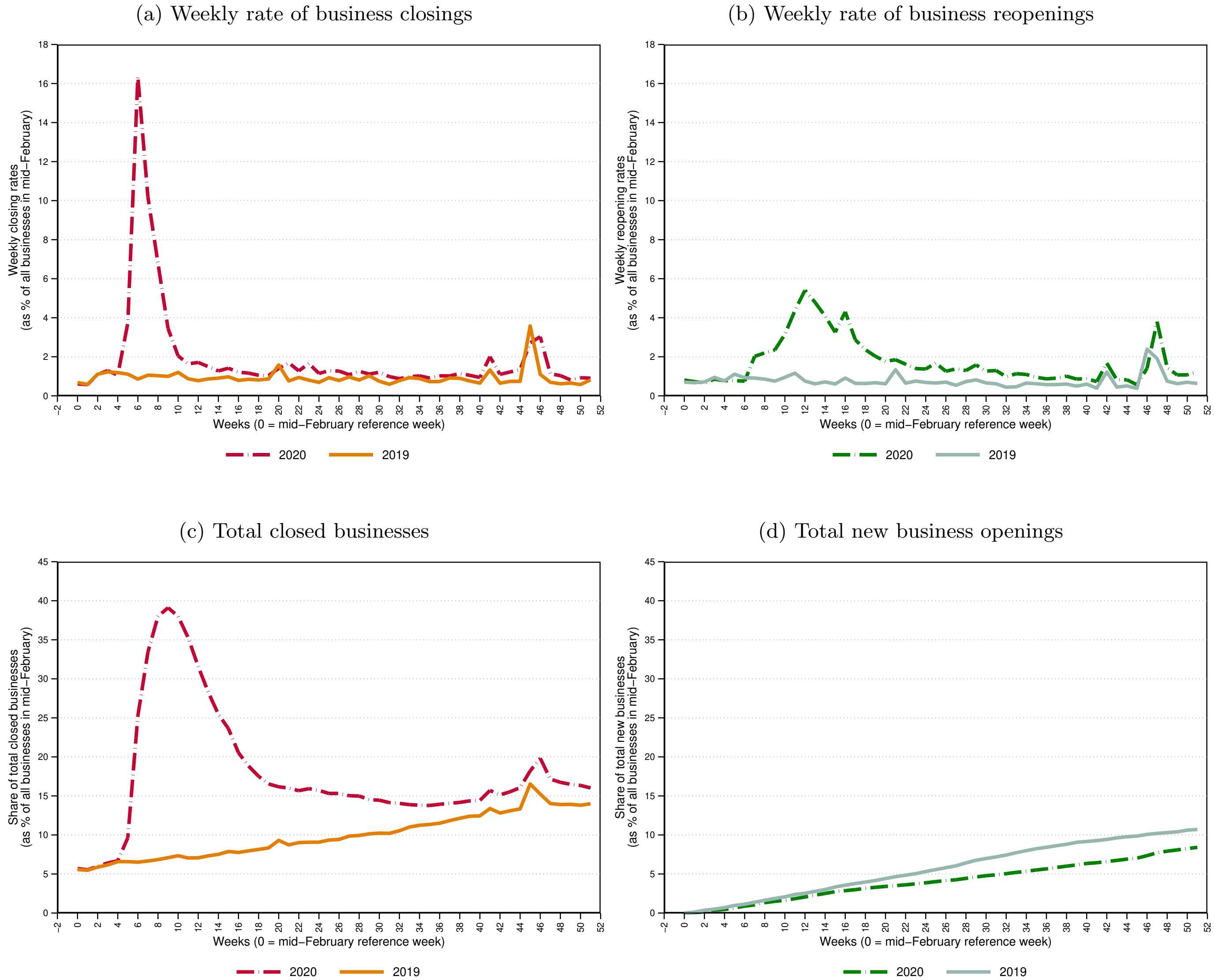
$$rate(\mathcal{I}_t) = \frac{\sum_i \omega_i \hat{n}_{i,t}^{\mathcal{I}_{i,t}}}{\sum_i \omega_i \left(\hat{n}_{i,0}^{\mathcal{A}_{i,1}} + \hat{n}_{i,0}^{\mathcal{C}_{i,1}} \right)}, \tag{4}$$

where $\hat{n}_{i,t}^{\mathcal{I}_{i,t}}$ denotes the count of establishments in industry-size-region cell i that closed either temporarily or permanently in week t ($\mathcal{I}_{i,t} = \mathcal{C}_{i,t}$), reopened in week t ($\mathcal{I}_{i,t} = \mathcal{R}_{i,t}$), or newly opened in week t ($\mathcal{I}_{i,t} = \mathcal{B}_{i,t}$), with $\mathcal{O}_{i,t} = \mathcal{R}_{i,t} \cup \mathcal{B}_{i,t}$ by definition; and $\hat{n}_{i,0}^{\mathcal{A}_{i,1}} + \hat{n}_{i,0}^{\mathcal{C}_{i,1}}$ denotes the count of active establishments

[32] It should also be noted that the BED/BDS benchmarks are computed under the assumption that the relative birth and death rates by size class taken from the 2019 BDS data remain constant during the pandemic. This assumption is unlikely to hold exactly and so, the BED/BDS benchmarks are themselves only an approximation.

in the reference week.[33]

Figure 6: Small business closings, reopenings, and new openings



Notes: Rates of closings, reopenings, total closings, and total new openings of small businesses with less than 50 employees in Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81). All rates are computed as a % of the total count of active businesses in mid-February. Week 0 denotes the mid-February reference week.

As shown in panels (a) and (b) of Figure 6, the weekly rate of small business closings across the four sectors considered spikes to 16% in the week of March 22-28, 2020 (week 6 after the mid-February reference week) and then sharply declines to about 2% by mid-April (week 10) before further declining to

[33]We define these rates relative to the count of active establishments in the reference week as opposed to the count of active establishments around week t because the count of active establishments varies dramatically in the beginning of the pandemic.

just above the pre-pandemic average of about 1% per week.[34] Concurrent with the decline in the rate of business closings in April of 2020, reopenings start to increase, reaching about 5% per week in early May before gradually declining back to the 1.5-2% range between July and September and then the 1-1.5% range thereafter, just slightly above the pre-pandemic rate.

Panel (c) displays the cumulative effect of these closings and reopenings on the rate of total closed businesses relative to active businesses in the mid-February reference week. Note first that the rate of total closed businesses averages about 6% in both mid-February 2019 and mid-February 2020, indicating that a substantial fraction of businesses are temporarily closed at any point in time (also see Table 1). From mid-March 2020 onward, total closings rise steeply and peak at 39% in mid-April. Thereafter, the cumulative closing rate declines, steeply initially as reopenings rise and then more gradually to a low of about 14% by November before rising to about 16% by mid-February 2021.[35] This suggests that only about one third of all closings in mid-March are permanent.[36] Moreover, and perhaps surprisingly, the cumulative rate of closings one year after the start of the pandemic is only about 2 percentage points higher than the cumulative closing rate from mid-February 2019 to mid-February 2020. This implies, perhaps surprisingly but consistent with concurrent analysis by Crane et al. (2022) based on alternative measures of business closures as well as Decker and Haltiwanger (2022) and Fairlie et al. (2022) based on more recently released administrative data, that the pandemic did not lead to more permanent small business closings.

Panel (d), finally, shows total new business openings relative to total active businesses in mid-February. This rate rises gradually throughout the year, even during the worst of the pandemic in March and April. Compared to 2019, the pace of new openings is clearly lower during the Spring and Summer of 2020 but then picks up somewhat in Fall and Winter, finishing at about 8% in mid-February 2021. This is only about 2 percentage points lower than the rate of new business openings a year earlier, implying that the pandemic only exerted a modest negative effect on new business openings.

At first sight, the lower rate of new business openings contrasts with recent evidence from the U.S. Census Bureau that shows record rates of new applications for likely employer business relative to pre-pandemic levels, in particular in Nonstore Retail (NAICS 454) and Personal & Laundry Services (NAICS 812), but also in Food Services & Drinking Places (NAICS 722) (see Haltiwanger, 2021; also see Fazio et al., 2021). However, there are a number of potential explanations for this difference. First, based

[34]The temporary upticks in closing rates in weeks 41 and 46 capture the weeks of Thanksgiving and Christmas.

[35]These cumulative closing rates are less than half of what Crane et al. (2022) report based on HB data. The reason for this difference is that their study treats all exits as closings.

[36]The majority of establishments that closed for more than 10 weeks remain closed.

on historical patterns, it typically takes four to eight quarters from business registration to employment of workers. Second, as we show below, while we see substantial employment gains from new business openings in the Leisure & Hospitality sector (of which Food Services and Drinking Places is a large part), new business openings play a more modest role for Retail Trade and Other Services. Since we exclude Nonstore Retail and many of the new small businesses in Personal & Laundry Services may not have a physical store location (and are therefore not part of our HB sample since they do not appear in Safegraph), this implies that our estimates may represent a lower bound of the recovery in small business employment in Retail Trade and Other Services.

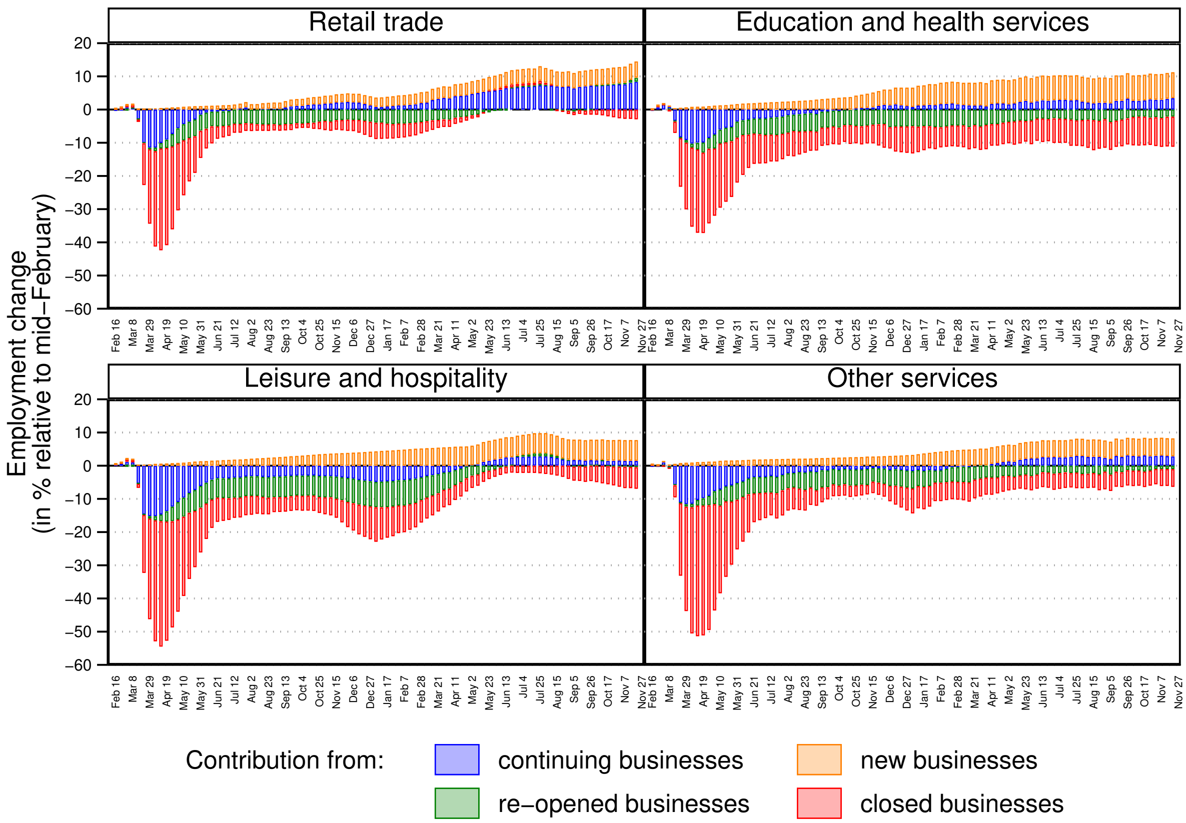
To quantify the role of closings and openings for small business employment, we decompose the employment change for each sector into the contributions from businesses that continue to operate from mid-February until at least week t (and possibly longer), businesses that closed at some point after the mid-February reference week but reopened by week t, businesses that operated in mid-February but are closed in week t (temporarily or permanently), and businesses that newly opened between mid-February and week t.[37]

As Figure 7 shows, business closings account for 70% or more of the initial employment decline from mid-March to mid-April across the four sectors (red bars) while job losses by continuing businesses account for rest (blue bars). Reopenings of closed businesses drive most of the rebound in employment between mid-April and mid-June (smaller red bars), even though the reopened businesses operate at lower employment than in mid-February (green bars). Finally, the recovery from mid-June onward is driven primarily by the combination of job gains by continuing businesses (smaller negative and positive blue bars) and new businesses (yellow bars). Overall, these new openings add almost 1.5 million new jobs by the end of the sample.

The decomposition also reveals interesting differences across sectors. Large job losses from closings persist through the end of the sample in Education & Health but reduce to almost zero in Retail Trade. At the same time, Retail Trade experiences substantial job gains by continuing businesses. Finally, job gains from new openings are more important in Leisure & Hospitality and Education & Health. These differences suggest that the pandemic led to varying degrees of restructuring within the different sectors.

[37] The Appendix provides details on the decomposition. The employment losses from closed business nets out gains from establishments that were active in HB prior to the mid-February reference week, temporarily closed in the reference week, and then reopened at some point thereafter (e.g. seasonal businesses; see Table 1). By netting out these gains, the contribution from closings represents the employment losses over and above the usual employment losses from business that temporarily close. See below for further discussion.

Figure 7: Contribution of closings, reopenings, and new openings to small business employment change



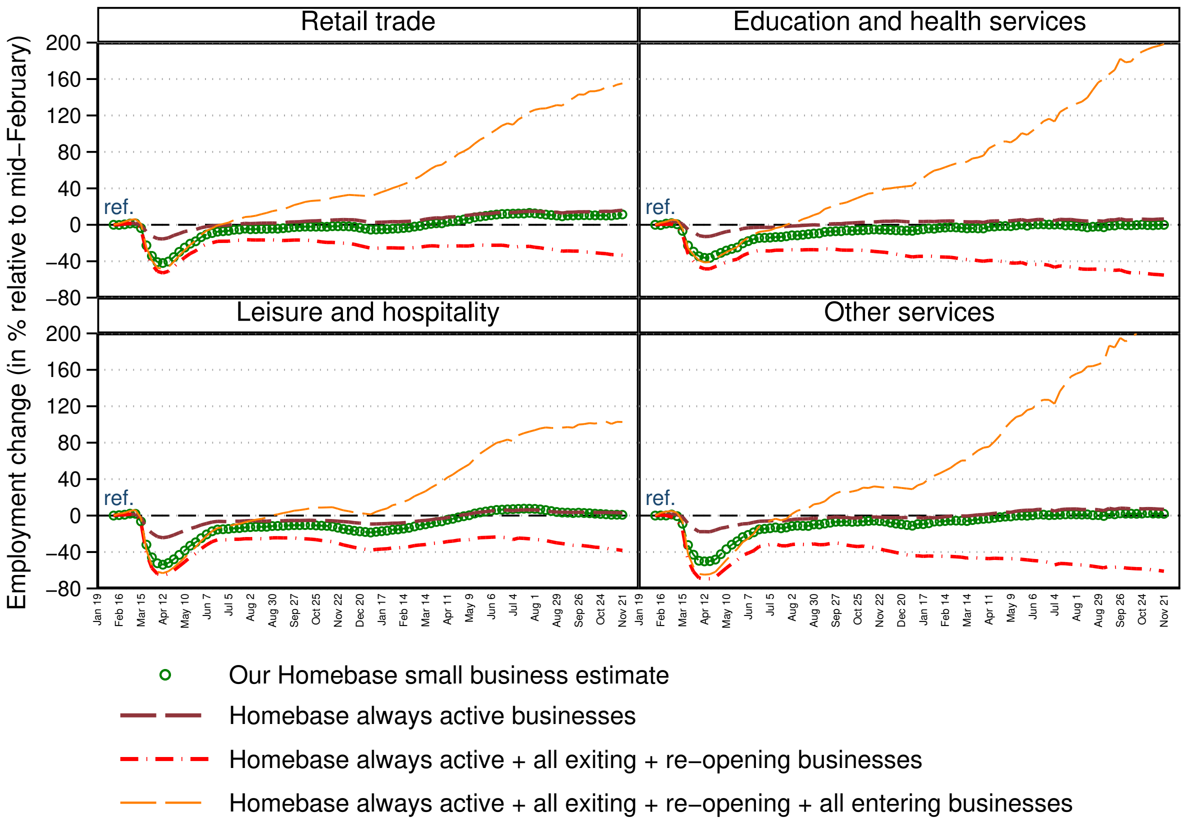
Notes: Contribution to percent employment change relative to mid-February 2020 in Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81) by businesses that continued operating from mid-February until at least week t (blue bars), businesses that closed at some point after mid-February 2020 and but reopened by week t (green bars), employment changes from businesses that operated in mid-February 2020 but are closed in week t (red bars), and employment changes from new businesses that opened between mid-February 2020 and week t (orange bars). The estimates for the weeks of Thanksgiving, Christmas, and New Year are smoothed by using the estimates of adjacent weeks.

5.3 The importance of controlling for sample churn revisited

To further illustrate the importance of distinguishing closings and openings from sample churn, Figure 8 reports different counterfactual employment estimates for the pandemic (similar counterfactual estimates for the pre-pandemic period are reported in the Appendix) and compares them with our baseline small business estimates (green circled line).

The brown short-dashed line shows the HB estimate if we abstracted completely from entry and exit and used only the set of establishments that are continuously active in HB from the beginning through the end of the sample (i.e., $\widehat{E}_t = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \hat{e}_{i,t}^{\mathcal{A}_{i,t}}}{\sum_i \omega_i \hat{e}_{i,t-1}^{\mathcal{A}_{i,t}}}$, where $\mathcal{A}_{i,t}$ denotes continuously active establishments). This estimate would miss much of the large decline and subsequent rebound of small business employment in the initial phase of the pandemic.

Figure 8: Comparison with counterfactual employment estimators



Notes: Estimated employment change in % relative to mid-February 2020 of small businesses with less than 50 employees in Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81) according to different estimation methods (see text). The estimates are constructed based on February 2020 CES employment estimates (week of Feb 9 – Feb 15) and QCEW shares of small business employment for the first quarter of 2020. The estimates for the weeks of Thanksgiving, Christmas, and New Year are smoothed by using the estimates of adjacent weeks.

The red dashed-dotted line reports what happens if we treated all exits as either temporary or permanent closings (i.e., $\widehat{E}_t = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{R}_{i,t}}\right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\text{exit}_{i,t}}\right)}$, where $\text{exit}_{i,t}$ denotes the set of all exiting establishments in week t and $\mathcal{R}_{i,t}$ the set of all returning establishments in week t). Since a substantial fraction of establishments that exit HB do not close, this estimate declines even more than our baseline estimate in the beginning of the pandemic and recovers much less thereafter. In fact, from 2021 onward, this estimate declines gradually as cumulative sample churn becomes increasingly important.

The orange dashed line, finally, adds all entries and treats them as new openings (i.e., $\widehat{E}_t = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,T}} + \hat{e}_{i,t}^{\text{entry}_{i,t}}\right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,T}} + \hat{e}_{i,t-1}^{\text{exit}_{i,t}}\right)}$, where $\text{entry}_{i,t}$ denotes the set of all entering establishments in week t, including the returning establishments). The resulting estimate shows a dramatic increase in small business employment through the end of the sample, far outweighing the negative effect of treating all exits as closings. This reflects the fact that even during the pandemic, HB managed to substantially expand its client base.

The different counterfactuals offer an interesting perspective relative to other research estimating the impact of the pandemic on small business employment with HB and other private-sector establishment level datasets (e.g. Bartik et al., 2020; Cajner et al., 2020; Chetty et al., 2020; or Dalton et al., 2020 among others). These studies do not systematically distinguish business closings and new openings from sample churn and instead used either the set of continuously active businesses or the set of continuously active businesses plus all exiting and reopening businesses or some combination thereof to estimate employment. While the consequences of doing so with other datasets may not be as extreme as with the HB data, the results are nevertheless affected by this choice. As such, our results offer a cautionary tale about the use of opportunity samples to estimate small business employment series without careful incorporation of employment gains from establishment openings, respectively employment losses from closings.

5.4 Average weekly hours

To provide further insights on the impact of the pandemic on workers employed by small businesses, we use the HB data to report changes in average weekly hours (AWH) worked. To do so, we start with the CES estimate from February 2020, $\widehat{AWH}_0$, and then use our HB data to estimate

$$\widehat{AWH}_t = \widehat{AWH}_{t-1} \times \frac{\left(\sum_i \omega_i \widehat{wh}_{i,t}\right) / \left(\sum_i \omega_i \hat{e}_{i,t}\right)}{\left(\sum_i \omega_i \widehat{wh}_{i,t-1}\right) / \left(\sum_i \omega_i \hat{e}_{i,t-1}\right)}, \tag{5}$$

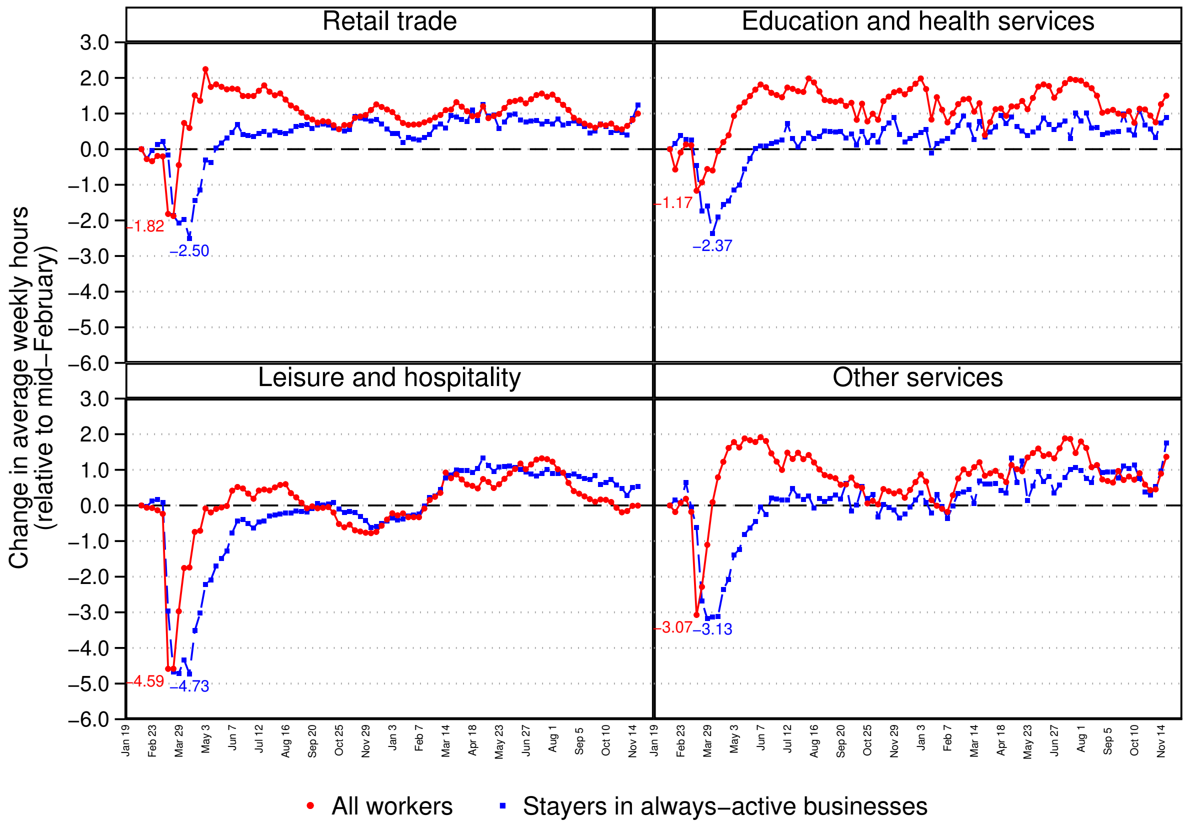
where $\widehat{wh}_{i,t}$ denotes estimated total weekly hours worked and $\hat{e}_{it}$ denotes estimated employment at establishments in industry-size-region cell i in week t. This estimation of AWH is different from the "link-and-taper technique" used to construct AWH in the CES, which adjusts the current estimate towards the previous estimate so as to keep it close to the overall sample average over time. The CES estimate may therefore not capture large changes in actual AWH that occur in times of economic disruptions, whereas our estimate does because it is based on current information only.[38]

Figure 9 shows two estimates of AWH, one for all workers employed in week t, and one for all job stayers who remained employed continuously in establishments that are active throughout the sample.

[38]The link-and-taper estimate used in the CES can be expressed as $\widehat{AWH}_t = 0.9\left(\widehat{AWH}_{t-1} - \widehat{awh}_{t-1}\right) + \widehat{awh}_t$, where $\widehat{AWH}_t$ is the official estimate and $\widehat{awh}_t = \left(\sum_i \omega_i wh_{i,t}\right) / \left(\sum_i \omega_i e_{i,t}\right)$. If $\widehat{AWH}_{t-1} > \widehat{awh}_{t-1}$ in the previous month, then the current month official estimate will be raised relative to actual data, and vice versa if $\widehat{AWH}_{t-1} < \widehat{awh}_{t-1}$. The CES makes a slight adjustment to this estimator to account for atypical reports although it is unclear what makes a report atypical.

While the former measure is affected by compositional change, the latter is not since it consists by definition of a balanced panel of workers.

Figure 9: Average Weekly Hours of Small Business Employees



Notes: Changes in average weekly hours of employees in small businesses with less than 50 employees in Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81) relative to the February 2020 CES estimate (week of Feb 9 – Feb 15). The solid red line shows the change in average weekly hours of all workers employed in all small businesses. The blue shows the change in average weekly hours of job stayers in businesses that remain active throughout the sample.

Across all four sectors, AWH declines sharply in March 2020 but then recovers quickly and exceeds the pre-pandemic level somewhat through the end of the sample. The sharp decline in AWH in the beginning of the pandemic is further evidence of the sudden impact of the pandemic on service-sector jobs. The larger drop as well as the smaller increase in AWH of job stayers relative to the all workers measure is due to the compositional change: the workers laid off or furloughed in March worked on average fewer hours than job stayers. As employment recovered a large part of the loss, this gap then mostly closed.

Remarkably, AWH of job stayers recovers fully within a few months of the start of the pandemic. This suggests that the labor market during the pandemic has not been characterized by a large increase in involuntary part-time employment as has been the case during previous recessions (e.g. Borowczyk-Martins and Lalé, 2019). This may be due to the particular nature of the pandemic and its outsize effect on in-person service sector jobs where part-time is unlikely to be as feasible as in, say, manufacturing that suffered more heavily during previous recessions. Alternatively, the health risks implied by the pandemic and the large extensions of unemployment insurance in response may have reduced the incentives for part-time work. Examining these questions is an interesting topic for future research.

5.5 Job separations, new hires, recalls, and excess turnover

The linked worker-establishment structure of the HB data also allows us to provide a detailed account of gross job flows. We decompose weekly employment changes into job separations, new hires, and recalls. As is common in the literature, recalls are defined as workers who are employed with the establishment at some point in the past, disappear for at least one time period (one week in our case), and then reappear as employees in the same establishment.

Figure 10 reports the different weekly gross flows as a rate of average employment in the same week and the preceding week (as for the previous results, all weighted by industry-size-region cells). As panels (a) and (b) show, the job separation rate spikes the week of March 22-28 (week 6), the same week as business closures spike, while the new hire rate declines. Both rates then return to their pre-pandemic average by mid-June and remain essentially the same as one year earlier.

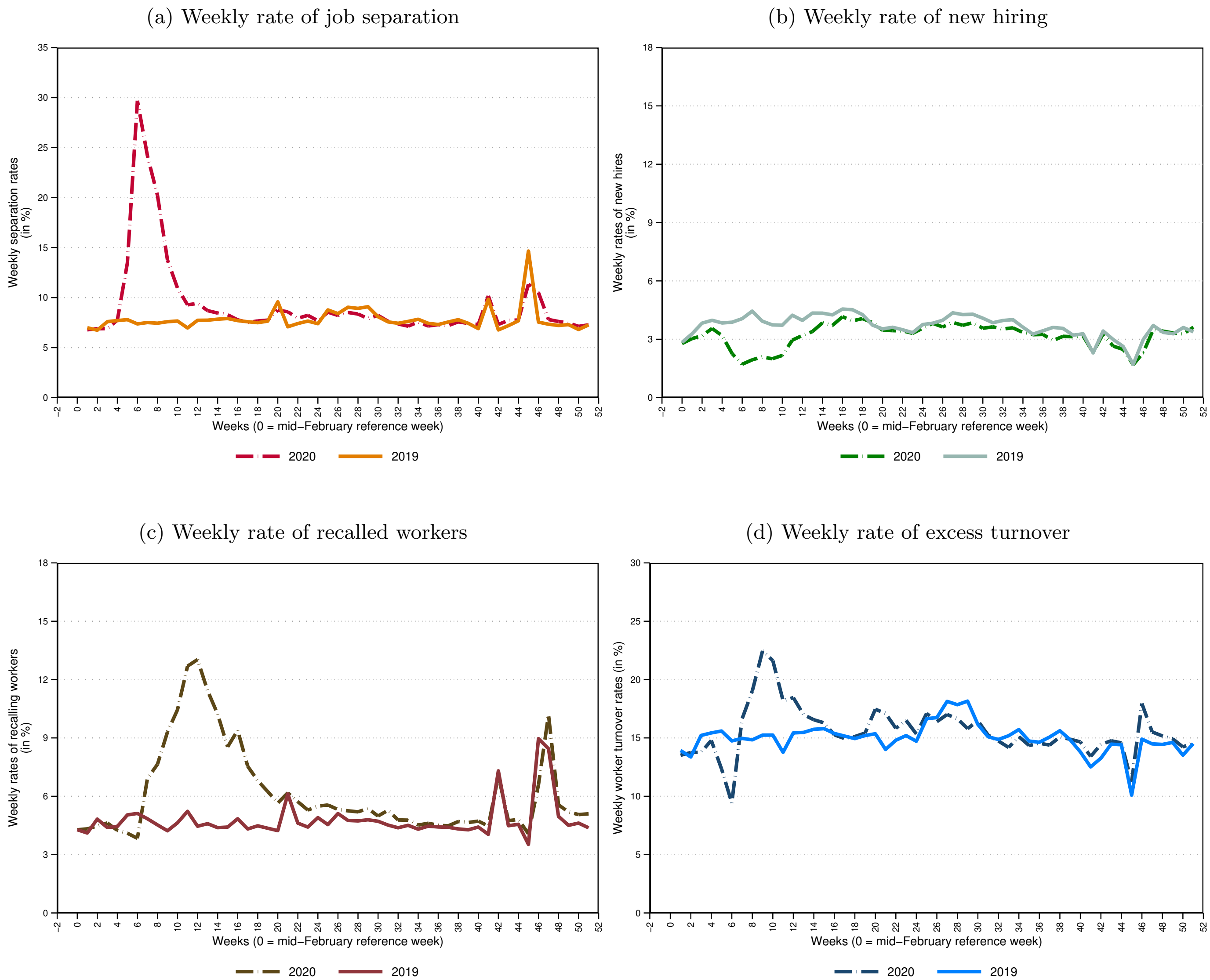
As shown in panel (c), the recall rate of workers previously employed in the same establishment increases substantially in the weeks following the initial spike in separations, peaking the week of May 3-9 (week 12). The recall rate then declines steadily through the week of June 28 - July 4 (week 20) and thereafter remains slightly elevated through the end of summer before essentially returning to the corresponding 2019 value (the spikes around week 42 and 47 are due to the Thanksgiving and Christmas holidays).[39]

Panel (d), finally, shows the excess turnover rate, which is computed as the difference between the sum of separations rate, new hiring rate, and recall rate minus (the absolute value of) net employment growth. The excess turnover rate drops briefly in the beginning of the pandemic as new hiring and recalls decline and then jumps up as recalls jump up while some businesses still show excess job separations.

[39] It is interesting to compare these recall numbers to recent results on recalls in the literature. In particular Fujita and Moscarini (2017) document based on monthly household survey data that on average about 40% of workers return to their previous employment after a jobless spell. Our estimates imply that the corresponding recall rate, measured as recalls divided by the total of recalls and new hires, averages about 55% for 2019 and rises as high as 85% in mid-April 2020. The higher average for 2019 is primarily due to time aggregation in monthly data (we observe non-trivial non-employment spells lasting less than one month with subsequent recall).

After mid-June, excess turnover averages about the same rate as one year earlier.

Figure 10: Job separations, new hires, recalls, and excess turnover in small businesses



Notes: Weekly rates of job separation, new hires, recalls, and excess turnover for small businesses with less than 50 employees in Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81). All rates are computed as a percent of average employment in the same week and the preceding week.

The results indicate that the rebound in small business employment following the sharp decline in the beginning of the pandemic is driven primarily by recalls of temporarily furloughed workers as opposed to new hires, which is in line with other estimates (e.g. Ganong et al., 2021). This is quite different from previous downturns (e.g. the Great Recession) where a larger share of separations was permanent and the recovery was more sluggish due to persistently lower new hiring rates. Furthermore, the quick return in the excess turnover rate to its 2019 average suggests that, at least within the four in-person service sectors considered, the pandemic has so far not led to major reallocations of labor.

6 Effects of the Paycheck Protection Program

As a second application, we use the HB data to investigate the extent to which small business dynamics and employment was affected by the Paycheck Protection Program (PPP), which provided loans to small businesses during the first months of pandemic. The program has been the subject of much controversy and intense research. The novelty of our investigation is that we exploit the high-frequency and detailed geographic dimension of the HB data to differentiate the effects of variations in timely access to PPP loans from the many other changes that occurred in the first months of the pandemic. As before, properly distinguishing the effects of business closings and openings from sample churn turns out to be central for the results.

6.1 Delayed access to PPP loans

The 2020 CARES Act that was signed into law on March 27, 2020 appropriated $349 billion in PPP loans to support firms with fewer than 500 employees prior to the pandemic.[40] To allow broad access, many of the usual eligibility criteria to access government loan programs were waived and the loans came with very favorable terms: qualifying businesses could apply for 2.5 times the average total monthly payroll for each employee up to a maximum of $10 million, and the loans had a duration of two years at a 1% annual interest rate but were forgivable if the business spent at least 75% on payroll within 8 weeks of loan disbursement.[41]

While the Small Business Administration (SBA) was responsible for oversight, firms applied for the loans through local lenders and the first loans were approved on April 3. The demand for loans was so overwhelming that by April 16, the appropriated funds were depleted. In response and after considerable uncertainty, Congress voted on an additional $321 billion in PPP funding that the President signed into law on April 24. Banks started issuing new loans on April 27 and demand spiked immediately, with 60% of the additional funds allocated within two weeks of reopening of the program. Thereafter, loan demand declined substantially and PPP stopped taking new applications on August 8, with almost $150 billion

[40] For multi-establishment firms in Accommodation & Food Services (NAICS 72), the 500 employee threshold applied to establishments within certain limits.

[41] Businesses also had to maintain or restore employee counts and pay for loans to be forgivable. On June 5, 2020 Congress relaxed the conditions for loan forgiveness, lowering the threshold on PPP funds used for payroll from 75% to 60% and increasing the number of weeks to use the funds from 8 to 24. See https://www.sba.gov/funding-programs/loans/covid-19-relief-options/paycheck-protection-program for details.

in unused funds remaining.[42]

As documented in detail by Bartik et al. (2021), Doniger and Kay (2021), and Granja et al. (2020) among others, the first round of PPP was subject to large geographic disparities in loan allocations, likely reflecting differences in the ability and willingness of local banks to process and approve the large initial influx of loan applications. Funds did not necessarily flow to areas of the country where the initial economic effects from the pandemic were largest but were instead driven by the local presence of the different lenders. In addition, the first loans were unusually large, made to relatively larger businesses. Hence, many of the smallest businesses – the ones that are the focus of our study – were subject to delayed access to PPP loans during the beginning of the pandemic, and the extent of this delay depended in large part on the quasi-random presence of different banks across localities.

6.2 Research design

We exploit the geographic variation in initial loan access to evaluate the effects of PPP for small business activity. Similar to Doniger and Kay (2021) we measure delayed access to PPP loans by the share of loans issued between April 26 and May 2 (the week when additional PPP funding became available) relative to the total amount of loans issued between April 12 and May 2 (the week when initial PPP funding ran out to the week when additional PPP funding became available); i.e. $\text{share PPP delayed}_c = \frac{(\text{loans April 26-May 2})_c}{(\text{loans April 12-May 2})_c}$, where c denotes the county of the businesses receiving the loans.[43] As Doniger and Kay (2021) argue, focusing on a relatively narrow window around the temporary exhaustion is important to avoid selection issues associated with the first few weeks of the program.

We construct $\text{share PPP delayed}_c$ using data on all PPP loans from the SBA. The loans made during these weeks account for about one third of all loans and for an even larger share of loans made to businesses with fewer than 50 employees. As shown in the Appendix, the variation in $\text{share PPP delayed}_c$ across counties is wide, with a median of 40% and a 10-to-90-percentile range of $[26\%, 60\%]$.[44]

[42] In December 2020, Congress voted for and the President signed into law a third round of PPP consisting of an additional \$285 billion in funding and new eligibility rules. Loans started in mid-January 2021 and the program ran through the end of May 2021. This third round is not the focus of our investigation.

[43] The weeks in our estimation run from Sunday to Saturday. April 12, 19 and 26 are Sundays. Doniger and Kay use a narrower 2-day window around the temporary exhaustion of PPP loans to measure the share of delayed PPP loans, and they compute the measure at the broader CBSA geographic level. Our estimates are robust to using their narrower time window and the broader CBSA level.

[44] In the regression, we use a subset of 1,956 counties for which we have reliable HB data (out of 3,143 counties for which we have PPP data). The distribution of $\text{share PPP delayed}_c$ for this subset of counties is almost identical to the distribution for the full set of counties.

We use the share PPP delayed$_c$ measure to estimate the following county-level regression

$$y_{c,t} = \sum_{t=0}^{57} \alpha_t \left(\mathbb{1}\left\{\text{week} = t\right\} \times \text{share PPP delayed}_c \right) + \mathbf{X'_{c,t}}\boldsymbol{\gamma} + \phi_t + \mu_c + \varepsilon_{c,t} \tag{6}$$

where $y_{c,t}$ is either the percent deviation of employment across establishments in county c in week t relative to its employment in the first week of 2020 ($t = 0$); the fraction of establishment in county c being closed in week t; or the fraction of establishments in county c being newly opened in week t. The vector $\mathbf{X_{c,t}}$ contains a vector of county-specific controls measuring weekly COVID cases and deaths, non-pharmaceutical interventions (NPIs), school closures, weather, as well as week fixed effects interacted with average county household income prior to the pandemic.[45] Finally, ϕ_t is a week fixed effect capturing time variations in average $y_{c,t}$; μ_c is a county fixed effect controlling for unobserved average differences across counties; and $\varepsilon_{c,t}$ is the error term. All regressions are weighted by county level employment prior to the pandemic in the four sectors considered, and standard errors are clustered at the county level.[46]

The α_t are the main coefficients of interest and measure the effect in week t of the share of delayed PPP loans in county c. The identifying assumption for estimates of these coefficients to have economic meaning is that conditional on controls, share PPP delayed$_c$ reflects the relative difficulty for small businesses located in county c to obtain a PPP loan and is independent of other factors affecting small business activity during the initial phase of the pandemic.

An obvious concern with this assumption is that share PPP delayed$_c$ may reflect at least partly systematic variations in PPP loan demand across counties. If, for instance, small businesses in counties affected more severely by the pandemic were also more likely to apply for a PPP loan earlier, then this would bias our estimates of α_t away from zero. As Doniger and Kay (2021) show, however, there is no clear geographic concentration in loan issuance within the narrow window around the temporary exhaustion of

[45]Weekly county-level COVID cases and deaths per capita are obtained from Covid Act Now. The NPIs included are weekly county-level indicators of industry closures, stay-at-home orders, and gathering bans from the Data Library on Industry Closures and Reopening by Atalay et al. (2020) and the Center for Disease Control and Prevention. In addition, we include a weekly containment index containing seven indicators of state-level policies from the Oxford Covid-19 Government Response Tracker (see Hale et al., 2021). We allow for differential impact of these controls over subperiods of the sample. Following Bravata et al. (2021), we proxy weekly school closures by the log change in county average school visits relative to one year prior using data on individual schools from Safegraph. For weather, we use the maximum county temperature in a given week from the Climatology Lab. We also controlled for other measures of weather and found very similar results. Finally, county level average household income are 2016-2019 estimates from the American Community Survey. The predictive value of these controls for small business activity is interesting in its own right and extends earlier results by Bartik et al. (2020), Chetty et al. (2020), and Goolsbee and Syverson (2020), among others. We discuss these results as well as further details on the different controls in the Appendix.

[46]County level employment prior to the pandemic is computed from the Quarterly Workforce Indicators (QWI). Results are robust to using county population as weights or estimating the regression at the establishment level that implicitly weighs counties by the count of HB establishments.

PPP considered here, and share PPP delayed_c varies substantially between adjacent counties in the same state. Furthermore, our regressions control for a host of county-specific time-varying factors that may have affected small business activity throughout the pandemic as well as week fixed effects interacted with a county's pre-pandemic average household income that absorb local demand effects related to a county's affluence.[47] Finally, as shown below, our estimations do not show significant effects on employment and revenue prior to the start of PPP, which provides direct evidence against this particular demand-side story.

Another threat to identification could be that there are systematic differences in the distribution of productivity of businesses across counties and that more productive businesses both grow faster and applied for a PPP loan earlier. Since average differences are absorbed by county fixed effects, this would bias estimates of α_t towards zero only insofar as these growth effects were time-varying during the pandemic. To address this possibility, we also estimate regression (6) at the establishment level and control for establishment fixed effects that differentiate out firm-specific factors including productivity differences. As shown in the Appendix, all the results are robust to these establishment-level regressions and are even somewhat stronger.[48]

6.3 Results

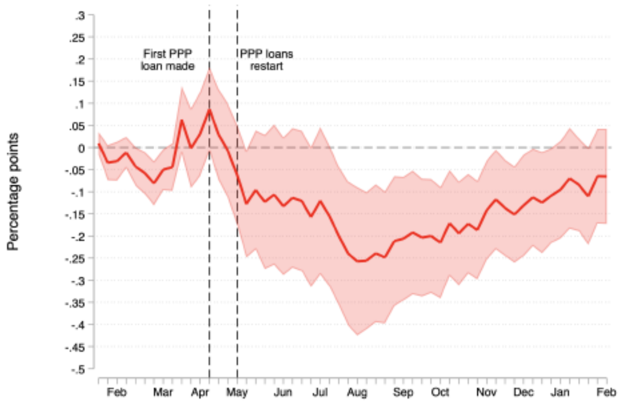
Figure 11 reports the point estimates for α_t together with 95% confidence bands. Panel (a) shows that counties with a higher share of delayed loans experienced lower employment growth starting the week after the exhaustion of the first round of PPP loans. This negative effect becomes stronger from May through August and remains significantly negative through the end of the sample, long after PPP ended. Prior to the temporary exhaustion of PPP in mid-April, the estimates are close to zero and insignificant, indicating that the negative effect is not driven by pretrends. As a further robustness check for pretrends, we estimate the same regression for small business revenue from Womply, made available at the county-week level by Chetty et al. (2020). We find no evidence of pretrends in this regression, either.

[47] As Chetty et al. (2020) document, more affluent localities suffered relatievely larger declines in spending on in-person services and employment in the beginning of the pandemic, presumably due to local demand declining by more in these localities.

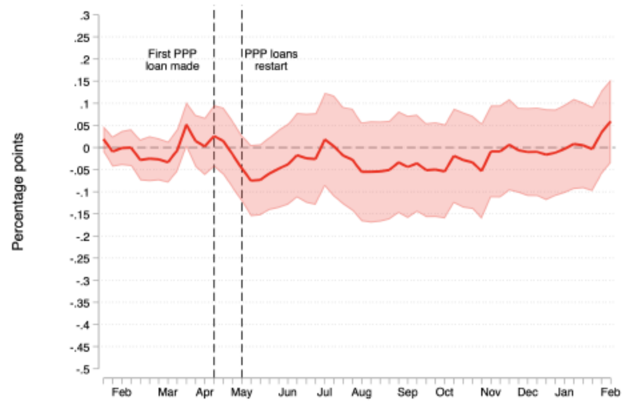
[48] The establishment-level regressions also measure share PPP delayed_c separately for each of the four service sector considered, control for week fixed effects interacted with average household income at the zip-code level, and cluster standard errors at the establishment level.

Figure 11: Effect of delayed PPP loans on small business activity

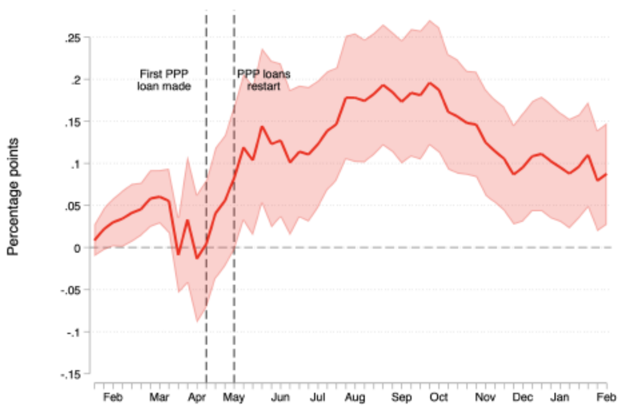
(a) County employment



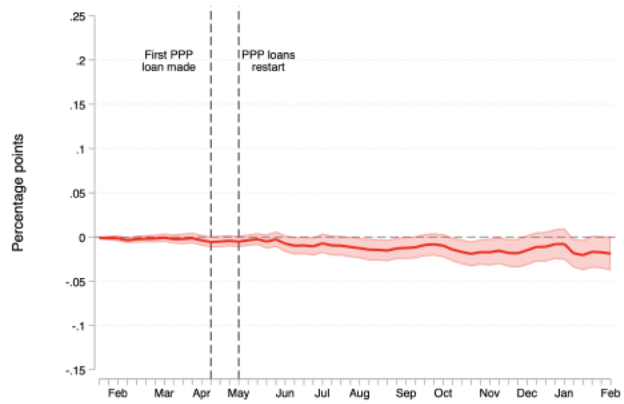
(b) Employment of always active businesses



(c) Business closings



(d) Newly opened businesses



Notes: Coefficient estimates of share PPP delayed$_c$ interacted with weekly fixed effects. Shaded areas show 95% confidence bands. All regressions are estimated over all weeks between January 5-11, 2020 and January 31 - February 6, 2021. share PPP delayed$_c$ is constructed as the amount of PPP loans issued in county c during the week of April 26 relative to the total amount of PPP loans issued per county during the weeks of April 12, April 19, and April 26. County employment in Panel (a) is the percent deviation relative to mid-February 2020 employment for all county-weeks for which the HB sample contains positive employment observations. Employment of always active businesses in Panel (b) is the percent deviation relative to mid-February 2020 employment for all establishments in a county that are continuously active throughout the entire sample. Business closings in Panel (c) is the percent ratio of the total count of establishments closed in a county in week t to the count of businesses in the reference week. Newly opened businesses in Panel (d) is the percent ratio of the total count of new establishments in a county in week t relative to the count of businesses in the reference week. All regressions control for county-specific time-varying controls as described in the text as well as week- and county fixed effects. Regressions are weighted by county employment prior to the pandemic in the four service sectors considered, and standard errors are clustered at the county level.

The fact that the exhaustion of PPP loans lasted for only 10 days and that the additional funding from PPP approved in late April 2020 was not used up by the time the program stopped taking applications in early August 2020, begs the question of why the estimated employment effects are so persistent. To shed light on this question, we run regression (6) separately for employment growth of businesses that are continuously active throughout the sample, the share of business closings, and the share of new business openings (computed, as above, using our methodology that distinguishes business closings and new openings from sample churn and then adjusts for selection).

As shown in panel (b), employment growth by always active businesses in counties with a larger share of delayed PPP loans is barely affected. There is a small decline around the week of the temporary exhaustion of PPP loans that turns marginally significant the week after PPP loans restart, but thereafter the effect is close to zero and insignificant (in the establishment-level regressions, this negative effect is somewhat larger and temporarily significant).

In contrast, as shown in panel (c), counties with a larger share of delayed PPP loans experience a significantly higher rate of business closings starting the week of the temporary exhaustion of PPP. This effect peaks in August 2020, then declines gradually through Fall of 2020, and stabilizes at about half of its peak value by 2021.

Panel (d), finally, shows that the share of delayed PPP loans has only a very small and generally insignificant effect on the rate of new business openings, which confirms the validity of the design since new businesses by definition did not qualify for PPP loans.[49]

To interpret the magnitude of the estimated coefficients, consider the difference in share PPP delayed_c between counties at the 90th and the 10th percentile of the distribution, which is 34% (= 60%−26%). The point estimate at the end of the sample in mid-February 2021 is about −0.15 percentage points for the effect on county employment and about −0.1 percentage points for business closings. This implies that a county at the 90th percentile of delayed PPP loans has about 5.1% lower small business employment (relative to mid-February 2020) and an about 3.4% higher rate of business closings than a county at the 10th percentile. Given that in mid-February 2021, average small business employment across the four sectors considered was about 8% below its pre-pandemic level and the average cumulative closing rate amounted to about 17%, these magnitudes are substantial.

The estimates imply that the negative and persistent employment effect of delays in PPP funding is in large part driven by higher rates of business closings, some of which appear to be permanent. This suggests that the exhaustion of PPP in mid-April 2020 occurred at a critical moment when many small business owners, faced with an unprecedented downturn amid COVID health concerns, stay-at-home orders, and business restrictions had to decide whether to continue operating and hope for loan relief from the government or cut their losses and close shop.[50]

[49] As argued by Acemoglu et al. (2018), supporting incumbent businesses could potentially suppress new openings. The slight negative estimates are consistent with this possibility, although the effect is very small.

[50] As an example of these difficulties, see the NPR Planet Money podcast episode 990 "The Big Small Business Rescue" from April 10, 2020.

6.4 Aggregate employment effects and comparison to the literature

To provide a sense of the the aggregate employment effects of PPP loan delay, we follow an approach similar to Mian and Sufi (2012) and Berger et al. (2020) that is also used in the PPP context by Granja et al. (2020). For each county c, we compute the difference between actual small business employment $E_{c,t}$ and counterfactual employment $\tilde{E}_{c,t}$ under the assumption that the county experienced zero delay in loans around the temporary exhaustion in PPP (which is the case for a small set of counties); i.e.

$$E_{c,t} - \tilde{E}_{c,t} = \frac{\hat{\alpha}_t}{100} \times \text{share PPP delayed}_c \times E_{c,0} \tag{7}$$

where $\hat{\alpha}_t$ are the regression estimates reported in Figure 11 and $E_{c,0}$ is small business employment in county c in the pre-pandemic reference period. We then aggregate across counties using pre-pandemic employment weights.[51] The approach implicitly assumes that $\text{share PPP delayed}_c$ is a good measure of the difficulty of small businesses in obtaining PPP funding during the first round of loans. The approach also abstracts from possible general equilibrium effects of more timely availability of PPP loans and any other differences across counties in the difficulty of obtaining loans that are unrelated to PPP. Nevertheless, the approach is illustrative because it provides a benchmark for the overall effect of PPP and allows us to compare our estimates to other results in the literature.

Given that small business employment in the four sectors considered was about 30 million prior to the pandemic, the point estimate of $\hat{\alpha}_t = -0.25$ for the last week of July 2020 implies that without delays in PPP loans, small business employment in the four sectors would have been about 3 million or 10% higher. In turn, for the last week of January 2021, the estimate of $\hat{\alpha}_t = -0.15$ implies that without delays in PPP loans, small business employment in the four sectors would have been about 1.8 million or 6% higher. Since the delays in PPP loans could have been avoided by appropriating a larger initial amount for PPP in the CARES Act, the costs of doing so would have been essentially zero. Vice versa, the estimates suggest that if PPP had not been part of the CARES Act, small business closings would have been substantially larger and the pandemic would have caused a larger and longer-lasting decline in service sector jobs.[52]

[51] Specifically, the aggregate employment effect relative to pre-pandemic employment is estimated as $\frac{E_t - \tilde{E}_t}{E_0} = \frac{\hat{\alpha}_t}{100} \sum_c \frac{E_{c,0}}{E_0} \text{share PPP delayed}_c$, where $E_{c,0}$ and E_0 denote pre-pandemic employment in county c and nationwide, respectively, for small businesses in the four sectors considered.

[52] We refrain from attempting to infer the overall number of jobs saved by PPP for two reasons. First, our estimates pertain to small businesses in four of the service sectors affected most by the pandemic. Larger businesses and businesses in other sectors that received PPP loans may have been less dependent on PPP funding, but the HB data do not allow us to quantify the extent of this treatment effect heterogeneity. Second, under the counterfactual assumption that the CARES Act or the subsequent COVID relief bills had not contained any funding for PPP, small businesses as a whole could have

Our results complement event-study estimates of the effects of PPP by Bartik et al. (2021), Doniger and Kay (2021), and Granja et al. (2020) among others. Granja et al. (2020) use Homebase data like us but apply a different research design that exploits local variations in the presence of banks that processed PPP loans at varying expediency. They find that over the months of April, May and June 2020, employment in small businesses would have been 4.5% higher if all banks had been equally expedient in making loans, which implicitly assumes that the initial PPP funding in the CARES Act would have been larger (i.e. no loan delays). This number is about half of our estimate. However, their estimation treats all exits from Homebase as business closings, which is likely to impart substantial error in the cross-regional variation in small business employment. Bartik et al. (2021) use data from a survey of small businesses owners in late April 2020 during the temporary exhaustion phase of PPP. Leveraging information on existing banking relationships as instrumental variables, they find that PPP loan approval led to a 14 to 30% increase in expected survival probability and had a positive but imprecisely estimated effect on employment. Our estimates provide confirmation of this finding with actual data on employment and closings. Using the same aggregation approach as for employment above, the estimated aggregate reduction in permanent closings if there had been no delay in PPP loans is about 5%. This is lower than the estimates in Bartik et al. (2021), which may be due to the fact that we use actual data as opposed to expectations formed in the initial phase of the pandemic when uncertainty was likely higher. Doniger and Kay (2021), whose research design we adopt here, use monthly household survey data from the CPS and find persistent effects of PPP loan delay on unemployment and non-participation. From their estimates, they infer that a reduction in the average share of delayed PPP loans by 20% would have increased private-sector employment in mid-May 2020 by nearly 2.8 million with much of the job savings concentrated in small businesses. This number is difficult to compare to our estimates since it pertains to all private sector employment while we focus on the four hardest-hit service sectors. At the same time, our estimates confirm that PPP loan delay had a substantial negative effect on small business employment and provide an explanation for why this effect is persistent: it is in largely due to business closings of which many appear to be permanent.[53]

This finding is confirmed in more recent studies by Dalton (2021) and Autor et al. (2022) who match establishment-level data from the QCEW and ADP, respectively, to individual PPP loan information and

reacted very differently from the present context where PPP funds were available but temporarily ran out for a relatively short period of time. This in turn could have led to important general equilibrium effects that are difficult to quantify.

[53]More generally, our results are consistent with a growing literature documenting that limited cash-on-hand and working capital adversely affects labor demand and makes small businesses more sensitive to negative shocks (e.g. Chodorow-Reich, 2014, Bacchetta et al., 2019, Barrot and Nanda, 2020, or Mehrotra and Sergeyev, 2020 among others). Our results, however, put increased emphasis on the effects that these financial frictions can have on the extensive margin – i.e. business closings – which likely has more permanent effects.

compare employment and closing probabilities of businesses who received a PPP loan earlier with those who recevied a loan later. Like us, they find sizable effects on employment that are concentrated among businesses with fewer than 50 employees, and that a large part of these employment effects are due to business closures.

Finally, Autor et al. (2020), Chetty et al. (2020), and Hubbard and Strain (2020) exploit the 500 employee threshold for PPP loan eligibility to estimate the overall impact of PPP. These studies find more modest employment effects, suggesting that businesses around the 500 employee threshold have been less dependent on PPP loan support, which is consistent with results by Chodorow-Reich et al. (2020). This suggests that the effectiveness of PPP could have been enhanced if the program had been restricted at least initially to the smallest businesses.

7 Conclusion

In this paper, we use establishment-level data from Homebase, a scheduling and time clock software provider, to construct weekly estimates of the effect of the COVID-19 pandemic on small business dynamics and employment in four of the hardest hit service sectors. The main methodological contribution relative to the many other studies using Homebase and other high-frequency establishment-level datasets is that we match the Homebase establishment records with independent data to assess the representativeness of the establishments in the Homebase sample and to distinguish business openings and closings from sample churn.

Our implementation uses data from Safegraph, Google, Facebook, but other datasets measuring business activity could be used as well. Similarly, while the Homebase data provides an excellent case to test the quality of our methods, other establishment-level datasets could be used. As such, we consider our paper as a proof of concept on how to combine different data sources to construct employment estimates that directly incorporate the effects of business openings and closings in almost real-time and that can be benchmarked to official statistics and used to measure the impact of rapidly disseminating shocks and economic policies.

References

Acemoglu, D., U. Akcigit, N. Bloom, and W. R. Kerr (2018). Innovation, Reallocation and Growth. *American Economic Review 108*, 3450–3491. 49

Atalay, E., S. Fujita, S. Mahadevan, R. Michaels, and T. Roded (2020). Reopening the Economy: What are the risks, and what have States done? Research brief, Federal Reserve Bank of Philadelphia. 45

Autor, D., D. Cho, L. D. Crane, M. Goldar, B. Lutz, J. Montes, W. B. Peterman, D. Ratner, D. V. Vallenas, and A. Yildirmaz (2020). An Evaluation of the Paycheck Protection Program Using Administrative Payroll Microdata. Working paper, MIT. 1, 6.4

Autor, D., D. Cho, L. D. Crane, M. Goldar, B. Lutz, J. Montes, W. B. Peterman, D. Ratner, D. V. Vallenas, and A. Yildirmaz (2022). The 800*BillionPaycheckProtectionProgram* : *WhereDidtheMoneyGoandWhyDidItGoThere?WorkingPaper*29669, *NationalBureauofEconomicResearch.*1,

Bacchetta, P., B. Kenza, and C. Poilly (2019). Corporate cash and employment. *American Economic Journal: Macroeconomics 11 (3): 30–66*. 53

Barrot, J.-N. and R. Nanda (2020). The Employment Effects of Faster Payment: Evidence from the Federal Quickpay Reform. *The Journal of Finance 75 (6):3139–3173*. 53

Bartik, A. W., M. Bertrand, Z. B. Cullen, E. L. Glaeser, M. Luca, and C. T. Stanton (2020). The Impact of COVID-19 on Small Business Outcomes and Expectations. *Proceedings of the National Academy of Sciences 117*(30), 17656–17666. 1

Bartik, A. W., M. Bertrand, F. Lin, J. Rothstein, and M. Unrath (2020). Measuring the labor market at the onset of the COVID-19 crisis. *Brookings Papers on Economic Activity forthcoming*. 1, 5.3, 45

Bartik, A. W., Z. Cullen, E. L. Glaeser, M. Luca, C. Stanton, and A. Sunderam (2021). The Targeting and Impact of Paycheck Protection Program Loans to Small Businesses. Working Paper 21-021, Harvard Business School. 1, 6.1, 6.4

Bartlett, R. P. and A. Morse (2020). Small Business Survival Capabilities and Policy Effectiveness: Evidence from Oakland. Working paper, National Bureau of Economic Research. 1

Berger, D., N. Turner, and E. Zwick (2020). Stimulating Housing Markets. *Journal of Finance 75*(1), 277–321. 6.4

Bick, A. and A. Blandin (2020, April). Real Time Labor Market Estimates During the 2020 Coronavirus Outbreak. Working paper, Arizona State University. 1

Borowczyk-Martins, D. and E. Lalé (2019). Employment Adjustment and Part-time Work: Lessons from the United States and the United Kingdom. *American Economic Journal: Macroeconomics 11*(1), 389–435. 5.4

Bravata, D., J. H. Cantor, N. Sood, and C. M. Whaley (2021). Back to School: The Effect of School Visits During Covid-19 on Covid-19 Transmission. Working paper, National Bureau of Economic Research. 45

Cajner, T., L. D. Crane, R. A. Decker, J. Grigsby, A. Hamins-Puertolas, E. Hurst, C. Kurz, and A. Yildirmaz (2020, May). The U.S. Labor Market during the Beginning of the Pandemic Recession. 1, 5.1, 5.3

Chetty, R., J. N. Friedman, N. Hendren, M. Stepner, and the Opportunity Insights Team (2020). The Economic Impacts of COVID-19: Evidence from a New Public Database Built from Private Sector Data. Working paper 26463, National Bureau of Economic Research. 3, 1, 5.3, 45, 6.3, 47, 6.4

Chodorow-Reich, G. (2014). The employment effects of credit market disruptions:Firm-level evidence from the 2008–9 financial crisis. *The Quarterly Journal of Economics129 (1): 1–59*. 53

Chodorow-Reich, G., O. Darmouni, S. Luck, and M. Plosser (2020). Bank Liquidity Provision Across the Firm Size Distribution. *Journal of Financial Economics forthcoming*. 1, 6.4

Coibon, O., Y. Gorodnichenko, and M. Weber (2020, April). Labor Markets During the COVID-19 Crisis: A Preliminary View. Working Paper 27017, National Bureau of Economic Research. 1

Crane, L. D., R. A. Decker, A. Flaeen, A. Hamins-Puertolas, and C. Kurz (2022). Business Exit During the COVID-19 Pandemic: Non-Traditional Measures in Historical Context. *Journal of Macroeconomics 72*, 103419. 1, 4.1, 5.2, 35

Dalton, M. (2021, November). Putting the Paycheck Protection Program into Perspective: An Analysis Using Administrative and Survey Data. Working Paper 542, Bureau of Labor Statistics. 1, 6.4

Dalton, M., E. W. Handwerker, and M. A. Loewenstein (2020, October). Employment changes by employer size during the COVID-19 pandemic: a look at the Current Employment Statistics survey microdata. *Monthly Labor Review*, 1–17. 1, 5.1, 5.3

Decker, R. A. and J. C. Haltiwanger (2022). Business Entry and Exit in the COVID-19 Pandemic: A Preliminary Look at Official Data. Feds notes, Board of Governors of the Federal Reserve System. 1, 31, 5.2

Doniger, C. L. and B. Kay (2021). Ten Days Late and Billions of Dollars Short: The Employment Effects of Delays in Paycheck Protection Program Financing. Finance and economics discussion series 2021-003, Board of Governors of the Federal Reserve System. 1, 6.1, 6.2, 6.2, 6.4

Fairlie, R. (2020). The impact of COVID-19 on small business owners: Evidence from the first three months of widespread social-distancing restrictions. *Journal of Economics Management Strategy 29*, 727–740. 1

Fairlie, R. W., F. M. Fossen, R. L. Johnsen, and G. Droboniku (2022). Were Small Businesses More Likely to Permanently Close in the Pandemic. Working paper 30285, National Bureau of Economic Research. 1, 5.2

Fazio, C. E., J. Guzman, Y. Liu, and S. Stern (2021). How is COVID Changin the Geography of Entrepreneurship? Evidence from the Startup Cartography Project. Working paper 28787, National Bureau of Economic Research. 1, 5.2

Fort, T. C., J. C. Haltiwanger, R. S. Jarmin, and J. Miranda (2013). How Firms Respond to Business Cycles: The Roel of Firm Age and Firm Size. *IMF Economic Review 61*(3), 520–559. 1, 5.1

Fujita, S. and G. Moscarini (2017). Recall and unemployment. *American Economic Review 107*(12), 3875–3916. 39

Ganong, P., F. Greig, M. Liebeskind, P. Noel, D. M. Sullivan, and J. Vavra (2021). Spending and Job Search Impacts of Expanded Unemployment Benefits: Evidence from Administrative Micro Data. Working paper, Becker Friedman Institute. 5.5

Goolsbee, A. and C. Syverson (2020). Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020. Working paper, Becker Friedman Institute. 45

Granja, J., C. Makridis, C. Yannelis, and E. Zwick (2020). Did the Paycheck Protection Program Hit the Target? Technical report. 1, 6.1, 6.4, 6.4

Hale, T., N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake,

L. Hallas, S. Majumdar, and H. Tatlow (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker. *Nature Human Behaviour 5*(529-538). 45

Haltiwanger, J. C. (2021). Entrepreneurship During the Covid-19 Pandemic: Evidence from the Business Formation Statistics. Working paper 28912, National Bureau of Economic Research. 1, 5.2

Haltiwanger, J. C., H. R. Hyatt, L. B. Kahn, and E. McEntarfer (2018). Cyclical Job Ladders by Firm Size and Firm Wage. *American Economic Journal: Macroeconomics 10*(2), 52–85. 1, 5.1

Hubbard, G. and M. R. Strain (2020). Has the Paycheck Protection Program Succeeded? *Brookings Papers on Economic Activity forthcoming*. 1, 6.4

Kahn, L. B., F. Lange, and D. G. Wiczer (2020, April). Labor Demand in the Time of COVID-19: Evidence from Vacancy Postings and UI Claims. Working Paper 27061, National Bureau of Economic Research. 1

Lewis, D., K. Mertens, J. H. Stock, and M. Trivedi (2021). Measuring Real Activity Using a Weekly Economic Index. *Journal of Applied Econometrics forthcoming*. 1

Mehrotra, N. and D. Sergeyev (2020). Financial shocks, firm credit and the Great Recession. *Journal of Monetary Economics In Press*. 53

Mian, A. and A. Sufi (2012). The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program. *Quarterly Journal of Economics 127*(1107-1142). 6.4

Moscarini, G. and F. Postel-Vinay (2012). The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment. *American Economic Review 102*(6), 2509–39. 1, 5.1

Rinz, K. (2020). Understanding Unemployment Insurance Claims and Other Labor Market Data During the COVID-19 Pandemic. Working paper. 4.3

Vaan, M. D., S. Mumtaz, A. Nagaraj, and S. B. Srivastava (2021). Social Learning in the COVID-19 Pandemic: Community Establishments' Closure Decisions Follow those of Nearby Chain Establishments. *Management Science 67*(7), 4446–4454. 3

Yelp (2020, September). Local economic impact report. https://www.yelpeconomicaverage.com/business-closures-update-sep-2020. 3

Supplementary material for:

Measuring Small Business Dynamics and Employment with Private-Sector Real-Time Data

André Kurmann
Drexel University

Etienne Lalé
Université du Québec à Montréal

Lien Ta
Drexel University

Contents

A Homebase data

The Homebase (HB) data contains employment and wage records for small businesses primarily engaged in customer-oriented services. In February 2020, the database covered 938,072 individual workers employed in 94,203 business locations (establishments) belonging to approximately 80,921 firms.

A.1 Description and definitions

- A HB customer (firm) is identified by a unique persistent `company_id`. Different business locations (establishments) belonging to the same firm are identified by a unique persistent `location_id`. The establishment is our main unit of observation.

- Some HB customers create several `location_id`'s for the same establishment in order to track different departments. These departments either have the same address or no address. We tag them by creating a common `parent_location_id`. In the sample construction, we collapse `location_id`'s with the same `parent_location_id` into a single `location_id`.[1]

- Each user of HB's services has a unique persistent `user_id`. Associated with each `user_id` is a `level` (employee, manager, general manager) and an `owner_status` (owner, non-owner). The values of these fields do not change over time.

[1] `parent_location_id`'s account for about 6% of all unique HB `location_id`'s., and about 15% of `location_id`'s have a `parent_location_id`.

- A user_id shows up in the database on a given day if the user had some activity; e.g. scheduling or logging hours, sending messages to another user, etc.[2]

- Each user_id comes with a hours_scheduled and a hours_worked field. If both fields are empty, we designate the user as having "untracked hours". If either the hours_scheduled or the hours_worked field contains data, then we designate the user as having "tracked hours". When a user has data in both fields, we use hours_worked to measure actual hours worked, although we find that the data in hours_scheduled and hours_worked correspond closely.[3]

- In a given week in 2020, user_id's with untracked hours make up one-fifth of all user_id's in the raw HB data. For those with tracked hours, 30% of user_id's have hours_worked and a missing hours_scheduled field. The remaining 70% of user_id's with tracked hours have entries in both the hours_scheduled and hours_worked fields. 42% of these user_id's have hours_worked that are exactly equal to hours_scheduled, suggesting they use HB for the purpose of scheduling hours and that the hours_worker field has been populated with information from the hours_scheduled field.

- Figure A1 shows the distribution of weekly hours worked on average over the first 10 weeks of 2020.[4] Among managers and general managers, there is a mode at 40 hours, and a substantial proportion of managers with 0 tracked hours (8.7% of managers, 3.3% of general managers). Median weekly hours worked equal 21.7 for employees, 28.5 for general managers, and 24.3 for managers. Employees make up 90% of all user_id's in the data with tracked hours in the first 10 weeks of 2020, while general managers account for 1% of these user_id's.

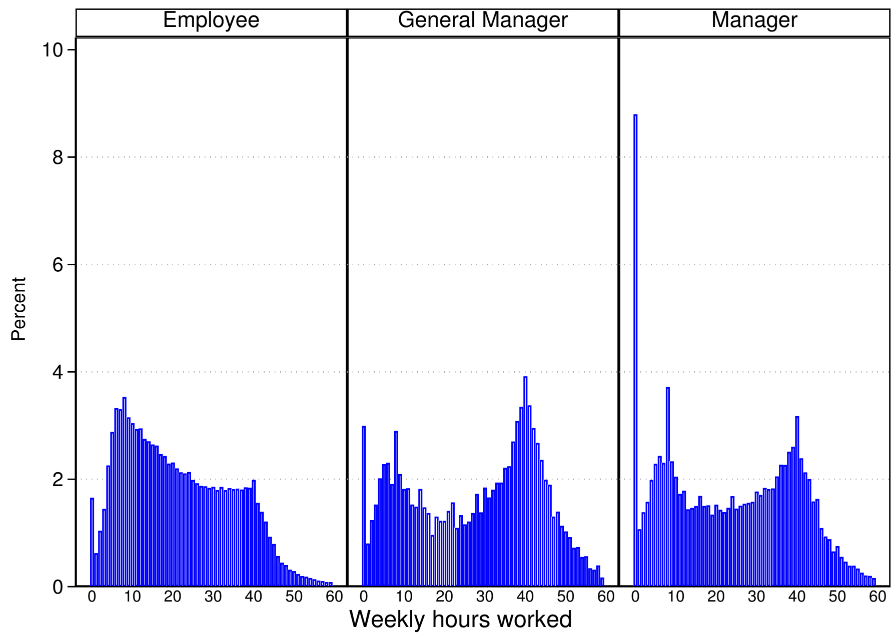
- Figure A2 shows the proportion of location_id's in 2020 with tracked hours only, untracked hours only, and both tracked and untracked hours. At the onset of the pandemic (mid-March 2020), location_id's with untracked hours make up 10% of all location_id's with some activity. The corresponding figure for location_id's with both untracked and tracked hours is 66%. During the first weeks of the pandemic, there is a clear shift towards location_id's with untracked hours only. The increase is mainly driven by an inflow of location_id's that used to have both untracked and tracked hours. Transition probabilities (not reported here) of switching across the 3 groups return

[2] But simply being logged in to the app on a cellular device is not counted as an activity.

[3] The difference between hours worked and hours scheduled, conditional on being different from each other, is symmetric, bell-shaped centered on zero. For 75% of user_id's with different hours worked and hours scheduled, the absolute difference between the two measurements is less than 1.2 hours.

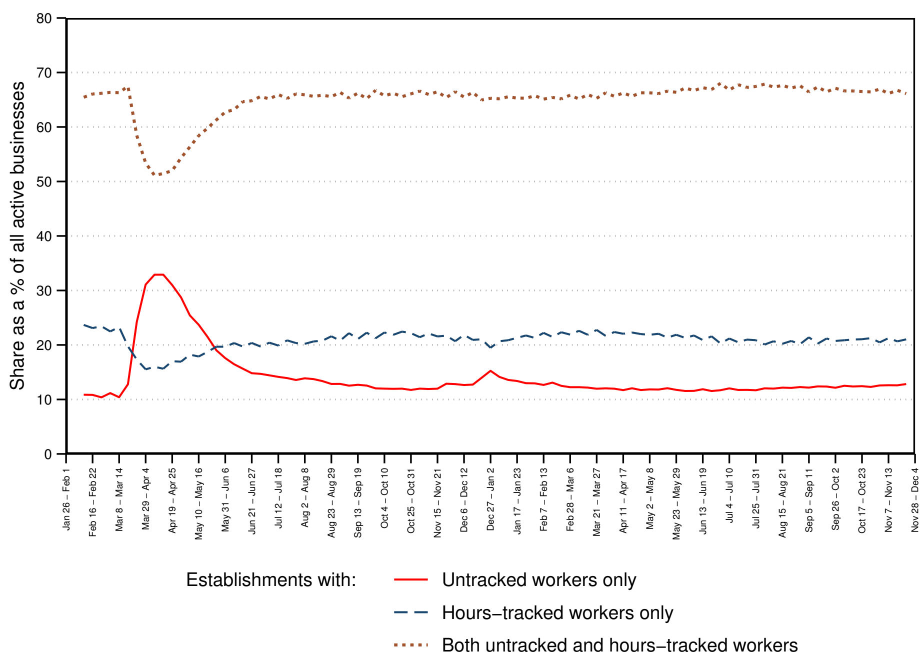
[4] In Figure A1, the sample is restricted to weekly hours worked (strictly) under 60 hours. Less than 1 percent of workers have weekly hours worked (averaged over this 10-week window of time) above the 60 hours cutoff.

Figure A1: Distribution of weekly hours worked in raw HB data



Notes: Weekly hours worked averaged over the first 10 weeks of 2020 for all HB `user_id`'s with tracked (worked or scheduled) hours.

Figure A2: Establishments with untracked and tracked hours



Notes: HB establishments with untracked only (solid line), hours-tracked only (dashed line), and both untracked and hours-tracked workers (dotted line), as a share (expressed in %) of all establishments with some activity in 2020.

to their pre-pandemic levels by mid-April. The distribution of `location_id`'s across the 3 groups takes longer to recover.

A.2 Employment, hours worked and active establishments

A key operational definition is that of employment. We measure employment of a location as the sum of all users with positive hours worked or hours scheduled plus active users with untracked hours. Employment is set to 0 if the establishment has zero tracked hours. In other words, we only count employment at establishments that have at least one hourly-tracked employee with positive hours worked or hours scheduled. Total weekly hours and average weekly hours per worker are computed only for `user_id`'s users with positive hours worked or hours scheduled.

We use these measurements of employment and hours worked to define *active* establishments. To be considered as active in a given week we require a location to have positive employment during that week and have at least 40 total weekly hours. Our goal in adopting this definition is to purge the data from "try-outs", i.e. locations that only show up in HB data for a short period of time and without significant tracked hours.[5]

A.3 Industry classification

The historical HB data comes with an industry category for each establishment, but the available categories do not directly line up with standard industry classification, and for more than 10% of the records the industry category is not usable because it is "Other" or "Unknown". Instead of using this industry classification in an ad hoc manner, we match all available HB locations by name and address to Points of Interest (POI) from Safegraph "Core Places data", which come with their own NAICS-6 code. The Safegraph Core Places data consists of more than 8 million POIs. A POI is defined as a location where individuals spend time or money. Matching the HB locations to Safegraph's POIs involves extensive data cleaning and standardization. The data and match procedure are described in detail in Sections B.1 and C.1. We only retain establishments that match either exactly or with a high match rate.[6] Table A1 reports the mapping of HB industry categories to 2-digit NAICS codes for the HB locations that are

[5] The total count of locations provided at the beginning of this section already purges the data from a few "try-out" businesses. Specifically, we remove locations that either have (i) less than 1 total weekly hours worked or scheduled in at least 50% of the weeks with tracked hours or (ii) total weekly hours worked or scheduled that are never higher than 5 hours.

[6] In December 2020, HB independently started publishing NAICS industry classification. This classification is available only for locations that are active from December 2020 onward. Since many establishments that were active in 2019 and 2020 are no longer in the HB sample, this NAICS classification is not directly useful for our estimation and benchmarking, which starts in 2019. However, we compare our industry classifications to the one provided by HB and find a high level of overlap, especially at the 2-digit level that we use for our estimation.

Table A1: Comparison between HB industry codes and 2-digit NAICS codes obtained through Safegraph

	NAICS														
	11-23	**31-33**	**42**	**44-45**	**48**	**51**	**52-53**	**54-56**	**61**	**62**	**71**	**72**	**81**	**92**	**Total**
Beauty & personal care	0	2	4	29	0	0	3	1	6	26	5	13	329	0	418
Charities, education & membership	10	6	12	174	6	32	80	11	312	439	165	82	219	30	1,578
Food & drink	35	1,883	57	1,754	38	37	330	49	198	402	435	20,435	262	12	25,927
Health care and fitness	8	26	57	496	8	11	149	237	85	2,001	545	131	615	15	4,384
Home and repair	85	26	25	328	13	0	70	60	18	38	22	34	330	5	1,054
Leisure and Entertainment	3	16	3	179	4	49	53	11	35	36	432	212	38	3	1,074
Professional Services	121	200	108	1,481	101	77	518	333	238	632	533	984	896	69	6,291
Retail	17	40	32	349	30	20	337	255	55	217	63	116	330	50	1,911
Transportation	42	210	300	5,692	63	107	369	82	116	187	201	912	519	12	8,812
Other	8	3	5	125	123	5	39	14	9	32	19	42	92	1	517
Unknown	57	439	75	1,404	53	58	350	177	210	702	530	2,039	508	20	6,622
Total	386	2,851	678	12,011	439	396	2,298	1,230	1,282	4,712	2,950	25,000	4,138	217	58,588

Notes: HB data for February 2020. The table compares HB industry codes (in the first column of each row) to NAICS codes (in each column) obtained using Safegraph data (see Section C). Each cell is the count of establishments (including establishments with 50+ employees) that are active during the reference week of 2020 (Feb 9 – Feb 15).

active in mid-February 2020.[7]

A.4 Base sample

For both 2019 and 2020, we form a base sample consisting of HB establishments that are active according to the above definition for each of the three weeks centered around the week containing February 12 (week 0) plus all establishments that are temporarily inactive during this period but had positive employment in at least three weeks at some point prior to the mid-February reference period and become active for at least three consecutive weeks at some point thereafter. For active establishments, we determine their size class (i.e., 1 to 4, 5 to 9, 10 to 19, and 20 to 49 employees) by taking average employment over the three week window centered on week 0. For temporarily inactive establishments, size class is determined by the average of employment over all weeks of activity prior to the reference period.

Table A2: Homebase sample counts

	2019		2020	
All estab. potentially in base sample	48,993	(100%)	64,807	(100%)
- active in mid-February	44,294	(90.4%)	59,207	(91.4%)
- temporarily inactive in mid-February	4,699	(9.6%)	5,600	(8.6%)
In-scope estab. in base sample	38,193	(100%)	49,268	(100%)
- active in mid-February	34,757	(91.0%)	45,454	(92.3%)
- temporarily inactive in mid-February	3,436	(9.0%)	3,814	(7.7%)

Notes: The table shows counts of establishments that (i) are either active or temporarily inactive in mid-February (top panel) of 2019 or 2020, and (ii) are in scope according to our industry (i.e. they belong to either Retail Trade, Education & Health Services, Leisure & Hospitality, or Other Services according to the NAICS codes obtained through Safegraph), size class (fewer than 50 workers), and geographic requirements (bottom panel).

The top panel of Table A2 reports the count of active and temporarily inactive establishments for both the 2019 and 2020 base periods. Interestingly, in both base samples there are close to 10% of temporarily inactive establishments, which is consistent with evidence from administrative data on temporarily closed establishments. From this sample, we further drop all establishments for which we do not have a sufficiently high quality match with Safegraph to confidently attribute a NAICS-6 industry code; establishments with a 2-digit NAICS code different from the four sectors that we study (44-45, 61-62, 71-72 and 81);[8] establishments with 50 employees or more; and establishments based in the U.S. Virgin

[7] The total count of establishments in Table A1 is different from the sample count of active establishments in mid-February in Table A2 (59,207 vs. 58,588 establishments in Table A1) because a few establishments match to a Safegraph POI with a missing NAICS code (and are thus included in Table A2 but not in Table A1). About 1% of POIs in Safegraph have a missing NAICS code.

[8] We retain a few (about 1,000 in both 2019 and 2020) businesses for which the quality of the match to SG data is low

Islands, Puerto Rico, or the island of Guam. As the lower panel of Table A2, this leaves us with an in-scope base sample of 38,193 establishments for 2019 and 49,268 establishments for 2020.

For our estimations, we weight establishments by 2-digit NAICS category × size class × geographic area cells. There are six 2-digit NAICS categories (44-45, 61, 62, 71, 72 and 81), four establishment size classes (1 to 4, 5 to 9, 10 to 19, and 20 to 49 employees) and thirteen geographic areas listed below in Table A4. Thus, we have a total of 312 cells. We drop from the analysis those cells where there are too few establishments in HB compared to the QCEW. In the 2020 base sample, for instance, we retain 296 industry-size-region cells i. The average number of HB establishments per cell i is 169, the median is 80, the 5th percentile is 11 and the 95th percentile is 640. The smallest cell i in the base sample of 2020 is NAICS 61 of size 5 to 9 employees in the Pacific region excluding California (that is to say Alaska, Hawaii, Oregon, Washington). This cell contains 7 establishments and its QCEW-HB weight ω_i is equal to 156. The largest cell is NAICS 72 of size 10 to 19 employees in the State of California; it contains 1,993 establishments and its QCEW-HB weight ω_i is equal to 11 (i.e., HB covers almost 10 percent of all establishments in this cell).

A.5 Benchmarking establishment counts and employment to the QCEW

Tables A3a and A3b display the count and distribution of establishments for each 2-digit NAICS categories in our 2020 base sample, and compares them with the corresponding data from the Quarterly Census of Employment and Wages (QCEW). Two features are noteworthy. First, the HB data is less skewed towards very small establishments than the QCEW. This is especially true in NAICS 61, 62, and 71. A large part of this difference is explained by the (very) large establishment counts in size class 1–4 of the QCEW. For instance, when we compare establishment counts in size class 1–4 of the QCEW with those from the Business Dynamics Statistics (BDS), we find enormous differences: for NAICS 61, 62 and 71, the QCEW counts are respectively 165%, 307%, and 145% higher than those from the BDS.[9] Second, the largest industry in the HB data is by far NAICS 72 (Accommodation and Food Services): it makes up for about half of all the businesses from the 2020 base sample. In the QCEW data, NAICS 72 accounts

but where the NAICS code obtained through the matching procedure is consistent with the industry code provided in HB data. Namely, we keep businesses with NAICS 44-45 and HB industry code “Retail”; businesses with NAICS 611 and HB industry code “Charities, Education & Membership”; businesses with NAICS 621, 623 or 624 and HB industry code “Health Care and Fitness”; businesses with NAICS 71 and HB industry code “Leisure and Entertainment”; businesses with NAICS 722 and HB industry code “Food & Drink”; businesses with NAICS 811 or 812 and HB industry code “Home and Repair”; and businesses with NAICS 813 and HB industry code “Charities, Education & Membership”.

[9] We run this comparison with data from 2019, which is the most recent available data for the BDS. Barnatchez et al. [2017] provide a detailed analysis of the discrepancies between these establishment counts. For establishment size 1–4 in NAICS 62 (Health care and social assistance) in 2019, the magnitude of the discrepancy is remarkable: 376,026 establishments according to the BDS vs. 1,154,994 in the QCEW.

Table A3a: Establishment counts and employment in HB and the QCEW

NAICS 44-45 - Retail Trade

	HB data				QCEW data					
	Estab.		Workers		Estab.			Workers		
	#	%	#	%	#	% all	% small	#	% all	% small
1–4	3,706	30.4	10,631	10.6	474,656	45.4	48.2	856,574	5.6	11.5
5–9	4,763	39.1	29,809	29.8	245,749	23.5	25.0	1,657,230	10.9	22.3
10–19	2,796	22.9	34,968	35.0	178,365	17.1	18.1	2,377,561	15.6	32.0
20–49	928	7.6	24,504	24.5	85,740	8.2	8.7	2,548,858	16.7	34.3
50–99	0	0.0	0	0.0	31,076	3.0	–	2,159,731	14.2	–
100+	0	0.0	0	0.0	29,303	2.8	–	5,658,339	37.1	–
Total	12,193	100	99,912	100	1,044,889	100	–	15,258,293	100	–

NAICS 61 - Educational Services

	HB data				QCEW data					
	Estab.		Workers		Estab.			Workers		
	#	%	#	%	#	% all	% small	#	% all	% small
1–4	149	14.2	408	3.2	73,977	56.4	60.9	94,839	3.1	11.2
5–9	342	32.6	2,272	18.0	18,558	14.1	15.3	124,591	4.1	14.7
10–19	372	35.5	4,707	37.4	15,674	11.9	12.9	213,644	7.1	25.3
20–49	186	17.7	5,209	41.4	13,351	10.2	11.0	412,494	13.7	48.8
50–99	0	0.0	0	0.0	5,331	4.1	–	368,226	12.2	–
100+	0	0.0	0	0.0	4,299	3.3	–	1,806,533	59.8	–
Total	1,049	100	12,596	100	131,190	100.0	–	3,020,327	100.0	–

NAICS 62 - Health Care and Social Assistance

	HB data				QCEW data					
	Estab.		Workers		Estab.			Workers		
	#	%	#	%	#	% all	% small	#	% all	% small
1–4	700	18.5	2,074	5.4	1,250,472	72.6	75.4	1,499,144	7.3	20.6
5–9	1,578	41.7	10,067	26.0	184,691	10.7	11.1	1,235,153	6.1	17.0
10–19	1,035	27.3	13,400	34.6	132,454	7.7	8.0	1,788,534	8.8	24.6
20–49	473	12.5	13,216	34.1	90,702	5.3	5.5	2,739,494	13.4	37.7
50–99	0	0.0	0	0.0	33,377	1.9	–	2,330,236	11.4	–
100+	0	0.0	0	0.0	30,056	1.7	–	10,810,686	53.0	–
Total	3,786	100	38,757	100	1,721,752	100.0	–	20,403,247	100.0	–

Notes: HB and QCEW data for February 2020. The columns titled "#" report the number of establishments by class size, and employment by establishment class size. In the "HB data" panels, the columns titled "%" show the distribution of establishments by class size and distribution of employment by establishment class size. In the "QCEW data" panels, the columns titled "% all" show the distribution of establishments by class size and distribution of employment by establishment class size, and the columns titled "% small" show the distribution among small establishments (establishments with fewer than 50 workers).

Table A3b: Establishment counts and employment in HB and the QCEW

NAICS 71 - Arts, Entertainment, and Recreation

	HB data				QCEW data					
	Estab.		Workers		Estab.			Workers		
	#	%	#	%	#	% all	% small	#	% all	% small
1–4	442	15.0	1,277	3.7	97,592	61.5	65.1	109,419	4.9	11.9
5–9	953	32.4	6,000	17.4	20,835	13.1	13.9	138,878	6.2	15.2
10–19	983	33.4	12,328	35.8	17,875	11.3	11.9	245,345	10.9	26.8
20–49	567	19.3	14,847	43.1	13,666	8.6	9.1	422,463	18.8	46.1
50–99	0	0.0	0	0.0	5,181	3.3	–	353,624	15.8	–
100+	0	0.0	0	0.0	3,432	2.2	–	971,527	43.3	–
Total	2,945	100	34,453	100	158,581	100.0	–	2,241,256	100.0	–

NAICS 72 - Accommodation and Food Services

	HB data				QCEW data					
	Estab.		Workers		Estab.			Workers		
	#	%	#	%	#	% all	% small	#	% all	% small
1–4	3,424	13.3	9,893	3.2	232,321	31.6	34.0	365,553	2.7	4.3
5–9	8,332	32.3	55,156	17.8	124,265	16.9	18.2	848,051	6.3	10.0
10–19	9,608	37.3	125,595	40.5	160,482	21.8	23.5	2,267,987	16.8	26.8
20–49	4,399	17.1	119,681	38.6	166,507	22.7	24.4	4,972,904	36.9	58.8
50–99	0	0.0	0	0.0	39,886	5.4	–	2,629,706	19.5	–
100+	0	0.0	0	0.0	11,309	1.5	–	2,391,440	17.7	–
Total	25,763	100	310,325	100	734,770	100.0	–	13,475,641	100.0	–

NAICS 81 - Other Services (except Public Administration)

	HB data				QCEW data					
	Estab.		Workers		Estab.			Workers		
	#	%	#	%	#	% all	% small	#	% all	% small
1–4	1,016	28.8	2,728	9.3	594,786	72.6	73.4	848,552	19.2	25.6
5–9	1,444	40.9	9,228	31.3	123,871	15.1	15.3	811,239	18.3	24.5
10–19	797	22.6	10,141	34.4	63,202	7.7	7.8	834,213	18.8	25.2
20–49	275	7.8	7,388	25.1	28,144	3.4	3.5	815,662	18.4	24.6
50–99	0	0.0	0	0.0	6,103	0.7	–	414,017	9.3	–
100+	0	0.0	0	0.0	3,560	0.4	–	706,529	15.9	–
Total	3,532	100	29,485	100	819,666	100.0	–	4,430,212	100.0	–

Notes: HB and QCEW data for February 2020. The columns titled "#" report the number of establishments by class size, and employment by establishment class size. In the "HB data" panels, the columns titled "%" show the distribution of establishments by class size and distribution of employment by establishment class size. In the "QCEW data" panels, the columns titled "% all" show the distribution of establishments by class size and distribution of employment by establishment class size, and the columns titled "% small" show the distribution among small establishments (establishments with fewer than 50 workers).

Table A4: Geographic distribution of establishments in the QCEW and HB data

	2019			2020		
	HB base sample		QCEW small estab.	HB base sample		QCEW small estab.
	#	%	%	#	%	%
Alaska, Hawaii, Oregon, Washington	2,015	5.3	4.9	2,527	5.1	4.9
California	6,218	16.3	21.4	7,942	16.2	21.7
Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming	3,377	8.8	6.1	4,388	8.9	6.2
Iowa, Kansas, Minnesota, Missouri, North Dakota, Nebraska, South Dakota	2,352	6.2	6.5	3,092	6.3	6.5
Illinois, Indiana, Michigan, Ohio, Wisconsin	4,241	11.1	11.3	5,729	11.6	11.3
Texas	3,585	9.4	6.4	4,711	9.6	6.4
Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Oklahoma, Tennessee	3,000	7.9	7.6	3,913	7.9	7.6
Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, Vermont	1,202	3.1	5.7	1,553	3.2	5.6
New York	1,534	4.0	6.2	1,950	4.0	6.0
Pennsylvania, New Jersey, Delaware	2,066	5.4	6.6	2,451	5.0	6.5
District of Columbia, Maryland, Virginia	1,729	4.5	5.4	2,124	4.3	5.3
Georgia, North Carolina, South Carolina	3,533	9.3	6.0	4,573	9.3	6.1
Florida	3,341	8.7	5.9	4,315	8.8	5.9
Total	38,193	100	100	49,268	100	100

Notes: HB and QCEW data for February 2019 and 2020. The columns titled "#" report the number of establishments in HB data. The columns titled "%" shows the distribution of establishments in HB data and according to QCEW data.

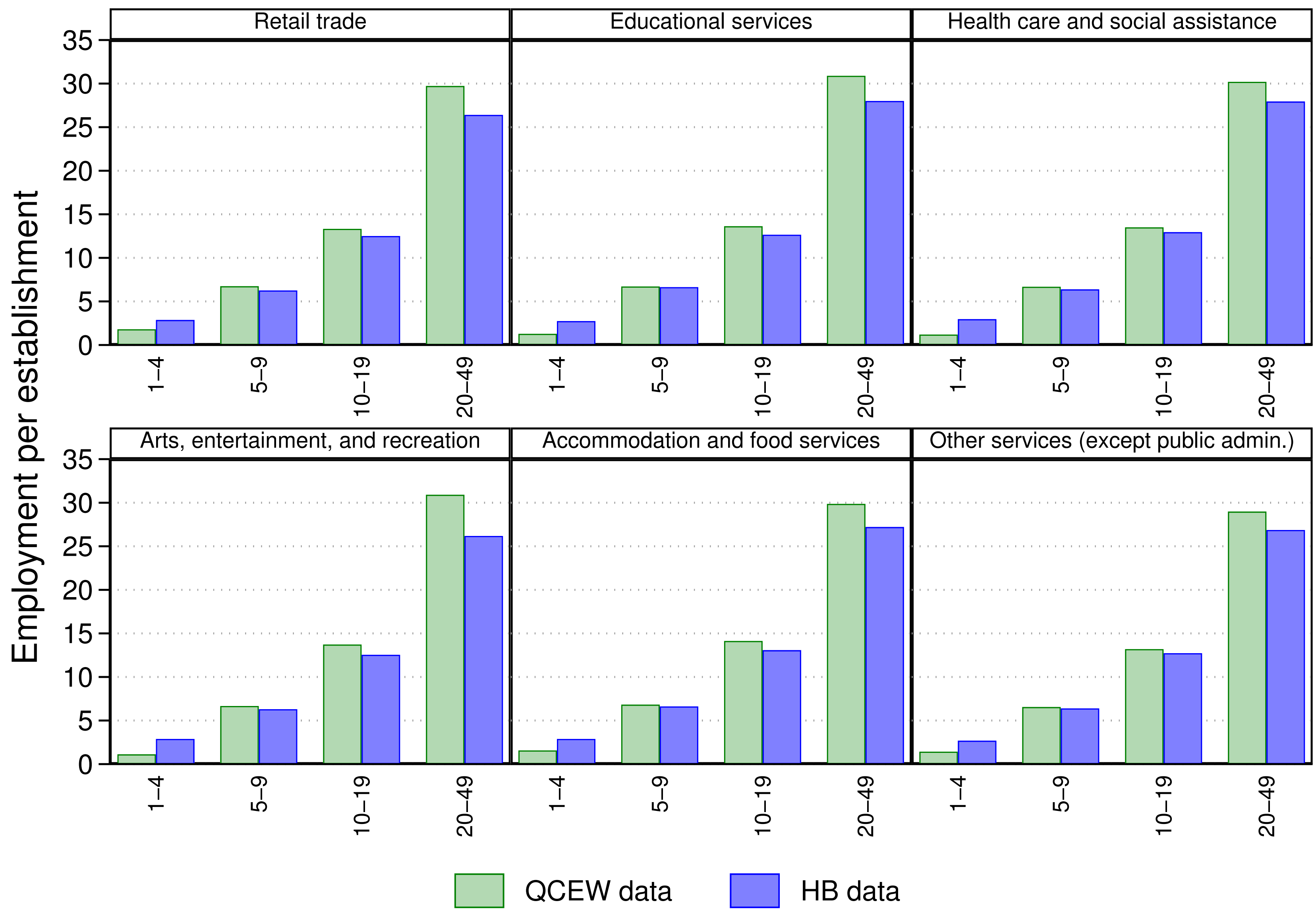
for 15% of establishments with fewer than 50 workers in the sectors of our analysis. Within NAICS 72, the distribution across size classes in HB is, again, less concentrated towards small businesses compared to QCEW data. In terms of employment, NAICS 72 accounts for an even larger share of the HB sample, namely 60% of employment in February 2020.

Table A4 reports the distribution of businesses with respect to U.S. states or groups of states. In grouping states together, we attempt to strike a balance between having 2-digit NAICS × size × region cells with sufficiently many HB establishments (as compared with QCEW data), while preserving some geographic variation. We group together states that are geographically close and, as it turns out, that have experienced similar employment outcomes during the crisis. As shown in Table A4, the geography distribution of HB data in both 2019 and 2020 is similar to the distribution of establishments according to the QCEW.

Figure A3 reports average employment by establishment within each establishment class size in HB data and in the QCEW in 2020. Establishments tend to be slightly larger (in employment terms) in the QCEW data, except in the first class size where establishments are always larger in HB data.[10] But

[10]Again, this difference could be related to the larger establishments counts of the QCEW in size class 1–4. According to

Figure A3: Average employment by establishment by establishment size



Notes: Average employment by establishment size class for NAICS-2 industries in HB data and in QCEW data for February 2020.

overall, the data match closely.

B Safegraph and Facebook data

B.1 Safegraph

We use Safegraph (SG) data on business locations, called Points of Interest (POIs) to attribute NAICS industry codes to HB establishments. SG contains information on more than 8 million POIs in the U.S. A POI is defined as a location where individuals can spend time or money. For a subset of POIs, SG reports anonymized visits at daily, weekly, and monthly frequency based on information from cell phone devices. The information in Safegraph is organized in three main datasets.

- **Core Places** contains basic information for every establishment including name, address, GPS coordinates (lat/lon), NAICS industry code, brand, etc. This is the main frame based on which Safegraph builds the other datasets.
- **Patterns** contains data for a subset of establishments including visit counts, visit duration, and mapping of visitors to their home Census block group.
- **Geometry** contains spatial hierarchy information for a subset of establishments. This information is important to understand and qualify the accuracy of the visits data.

We note that the Core Places files contain separate `opened_on` and `closed_on` fields that are supposed to carry the opening date of a new POI, respectively the closing date for a permanently closed establishment. However, as Safegraph acknowledges in its documentation, coverage and accuracy of this information is far from perfect. We confirm this through the following checks: (i) there is bunching of opening and closing in particular months, (ii) we find discrepancies between the `closed_on` field and both our HB data and Google's permanently closed indicator, (iii) the rates of business openings and closings implied by SG's `opened_on` and `closed_on` fields are too low compared to BED data. We therefore do not rely on SG's `opened_on` and `closed_on` variables.

Description and definitions. Each POI is identified by a unique persistent `safegraph_place_id`. This is our main unit of observation when working with SG data. In some cases, which are rare in

Barnatchez et al. [2017], one reason why these counts are larger is that QCEW includes some non-employers in establishments counts, which would bias downward the ratio between total employment and number of establishments. In BDS data, average employment per establishment in size class 1–4 is 2.4 workers (vs. 1.4 in the QCEW), which is very close to the HB data in Figure A3.

our matched dataset, a POI may appear under multiple `safegraph_place_id`'s. This happens when SG changes details about some time-invariant attribute of the POI, such as the address or the NAICS code. We identify duplicated `safegraph_place_id`'s through a deduplication procedure described in the paragraph "Deduplication of POIs" below. We keep the time-invariant attributes provided in the most recent release of the Core Places data for POIs that are duplicate of each other. We combine visits data to duplicated POIs by adding them up.

Industry codes and visits. Except for a handful of POIs (about 1% of the universe of SG POIs), each `safegraph_place_id` comes with a 6-digit industry NAICS code that SG attributes based on an algorithm. For details about SG's methods, see this Documentation.

About 80-85% of `safegraph_place_id`'s come with information on visits. Safegraph constructs visits data by attributing cell-phone pings (visits) to a POI's polygon.[11] In Section E, we use a weekly aggregate of visits, which is provided in the Weekly Patterns file, to run several checks on selection into Homebase, to compare the dynamics of Homebase employment with visits, and to contrast the visits data of continuing establishments vs. entry and exit churners in Homebase. SG weekly visits are also available in a "bucketed dwell time" format, which yields similar results to the results presented in those sections.

Deduplication of POIs. Our algorithm to identify and deduplicate Safegraph POIs is as follows:

1. Find all POIs that have the same location name (normalized using Step 1a of our matching algorithm for SG data; see Section C) and same GPS coordinates rounded up to one decimal place. This defines sets of *potential* duplicate POIs.

2. Within each set of potential duplicates, perform all pairwise comparisons to identify subsets of POIs that are duplicates of each other. Specifically:

   (a) Given two POIs that belong to the same set, compute the geographic distance between them using GPS coordinates, and the string distance between their street addresses (normalized using Step 1b of our SG matching algorithm; see Section C) concatenated with the 5-digit zip code. We use Levenshtein distance normalized by the length of the longest string to define string distance. If either the geographic distance is less than 250 meters or the string distance is under 0.250, we tag the two POIs as duplicates of each other.

[11]Safegraph's documentation provides information about spacial hierarchy for each polygon that is important for visit attribution: see Places Manual as well as this blog post for details.

(b) Form subsets of all POIs that are direct duplicates of each other. Indirect duplicates, as opposed to direct duplicates, refer to instances such as this one: A is a duplicate of B but not a duplicate of C, but B is a duplicate of C, making A and C indirect duplicates of each other. A and B could be included in the same subset while C is left aside. Alternatively, A could be left aside while B and C are included in the same subset. In such instances (which are very rare, as we explain below) we break the tie by assigning B to the closest POI (A or C) as measured by the geographic and Levenshtein distances.

3. For POIs that are duplicates of each other, we construct a `dedup_safegraph_place_id` to overwrite their time-invariant attributes (address, zip code, and most importantly NAICS codes) with those of the `dedup_safegraph_place_id`. We set `dedup_safegraph_place_id` equal to the `safegraph_place_id` of the POI that (i) was seen most recently in the Core place files and (ii) has a missing `closed_on` field (since otherwise SG will no longer attribute visits to this POI).

4. In our deduplicated SG data, visits for a POI that has duplicates in the raw SG data are taken to be the sum of the POI's own visits and visits at its duplicates POIs.

In Step 1 of our deduplication algorithm, we have 1,105,224 POIs (about 15% of SG Core place POIs) that belong to a set with more than one POI. 70% of these sets contain exactly two POIs and another 15% have only three POIs. The largest set contains 281 POIs. After Step 2, 476,770 POIs belong to a subset with more than one POI, meaning that roughly 50% of POIs from Step 1 have been found to have no duplicate. There are 197,280 subsets with more than one POI, 80 percent of which contain exactly two POIs and 10 percent of which contain exactly three POIs.

B.2 Facebook

We use information on Facebook (FB) posting activity by establishments in HB data to estimate new business openings and closings, as described in Section D. The FB data comes from CrowdTangle (https://www.crowdtangle.com/), FB's research tool to analyze social media activity.

Description and definitions. We use Google to search for a Facebook URL address for each HB establishment. A Facebook URL can be linked to a unique `facebookid`, which we can then use to retrieve information on FB posts. However, there are some caveats:

- In some cases, we obtain the same Facebook URL for several establishments. It so happens when different establishments that belong to the same company use and manage the same Facebook page. We drop these establishments from the analysis of FB posts at a later stage, upon getting information on posting activity;

- If the Facebook account is deactivated by the user, CrowdTangle no longer tracks the page's activity. When it so happens, we cannot attribute a `facebookid` to the establishment.

We access CrowdTangle information using two main datasets:

- **Leaderboard**, which contains information for every uploaded account, including a user name, a unique `facebookid`, page growth, etc.
- **Historical data**, which contains posting data for a subset of `facebookid` from the Leaderboard dataset.

Historical posts. CrowdTangle tracks public content, including the date when a users posts some content on Facebook, and from which `facebookid` the content is posted. We obtain historical posting data of all establishments with a valid Facebook URL in our sample. We aggregate counts of FB posts to the weekly frequency. We define a HB establishment as being an active user of FB if its posting history averages at least one post per week during the weeks when the establishment is also active in HB.

C Matching procedures

We augment HB's establishment records with information from Google and then match the records with SG's POIs based on name and geography. In addition, for the purpose of determining birth and death, we match the data with information from CrowdTangle. Below we describe both match algorithms and provide basic match statistics.

C.1 Matching with Safegraph

To match HB locations to Safegraph, we consider the entire catalog of SG POIs that ever appear in the Core place files between March 2020 and November 2022. The total number of SG POIs is 8,537,035 – out of which we remove duplicates, as described in Step 1 below. Our algorithm to match HB locations is as follows:

1. Pre-treat the data by cleaning and standardizing names, and deduplicate SG data:

(a) Clean company and location names in HB, and location names in Safegraph and Google by: (i) removing company titles, such as "inc", "incorporated", "corp", "corporation", "llc", (ii) removing "and" and "the", and (iii) removing any spaces and keeping only numeric and alphabetic characters.

(b) Clean addresses in HB, Safegraph and Google using Stata's `stnd_address` command. Then standardize addresses and city names by removing any spaces and keeping only numeric and alphabetic characters.

(c) Deduplicate SG data using the procedure described in Section B.1.

2. Merge or match HB data:

(a) Merge/Match is performed using 3 possible names for the establishment: HB location name, HB company name, and Google name retrieved using the Google place identifier.

(b) Try to merge using each of the name (sequentially in this order: HB location name, HB company name, and Google name) combined with the following information (again, sequentially in this order): (i) Latitude and longitude (rounded up to 3 decimal places), (ii) Address and zip code, (iii) Address and city, (iv) Address and State, (v) Address only.

(At each level of the merge, we only keep the unique merges: i.e., a HB establishment gets linked to a unique Safegraph POI. We discard merges when a HB establishment merges to more than one Safegraph POI.)

(c) Try to match using each of the names and the geographical information as described in the previous step. A matching score is assigned to each pair of HB establishment and Safegraph POI, representing how similar their names are. We consider matching to be successful if the HB establishment and the Safegraph POI have (1) the same geographical information, and (2) a matching score of 80 or higher.

(d) Try to merge using each of the name and broader geographical information (sequentially in this order): (i) Zip code, (ii) City, (iii) State.

(e) Try to match using each of the name and broader geographical information as described in the previous step.

Table C1 presents the outcomes of the algorithm for the pooled 2019 and 2020-2021 data. 45% of the sample is made up of locations that we *merge* to SG based on names and GPS coordinates. Table C1 also shows that there is some variation by sector, in that the high-level merge/matches (those based on the name combined with either GPS coordinates or street address) are likely to be in Leisure & Hospitality while the low-quality matches (those based on the name combined with either city or state) are more likely to be in Retail Trade.

Table C1: Results of matching HB establishment to Safegraph

	Sample		NAICS sector (%)			
	#	%	44-45	61-62	71-72	81
Merge on name and GPS coordinates	46,925	44.8	22.3	7.8	63.9	6.1
Merge on name and address	3,221	3.1	23.8	13.2	56.9	6.2
Match on name and address	12,772	12.2	21.8	12.0	59.7	6.5
Merge on name and zip code	6,488	6.2	24.9	8.7	59.9	6.6
Merge on name and city	2,828	2.7	28.4	12.6	51.3	7.8
Merge on name and state	5,607	5.4	31.5	14.6	43.2	10.7
Match on name and zip code	2,147	2.1	26.5	17.8	44.5	11.2
Match on name and city	5,044	4.8	31.7	21.6	33.5	13.2
Match on name and state	17,644	16.9	37.2	18.2	31.6	12.9
Others	2,039	2.0	34.4	15.1	45.8	4.7
Total	104,715	100	26.4	11.8	53.8	8.1

Notes: The table reports counts (#) and distribution (%) of HB locations in the 2019 base sample 2020 base sample, and all new entrants (entering after mid-February 2019 but no later than end of November 2021) across the outcomes of the algorithm for matching HB locations to Safegraph, and distribution across sectors for the base sample and new entrants sample combined. The last outcome category ("Others") refers to HB locations that match to Safegraph with a low quality but have a SG NAICS code that matches the industry code provided in the raw HB data; see Footnote 8 for details. The four sectors are Retail Trade (NAICS 44-45), Education and Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81).

C.2 Matching with CrowdTangle

Our algorithm to match HB establishments to CrowdTangle is as follows:

1. Pre-treat the data by cleaning and standardizing names:

   (a) Clean company and location names in HB, and location names in Google by: (i) removing company titles, such as "inc", "incorporated", "corp", "corporation", "llc", (ii) removing "and" and "the", and (iii) removing any spaces and keeping only numeric and alphabetic characters.

   (b) Clean addresses in HB and Google using Stata's `stnd_address` command. Then standardize addresses and city names by removing any spaces and keeping only numeric and alphabetic characters.

2. Use Google to find a Facebook address for each HB location, and clean the Facebook address by

   (a) Removing links that are posted in public pages or groups, such as "event", "careers", "places", "marketplace", etc.;

(b) Extracting the Facebook page name from the Facebook address, e.g. extract the end part of `http://www.facebook.com/page_name`.

3. Batch upload locations with a valid Facebook URL into CrowdTangle.

4. Merge or match the resulting CrowdTangle dataset to HB data. Specifically,

   (a) From CrowdTangle, we obtain the "Leaderboard" dataset (Section B.2), which contains a unique user name and `facebookid` for each Facebook account;

   (b) Try to merge the FB `page_name` obtained in Step 2b to CrowdTangle's user name ;

   (c) Otherwise, try to match the FB `page_name` to CrowdTangle's `facebookid`.

Table C2 presents the outcomes of matching HB establishments to CrowdTangle for the pooled 2019 and 2020-2021 data. Most HB establishments can be linked to a valid Facebook URL: more than 90% for both new entrants and permanent exits during this period. We manage to upload almost 30% of new entrants and permanent exits to CrowdTangle, either by merging or matching on `page_name`. Note, however, that we only use a small subset of these establishments in our analysis of FB posts (see Tables D1 and D2) because many establishments are not actively posting on FB.

Table C2: Results of matching HB establishment to CrowdTangle

	New entrants		Exits without return	
All estabs.	62,926	(100%)	41,545	(100%)
Establishments with a Facebook URL	57,491	(91.4%)	37,970	(91.4%)
Establishments uploaded to CrowdTangle	17,698	(28.1%)	10,632	(25.6%)

Notes: The table shows counts of establishments that newly enter HB after mid-February 2019 and no later than end of November 2021 (new entrants), and establishments that exit HB with no return to activity before the end of November 2021 (exits without return). For these establishments, the table reports counts of establishments with a valid Facebook URL and establishments that successfully upload to CrowdTangle.

D Closings and openings

Recall our employment estimator from (F.1)

$$\widehat{E}_t = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}} \right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{C}_{i,t}} \right)}, \tag{D.1}$$

where $\hat{e}_{i,t}^{\mathcal{A}_{i,t}}$ denotes week t employment of the set of establishments $\mathcal{A}_{i,t}$ that were active in both week $t-1$ and week t; $\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}}$ denotes week $t-1$ employment of the set of establishments $\mathcal{C}_{i,t}$ that are closing in week t; and $\hat{e}_{i,t}^{\mathcal{O}_{i,t}}$ denotes week t employment of the set of establishments $\mathcal{O}_{i,t}$ that are opening in week t. Using $\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}} = \hat{e}_{i,t-1}^{\mathcal{T}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{D}_{i,t}}$ and $\hat{e}_{i,t}^{\mathcal{O}_{i,t}} = \hat{e}_{i,t}^{\mathcal{R}_{i,t}} + \hat{e}_{i,t}^{\mathcal{B}_{i,t}}$, this estimator can be written as

$$\widehat{E}_t = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{R}_{i,t}}\right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{T}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{D}_{i,t}}\right)} + \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \hat{e}_{i,t}^{\mathcal{B}_{i,t}}}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{T}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{D}_{i,t}}\right)}, \tag{D.2}$$

where $\hat{e}_{i,t-1}^{\mathcal{T}_{i,t}}$ denotes week $t-1$ employment of the set of establishments $\mathcal{T}_{i,t}$ that closed temporarily in week t; $\hat{e}_{i,t-1}^{\mathcal{D}_{i,t}}$ denotes week $t-1$ employment of the set of establishments $\mathcal{D}_{i,t}$ that closed permanently (i.e. deaths) in week t; $\hat{e}_{i,t}^{\mathcal{R}_{i,t}}$ denotes week t employment of the set of establishments $\mathcal{R}_{i,t}$ that reopen in week t after being temporarily closed; and $\hat{e}_{i,t}^{\mathcal{B}_{i,t}}$ denotes week t employment of the set of establishments $\mathcal{B}_{i,t}$ that newly open (births) in week t. Notice that establishments that cease to be active in week t and become active again at some later date $t+n$ belong to the set of temporarily closed establishments $\mathcal{T}_{i,t} \subseteq \mathcal{C}_{i,t}$ in week t, and later on they are added to the set of re-opening establishments $\mathcal{R}_{i,t+n} \subseteq \mathcal{O}_{i,t+n}$.[12] The key challenge facing the implementation of (D.1) is that sample churn prevents us from directly observing $\mathcal{D}_{i,t}$ and $\mathcal{B}_{i,t}$. That is, among the establishments that cease to be active in some week t and are not part of $\mathcal{T}_{i,t}$, some continue to operate outside of HB and must therefore not be included in $\mathcal{D}_{i,t}$, Likewise, among establishments that open during week t and do not belong to $\mathcal{R}_{i,t}$, those establishments that were operating before entry into HB in week t must not be included in $\mathcal{B}_{i,t}$. The next sections present our methodology to address this challenge. Section D.1 explains how we use Google and Facebook to disentangle sample churn from (permanent) closings and (new) openings, while Section D.2 explains how we subsequently incorporate this information for the estimation of $\hat{e}_{i,t-1}^{\mathcal{D}_{i,t}}$ and $\hat{e}_{i,t}^{\mathcal{B}_{i,t}}$.

D.1 Google/Facebook approach to determine business closings and openings

D.1.1 Permanent closings. We use information from Google and Facebook to estimate whether establishments that exit HB and do not re-enter before the end of the sample are closed or continue to operate outside of HB. We match HB establishments to Google Places using their API. At the same time,

[12] The maintained assumption is that these establishments do not continue to operate while temporarily abstaining from using the HB service. Notice that as t gets closer to the end of the sample, this approach implies that we miss some temporary closings. This is not a problem as long as the closing probability (subsection D.1) correctly identifies these establishments as being closed, and since the employment estimator in Equation (D.1) does not require us to distinguish between temporary and permanent closings. By the same token, the definition of temporarily closed establishments implies that past real-time employment estimates are subjected to revisions upon using more recent data to estimate equation (D.1). In practice, revisions of our estimates have been unimportant.

we use Google to search for a Facebook address for each HB establishment based on name and address and then use CrowdTangle, Facebook's research utility, to extract the history of posts for each available Facebook address. We then proceed in 3 steps:

1. For HB establishments that can be matched to Google Places, we identify establishment ℓ that exits HB in week as closed if Google attributes a "closed" indicator.

2. For HB establishments that either cannot be matched to Google Places or are not flagged as Google-closed, we retain all establishments with a unique Facebook address that average at least one post per week during the weeks when they are active in HB. Then, for any establishment ℓ that exits during week t and satisfies these criteria,

   (a) if Facebook posts continue for more than 4 weeks after the establishment exits HB, we identify it as an establishment that continues to operate outside of HB;

   (b) otherwise, we identify establishment ℓ as a business closing.

3. For all other exiting establishments that cannot be matched to either Google Places or Facebook, we identify them as closed with probability equal to the proportion of closings estimated in Step 2. We compute this proportion separately by quarter q for each of the four sectors i.

Concretely, the procedure means that for all establishments ℓ belonging to cell i that exit HB permanently in week t of quarter q, we define:

$$\hat{p}(\mathcal{D}_{\ell,t}|\text{exit}_{\ell,t}) \begin{cases} = 1 \text{ if } \ell \text{ is Google-closed, or not Google-closed but matched to FB and closed} \\ \quad \text{based on FB posts} \\ = 0 \text{ if } \ell \text{ is not Google-closed but matched to FB and operating outside of HB} \\ \quad \text{based on FB posts} \\ = \text{cell-}i \text{ probability } (\in [0,1]) \text{ of closing conditional on exit from HB in the} \\ \quad \text{current quarter } q \end{cases}$$

$\hat{p}(\ell \in \mathcal{D}_{\ell,t}|\text{exit}_{\ell,t})$ is a key ingredient of our estimation approach, summarized in Table D3 at the end of this section. Notice that we rely on the set of establishments that are either Google-closed or matched to Facebook to compute cell-i's probability of closing conditional on exit from HB. When compared to establishments that permanently exit HB and are neither Google-closed nor matched to Facebook, we find no evidence of systematic differences between the two sets of establishments: they share similar industry-size-region distributions and have similar employment and hours dynamics while active in HB.[13] Since we have few permanent exits from HB that are either Google-closed or matched to Facebook in a given *week*, we pool these establishments together by quarter to compute cell-i's probability of closing conditional on exit from HB.

Table D1: Google / Facebook procedure to determine closings

	2019		2020	
Exiting estab. that do not reopen	13,289	(100%)	28,256	(100%)
- Google closed	3,011	(22.7%)	4,031	(14.3%)
- Not Google-closed and matched to FB	2,197	(16.5%)	6,826	(24.2%)
- Closed from FB posts	419	(3.2%)	2,000	(7.1%)
Proportion estimated as closed	37.4%		39.0%	

Notes: For 2019, exiting establishments that do not reopen refer to the set of establishments included in the base sample or new entrants sample of 2019 that cease to be active at some point before mid-February 2020 and do not return to activity in HB by the end of the sample (currently end of November 2021). For 2020, exiting establishments that do not reopen refer to the set of establishments included in the base sample or new entrants sample of 2020 that cease to be active and do not return to activity in HB by the end of the sample. Establishments that are Google-closed are establishments that can be matched to Google Places and are flagged as either "temporary closed" or "permanently closed". Establishments that can be matched to Facebook are establishments that can be matched to Facebook pages, do not belong to multi-establishment `company_id`'s and post on average at least once a week while active in HB. Establishments matched to Facebook are flagged as "closed" if their Facebook posts do not continue for more than 4 weeks after exit from HB. The proportion estimated as closed is the proportion of Google closed plus the proportion of establishments that are either identified (if they are not Google-closed and are included in the FB-HB matched sample) or estimated as closed based on the FB estimation.

Table D1 provides statistics on the procedure to deal with permanent closings. Consider for instance the 13,289 establishments that exit HB in 2019 without return before the end of the sample (currently end of November 2021). 22.7% of these establishments are estimated as closed because they receive a "closed" flag when matched to Google Places (Step 1 of the above procedure). The remaining establishments that can be matched to Facebook and are actively posting while being active in HB account for 16.5% of all permanent exists in 2019. Our procedure flags 19.1% of these establishments (419/2,197) as "closed" because their Facebook posts stop within less than 4 weeks after exit from HB. For 60.8% of exiting establishments in 2019 (100% minus 22.7+16.5%), the Google/Facebook approach is not applicable (Step 3 of the above procedure), and we estimated that 19.1% of these establishments are closed (based on the proportion of "closed" among establishments matched to FB). Put together, these numbers imply that the proportion estimated as closed among all 2019 exits from HB is 37.4%.[14]

[13]For the most part, it seems that these establishments cannot be matched to either Google Places or Facebook due to idiosyncratic details in company names or addresses that our algorithm fails to take into account (see Section C.2), or due to inexistent or erratic posting behaviors on Facebook that prevent us from relying on this source of information (as we use Facebook only for establishments that average at least one post per week while in HB).

[14]Table D1 in addition shows that we identify 13.4% of exiting establishments in 2019 as sample churn ($\ell \notin \mathcal{D}_{i,t}$) because

Table D1 shows that in 2020 our Google / Facebook approach yields a similar proportion of closings conditional on exit from HB, namely 39.0% in 2020 against 37.4% in 2019. There are fewer "temporary closed" or "permanently closed" based on the information obtained from Google places, and a higher proportion of establishments matched to Facebook, among which we flag 29.3% (2,000/6,826) of the establishments as closed based on their Facebook posts.

D.1.2 New openings. For establishments that become active in HB for the first time after the reference week (i.e. new entries, not reopenings), we proceed similarly as for closings. One difference is that we only exploit the information coming from Facebook posts since Google does not contain an indicator for new openings as it does for closings. We use Google to search for a Facebook address for all newly entering HB establishments based on name and address and then use CrowdTangle to extract the history of posts for each available Facebook address. We then proceed in 2 steps:

1. We retain all newly entering establishments with a unique Facebook address that average at least one post per week during the weeks when they are active in HB. Then, for any establishment ℓ that enters HB in week t and satisfies these criteria,
   (a) if Facebook posts start before the base period (mid-February of the respective year), we identify establishment ℓ as one that operated already prior to entering HB;
   (b) otherwise, we identify establishment ℓ as a new opening.
2. For all other newly entering establishments, we identify them as new openings with probability equal to the proportion of new openings estimated in Step 1.

For all establishments ℓ that become active in cell i for the first time in week t, this procedure yields:

$$\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t}) \begin{cases} = 1 \text{ if } \ell \text{ is matched to FB and new based on FB posts} \\ = 0 \text{ if } \ell \text{ is matched to FB and operating prior to entry based on FB posts} \\ = \text{cell-}i \text{ probability } (\in [0,1]) \text{ of birth conditional on entry into HB in the} \\ \quad \text{current quarter } q \end{cases}$$

As for closings, we find little evidence that those establishments that can be matched to Facebook and post actively while being active in HB are a selected sample among the set of all new entrants in HB, in the sense that they have similar industry-size-region distributions as well as similar employment and hours dynamics while active in HB.

they either do not receive a "closed" indicator from Google, and they post actively in Facebook and continue to do so for more than 4 weeks after exit from HB.

Table D2 provides statistics on the procedure for new openings. Of the 25,149 establishments that newly entered in 2019, 31.7% can be matched to Facebook and post regularly while being active in HB. Then, we estimate the proportion of new establishments among newly entering HB establishments to be equal to 9.1% (727 divided by 7,960 establishments). Notice that the remaining establishments matched to FB that were posting before the base period (7,233 establishments, or 28.8% of the sample of newly entrants in 2019) are considered to be part of sample churn. In 2020, the Facebook approach can be implemented for a lower portion of newly-entering establishments (25.3%), and the proportion that is estimated as new conditional on entry in HB (7.5%) is very similar to that for 2019.

Table D2: Facebook procedure to determine new openings

	2019		2020	
Newly entering establishments	25,149	(100%)	37,777	(100%)
- Matched to FB	7,969	(31.7%)	9,566	(25.3%)
- Newly opened from FB posts	726	(2.9%)	718	(1.9%)
Proportion estimated as new	9.1%		7.5%	

Notes: For 2019, newly entering establishments refer to the set of establishments that are active in HB for the first time after mid-February 2019 but no later than mid-February 2020. For 2020, newly entering establishments refer to the set of establishments that are active in HB for the first time after mid-February 2020 but no later than late-November 2021. Establishments matched to Facebook includes all establishments that can be matched to unique Facebook pages and post on average at least once a week while active in HB. Establishments matched to Facebook are flagged as "new" if their Facebook posts do not start before mid-February of the corresponding week. The proportion estimated as new is the number of newly opened establishments based on FB posts divided by the number of newly entering, FB-HB matched establishments.

D.2 Incorporating closings and openings into our estimation

Having described the identification of closings and reopenings / new openings, we now explain how this information is incorporated into our employment estimator. For establishments identified as temporary closings, respectively reopenings, we can directly measure $\hat{e}_{i,t-1}^{\mathcal{T}_{i,t}}$ and $\hat{e}_{i,t}^{\mathcal{R}_{i,t}}$. To estimate employment losses from establishments closing permanently $\hat{e}_{i,t-1}^{\mathcal{D}_{i,t}}$ (deaths) and employment gains from newly opened establishments $\hat{e}_{i,t}^{\mathcal{B}_{i,t}}$ (births), we need to perform several adjustments as described in what follows.

D.2.1 Employment at establishments closing permanently. We estimate employment for permanent closings (death) in industry-size cell i in week t as

$$\hat{e}_{i,t-1}^{\mathcal{D}_{i,t}} = \sum_{\ell \in i} \theta_{\ell,t}^{\mathcal{D}} \times \hat{p}(\mathcal{D}_{\ell,t}|\text{exit}_{\ell,t}) \times \hat{e}_{\ell,t-1}, \tag{D.3}$$

where $\hat{e}_{\ell,t-1}$ denotes employment of exiting establishments in the week prior to exit from HB; $\hat{p}(\ell \in \mathcal{D}_{i,t}|\text{exit}_{it})$ denotes the probability estimated from the above Google / Facebook approach; and $\theta^{\mathcal{D}}_{\ell,t}$ is an adjustment factor that corrects for possible selection issues, namely that the survival probability of the average HB business may differ systematically from the population survival probability of the average small business and that survival rates conditional on exit as implied by our Google / Facebook approach may differ systematically from survival rates of the average HB business.

The adjustment factor $\theta^{\mathcal{D}}_{\ell,t}$ is calculated to fit the unconditional average death rate for 2019 for cell i in the BED/BDS. Specifically, for each quarter q, we compute the unconditional death rate in our HB data as

$$\hat{p}_{\text{HB}}(\text{death}_{i,q}) = \frac{\hat{N}^{\mathcal{D}}_{i,q}}{\frac{1}{2}\left(\hat{N}^{\mathcal{A}}_{i,q} + \hat{N}^{\mathcal{A}}_{i,q-1}\right)}, \tag{D.4}$$

where $\hat{N}^{\mathcal{D}}_{i,q}$ is the count of HB establishments that permanently close in quarter q as implied by our Google / Facebook approach (i.e., establishment ℓ is multiplied by its $\hat{p}(\mathcal{D}_{\ell,t}|\text{exit}_{\ell,t})$); and $\hat{N}^{\mathcal{A}}_{i,q}$ and is the count of HB establishments active in quarter q. We define these counts analogous to how they are defined by the BLS to construct quarterly BED death rates (see Section D.4). In particular, $N^{\mathcal{A}}_{i,q}$ is the count of all establishments with positive employment in the third month of quarter q; and $\hat{N}^{\mathcal{D}}_{i,q}$ is defined as the count of establishments with positive employment in the third month of quarter $q-1$ but not in the third month of quarter q, and estimated to represent a permanent closing according to Google / Facebook. Then, we average the resulting quarterly death rates for 2019, and compute the adjustment factors as

$$\theta^{\mathcal{D}}_{i} = \frac{\frac{1}{4}\sum_{q\in 2019}\hat{p}_{\text{BED/BDS}}(\text{death}_{i,q})}{\frac{1}{4}\sum_{q\in 2019}\hat{p}_{\text{HB}}(\text{death}_{i,q})}, \tag{D.5}$$

where $\hat{p}_{\text{BED/BDS}}(\text{death}_{i,q})$ denotes the BED/BDS quarterly death rate for industry-size cell i in quarter q.[15] Finally, we set $\theta^{\mathcal{D}}_{\ell,t} = \theta^{\mathcal{D}}_{i}$. Thus, by construction of the $\theta^{\mathcal{D}}_{\ell,t}$, the adjusted HB death rates (i.e., $\sum_{\ell\in i}\theta^{\mathcal{D}}_{\ell,t}\times\hat{p}(\mathcal{D}_{\ell,t}|\text{exit}_{\ell,t})$) averaged over the four quarters of 2019 matches the average BED/BDS quarterly death rates over this period.

To allow for sufficient sample size, we compute the adjustment factor $\theta^{\mathcal{D}}_{i}$ for Retail Trade, Education & Health Services, and Other Services separately for establishment size 1 to 4 and establishment size 5-10, 11-19, and 20-49 pooled together. As shown in the main text, the fit with the BED death rate is excellent for all size classes despite this pooling. For Leisure & Hospitality, we compute $\theta^{\mathcal{D}}_{i}$ separately for all size classes, both because our sample is larger and because death rates vary more substantially across

[15]BED death and birth rates are available by industry but not by size class. We use information from the BDS to derive size-adjusted BED death and birth rates; see Section D.4.

size classes 5-10, 11-19, and 20-49 than for the other sectors. See the main text for results.

D.2.2 Employment at newly opened establishments. Conceptually, employment gains from new openings (births) in industry-size cell i in week t can be estimated similarly as employment losses from permanent closings; i.e.

$$\hat{e}_{i,t}^{\mathcal{B}_{i,t}} = \sum_{\ell \in i} \theta_{\ell,t}^{\mathcal{B}} \times \hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t}) \times \hat{e}_{\ell,t} \tag{D.6}$$

where $\hat{e}_{\ell,t}$ denotes employment of entering establishments during the week of entry into HB, $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t})$ denotes the estimated probability of new openings (births) conditional on entry into HB obtained from Facebook; and $\theta_{\ell,t}^{\mathcal{B}}$ is the adjustment factor. However, the computation of $\theta_{\ell,t}^{\mathcal{B}}$ is more involved than that of $\theta_{\ell,t}^{\mathcal{D}}$ because entry of establishments in HB may vary in ways that are not necessarily taken into account by corresponding changes in our Google / Facebook probability of new openings, $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t})$.

To illustrate this issue and motivate our strategy to compute $\theta_{\ell,t}^{\mathcal{B}}$, consider the hypothetical case where Homebase samples randomly from the population such that $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t}) = p(\mathcal{B}_{\ell,t})$; i.e. there are no selection issues. But then, according to Equation (D.6), employment from new openings $\hat{e}_{i,t}^{\mathcal{B}_{i,t}}$ in week t would be entirely driven by $\sum_{\ell} p(\mathcal{B}_{\ell,t})\hat{e}_{\ell t}$, which in turn is driven by the number of new entrants during that week. This number may vary independently of the size of the sample used to run the estimation of (D.1) as a result of, e.g., changes in HB's efforts to attract new customers or changes in HB's competitive environment.[16] In the present case, we can use an approach akin to inverse-probability weighting to adjust for this issue. Suppose that multiply $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t}) \times \hat{e}_{\ell,t}$ by the inverse of the ratio of new entrants to existing HB establishments (normalized by this ratio at $t = 0$), $\frac{\hat{n}_{i,t}^{\text{entry}}/\hat{n}_{i,t}^{\mathcal{A}}}{\hat{n}_{i,0}^{\text{entry}}/\hat{n}_{i,0}^{\mathcal{A}}}$. Then, employment from new openings would be calculated as $\hat{e}_{i,t}^{\mathcal{B}_{i,t}} = \sum_{\ell \in i} (\frac{\hat{n}_{i,t}^{\text{entry}}/\hat{n}_{i,t}^{\mathcal{A}}}{\hat{n}_{i,0}^{\text{entry}}/\hat{n}_{i,0}^{\mathcal{A}}})^{-1} \times \hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t}) \times \hat{e}_{\ell,t}$, or equivalently $\hat{e}_{i,t}^{\mathcal{B}_{i,t}} = \sum_{\ell \in i} \theta_i^{\mathcal{B}} \times p(\mathcal{B}_{i,t}) \times \hat{n}_{i,t}^{\mathcal{A}} \times \hat{e}_{\ell,t}/\hat{n}_{i,t}^{\text{entry}}$, with $\theta_i^{\mathcal{B}} = \hat{n}_{i,0}^{\text{entry}}/\hat{n}_{i,0}^{\mathcal{A}}$ Thus $\hat{e}_{i,t}^{\mathcal{B}_{i,t}}$ would be the estimated number of new births $(p(\mathcal{B}_{i,t}) \times \hat{n}_{i,t}^{\mathcal{A}})$ times average employment of newly entering establishments $(\sum_{\ell \in i} \hat{e}_{\ell,t}/\hat{n}_{i,t}^{\text{entry}})$.

In practice, we also want to adjust for possible selection issues, namely that HB establishments may not represent a random sample and that conditional birth rates as implied by our Google / Facebook approach may differ systematically from survival rates of the average HB establishment. We do by setting $\theta_{\ell,t}^{\mathcal{B}} = \theta_i^{\mathcal{B}} \times (\frac{\hat{n}_{i,t}^{\text{entry}}/\hat{n}_{i,t}^{\mathcal{A}}}{\hat{n}_{i,0}^{\text{entry}}/\hat{n}_{i,0}^{\mathcal{A}}})^{-1}$ and calculating the adjustment factor $\theta_i^{\mathcal{B}}$ to fit the unconditional average birth rate for 2019 for cell i in the BED. We pool class sizes in the same way as for $\theta_i^{\mathcal{D}}$: establishments with 5 to 49 employees are pooled together for Retail Trade, Education & Health Services, and Other Services,

[16] Put differently, while exits are naturally bounded by HB sample size (i.e. the total number of establishments in HB), the upper bound for entry is theoretically the population of establishments not already covered by HB. Hence, differences in HB's efforts to attract new customers or changes by competing service providers may lead to larger fluctuations in entry of new establishments.

while for Leisure & Hospitality we keep the four class sizes separate from each other. We compute the unconditional birth rate of the HB data for each quarter q of 2019 as

$$\hat{p}_{\mathrm{HB}}(\mathrm{birth}_{i,q}) = \frac{\hat{N}^{\mathcal{B}}_{i,q}}{\frac{1}{2}\left(\hat{N}^{\mathcal{A}}_{i,q} + \hat{N}^{\mathcal{A}}_{i,q-1}\right)}, \tag{D.7}$$

where $\hat{N}^{\mathcal{B}}_{i,q}$ is the count of newly entering establishments with positive employment in the third month of quarter q and multiplied by the estimated probability $\hat{p}(\mathcal{B}_{\ell,t}|\mathrm{entry}_{\ell,t})$. One difference with $\theta^{\mathcal{D}}_i$ is that we compute $\theta^{\mathcal{B}}_i$ by regressing the four quarterly values of the BED/BDS birth rates for 2019, $\hat{p}_{\mathrm{BED/BDS}}(\mathrm{birth}_{i,q})$, on the $\hat{p}_{\mathrm{HB}}(\mathrm{birth}_{i,q})$'s. We find that this approach does a better job at controlling for changes in $\hat{p}(\mathcal{B}_{\ell,t}|\mathrm{entry}_{\ell,t})$ induced by the large swings in the number of new entrants in HB in the first and last quarters of the year. While the regression implies that the average of the adjusted HB birth rates will be different from the average BED quarterly birth rates for 2019, Figure 1 in the main text shows that they are very close to each other.

D.3 Recap of our approach

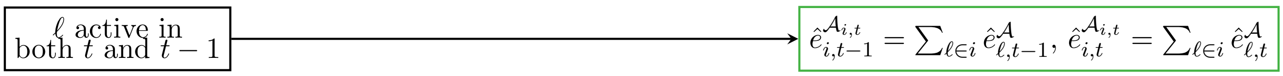
Table D3 recaps and summarizes our approach. For **permanent closings**, which are establishments that do not return at a later date: Each of these establishments is multiplied by the probability $\hat{p}(\mathcal{D}_{\ell,t}|\mathrm{exit}_{\ell,t})$ from our Google / Facebook approach, and multiplied by the industry-size adjustment factor for closing $\theta^{\mathcal{D}}_i$. The adjustment factor is calibrated to match average BED death rates in 2019. For any subsequent period, since $\theta^{\mathcal{D}}_i$ constant, all the variations in closing rates in our estimates come from exits from HB data and from the Google / Facebook indicator and probabilities of permanent closings. As shown by Figure 5 in the main text, the estimates come very close to the BED/BDS death rates for 2020, despite our assumption of constant $\theta^{\mathcal{D}}_i$'s. For **new openings**, we multiply new entrants in HB in a given week t by the Google / Facebook probability $\hat{p}(\mathcal{B}_{\ell,t}|\mathrm{entry}_{\ell,t})$, weight them by the inverse of $\frac{\hat{n}^{\mathrm{entry}}_{i,t}/\hat{n}^{\mathcal{A}}_{i,t}}{\hat{n}^{\mathrm{entry}}_{i,0}/\hat{n}^{\mathcal{A}}_{i,0}}$ to take account of changes in new entries in t relative to the sample size of establishments that contributed to the estimation in week $t-1$, and multiply them by the industry-size adjustment factor for openings $\theta^{\mathcal{B}}_i$. $\theta^{\mathcal{B}}_i$ is calibrated to BED/BDS birth rates in 2019. For any subsequent period, the source of variations in new openings is coming from variations in the Google / Facebook indicator and probabilities and controlling for variations in total counts of new entrants in HB data, $\hat{n}^{\mathrm{entry}}_{i,t}$, relative to $\hat{n}^{\mathcal{A}}_{i,t-1}$.

We now present implementation results on our approach.

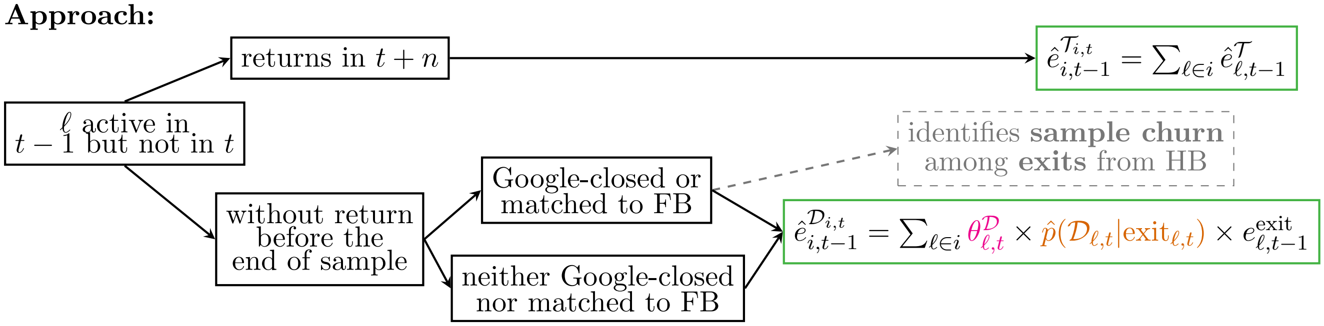
For **permanent closings**, the probability $\hat{p}(\mathcal{D}_{\ell,t}|\mathrm{exit}_{\ell,t})$ derived from Google and Facebook is de-

Table D3: Implementing $\widehat{E}_t = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}}\right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{C}_{i,t}}\right)} = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{R}_{i,t}} + \hat{e}_{i,t}^{\mathcal{B}_{i,t}}\right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{T}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{D}_{i,t}}\right)}$

Approach:

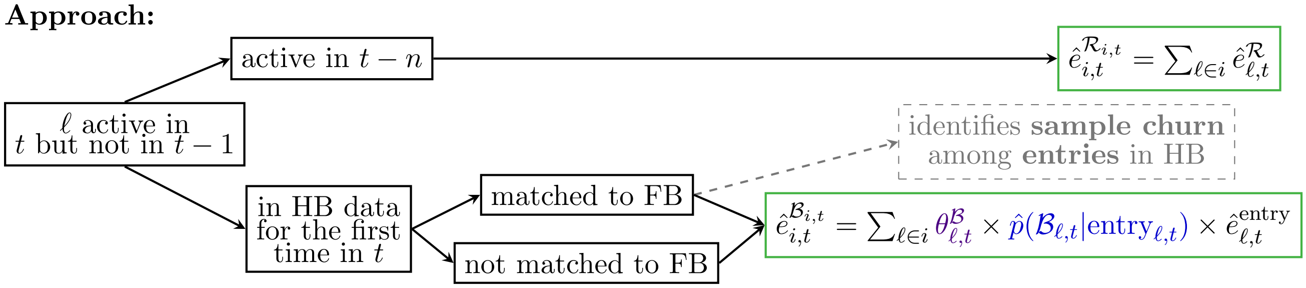


Approach:



Properties:	Varying over time t?	Specific to estab. ℓ?	Specific to cell i?
$\hat{p}(\mathcal{D}_{\ell,t} \| \text{exit}_{\ell,t})$	**Yes** weekly if ℓ is Google-closed or matched to FB	**Yes** by construction of matching at the level of ℓ	**Yes** by construction of ℓ belonging to cell i
	Yes quarterly if ℓ is not Google-closed and not matched to FB	**No**	**Yes** computed at the sector level
$\theta_{\ell,t}^{\mathcal{D}} = \theta_i^{\mathcal{D}}$	**No**	**No**	**Yes** matches sector-size BED 2019 death rates

Approach:



Properties:	Varying over time t?	Specific to estab. ℓ?	Specific to cell i?
$\hat{p}(\mathcal{B}_{\ell,t} \| \text{entry}_{\ell,t})$	**Yes** weekly if ℓ is matched to FB	**Yes** by construction of matching at the level of ℓ	**Yes** by construction of ℓ belonging to cell i
	Yes quarterly if ℓ is not not matched to FB	**No**	**Yes** computed at the sector level
$\theta_{\ell,t}^{\mathcal{B}} = \theta_i^{\mathcal{B}} \times \left(\frac{\hat{n}_{i,t}^{\text{entry}}/\hat{n}_{i,t}^{\mathcal{A}}}{\hat{n}_{i,0}^{\text{entry}}/\hat{n}_{i,0}^{\mathcal{A}}}\right)^{-1}$	**Yes** through $\hat{n}_{i,t}^{\text{entry}}/\hat{n}_{i,t}^{\mathcal{A}}$ computed at the weekly frequency	**No**	**Yes** matches sector-size BED 2019 birth rates

Notes: "matched to FB" means that establishment ℓ has a unique Facebook address that averages at least one post per week during the weeks when they are active in HB.

scribed in Table D1. The left panel of Table D4 presents the adjustment factors $\theta^{\mathcal{D}}_{\ell,t} = \theta^{\mathcal{D}}_i$ that multiply this probability to align the unadjusted HB death rates to BED/BDS death rates in 2019. These factors show a very consistent pattern: they are larger for small than for larger establishments, meaning that the unadjusted data would underestimate permanent closings for small and overestimate them for larger establishments. We note, however, that across size classes the unadjusted data comes close to the BED/BDS quarterly death rates.

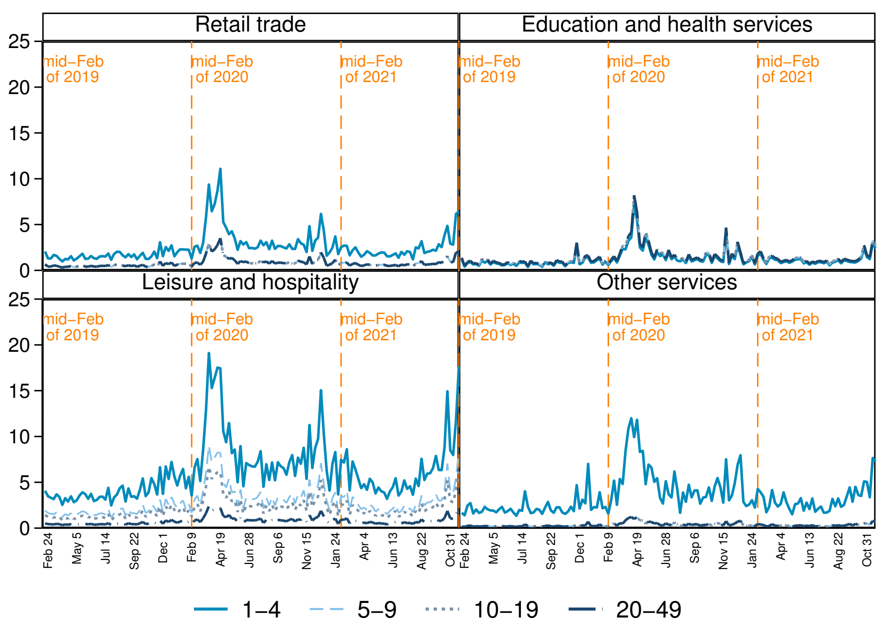
For **new openings**, the probability $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t})$ is described in Table D2 and the adjustment factors $\theta^{\mathcal{B}}_{\ell,t}$, averaged over the weeks of 2019, is described in Table D4. Similarly as for closings, the adjustment factors tend to be larger for small than for larger establishments, although not Education & Health Services, where in addition they are lower than one for all size classes, i.e. the unadjusted birth rates would always overestimate the BED/BDS birth rates for this sector. Figure D1 complements Table D4 by showing the $\theta^{\mathcal{B}}_{\ell,t}$'s for each sector and size class over time. Recall that the time variation is coming from $(\frac{\hat{n}^{\text{entry}}_{i,t}/\hat{n}^{\mathcal{A}}_{i,t}}{\hat{n}^{\text{entry}}_{i,0}/\hat{n}^{\mathcal{A}}_{i,0}})^{-1}$. During the pandemic, we observe a significant drop in the counts of new entrants in HB across all four sectors, which leads to an increase in the adjustment factors $\theta^{\mathcal{B}}_{\ell,t}$. Although not reported here, this effect is partly offset by a reduction in the Google / Facebook probability, $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t})$, during this period. For smaller establishments in Leisure & Hospitality and Other Services, and to a lower extent for Retail Trade, the adjustment factors $\theta^{\mathcal{B}}_{\ell,t}$ are noticeably higher during 2020 compared to the other time periods, meaning that birth rates would be substantially underestimated without the adjustment factors.

Table D4: Adjustment factors $\theta^{\mathcal{D}}_{\ell,t}$ and $\theta^{\mathcal{B}}_{\ell,t}$

$\theta^{\mathcal{D}}_{\ell,t}$ $(= \theta^{\mathcal{D}}_{i})$ **for permanent closings**					$\theta^{\mathcal{B}}_{\ell,t}$ **(average for 2019) for new openings**				
Class size	Retail Trade	Education & Health	Leisure & Hospitality	Other Services	Class size	Retail Trade	Education & Health	Leisure & Hospitality	Other Services
1–4	1.52	2.57	1.51	1.26	1–4	1.70	0.79	4.16	2.23
5–9	0.40	0.79	0.78	0.33	5–9	0.54	0.92	1.96	0.23
10–19	0.40	0.79	0.58	0.33	10–19	0.54	0.92	1.46	0.23
20–49	0.40	0.79	0.30	0.33	20–49	0.54	0.92	0.52	0.23

Notes: The table reports the adjustment factors for permanent closings, $\theta^{\mathcal{D}}_{\ell,t}$, and new openings, $\theta^{\mathcal{B}}_{\ell,t}$, calibrated to make the quarterly HB death and birth rates match the quarterly BED/BDS rates on average for 2019. For permanent closings, the adjustment factor are constant over time. For new openings since the adjustment factors vary over time, the table report the average value for 2019. Adjustment factors are identical for establishments with 5 to 49 employees within Retail Trade, Education & Health Services, and Other Services, since we pool class sizes together to avoid small cell issues.

Figure D1: Adjustment factors for new openings $\theta^{\mathcal{B}}_{\ell,t}$



Notes: Adjustment factor for $\theta^{\mathcal{B}}_{\ell,t}$, calibrated to make the quarterly HB birth rates match the quarterly BED/BDS rates on average for 2019. Adjustment factors are identical for establishments with 5 to 49 employees within Retail Trade, Education & Health Services, and Other Services, since we pool class sizes together to avoid small cell issues.

D.4 Benchmarking to BED establishment births and deaths

The adjustments presented in the previous section and benchmarking of our HB data rely on establishment births and death rates from the Business Employment Dynamics (BED). The BLS generates these rates by longitudinally linking establishment records of the QCEW. The BED reports quarterly rates of business closings and openings and business births and deaths by industry as well as employment gains and losses associated with these events. These rates are computed using the following definitions:[17]

- BED openings in quarter q are establishments with positive employment in the third month of quarter q and no employment in the third month of the previous quarter $(q-1)$;[18]
- BED closings in quarter q are establishments with zero employment in the third month of quarter q and positive employment in the third month of the previous quarter $(q-1)$;
- BED births in quarter q are establishments with positive employment in the third month of quarter q and no employment in the third month of the preceding four quarters $(q-4, q-3, q-2, q-1)$;
- BED deaths in quarter q are establishments with positive employment in the third month of quarter $q-1$ and no employment in the third month of the subsequent four quarters $(q, q+1, q+2, q+3)$.

As of this writing, the last BED data release was in January 2022 and includes data up to the second quarter of 2021. This means that the last available death rates are for the third quarter of 2020.

Adjustment of BED birth and death rates by size class. One important issue with comparing our HB data with the BED is that, at the industry level, the BED does not report statistics by establishment size class. This issue matters because entry and exit rates of small establishments are substantially higher than for larger establishments. We resolve this problem by using data from the Business Dynamics Statistics (BDS) of the U.S. Census Bureau, which contains entry and exit rates by industry and establishment size class. This data is only available annually up to 2019, and exit and entry rates are computed somewhat differently than BED entry and exit rates.[19] As Figure D2 shows, BDS and BED annual entry

[17]See https://www.bls.gov/news.release/cewbd.tn.htm for details on the BED. Rates are computed by dividing flows in quarter q by the average count of establishments in quarters q and $q-1$.

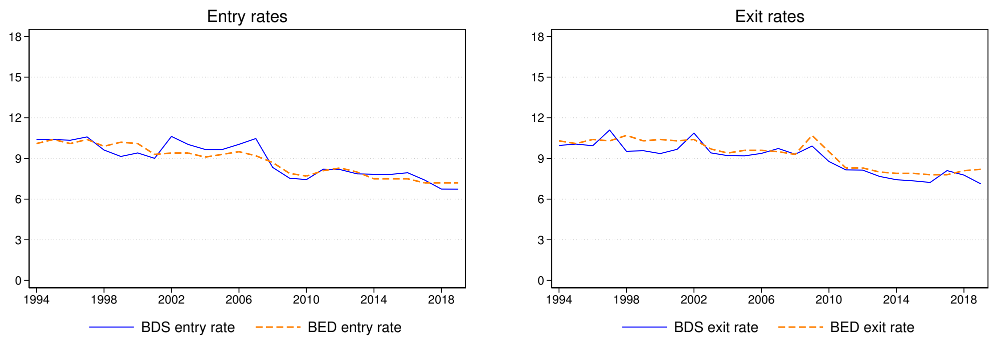
[18]No employment means either zero reported employment or no reported employment (e.g. because establishment appears for the first time in that quarter).

[19]BDS establishment entry and exits are defined as rates in a way similar to the BED's *annualized* opening and closing rates, defined as

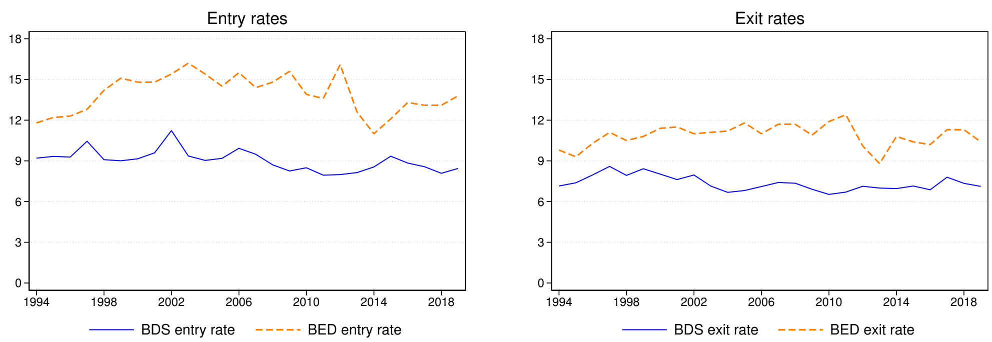
- annualized openings refer to establishments with positive employment in the third month of quarter q that were not present (or had zero employment) in the third month of quarter $q-4$,
- annualized closings refer to establishments with positive employment in the third month of quarter $q-4$ that are no

Figure D2: BDS and BED annual entry and exit rates

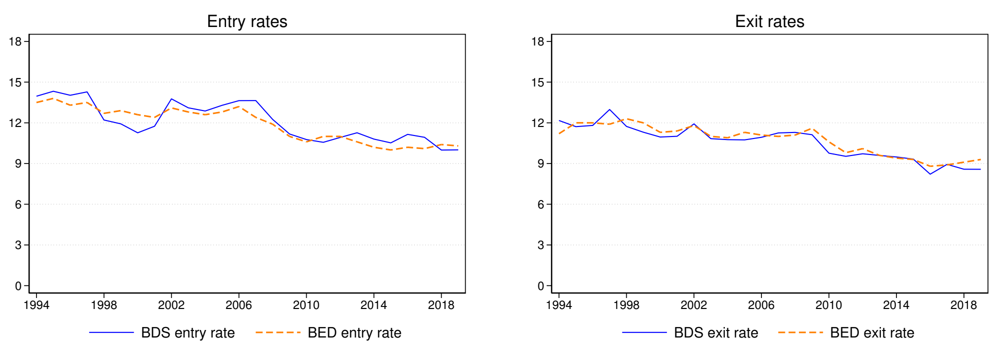
(a) Retail Trade



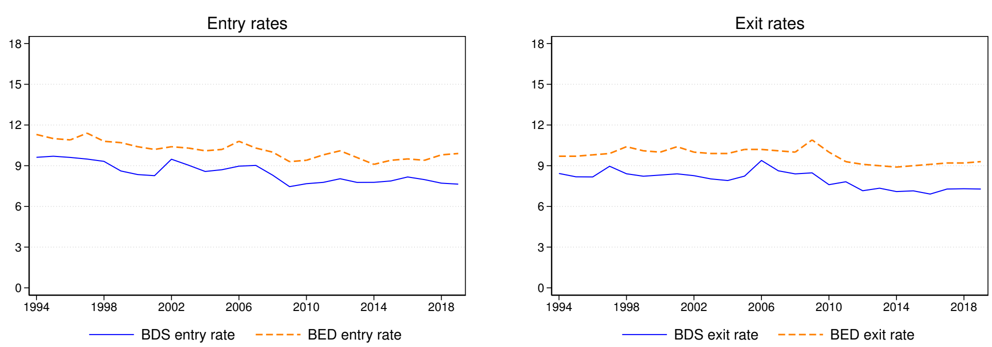
(b) Education & Health Services



(c) Leisure & Hospitality



(d) Other Services



Notes: BDS (solid lines) and BED (dashed lines) annual entry and exit rates, 1994-2019.

Table D5: BDS conversion factors

	Entry					Exit			
Class size	Retail Trade	Education & Health	Leisure & Hospitality	Other Services	Class size	Retail Trade	Education & Health	Leisure & Hospitality	Other Services
1–4	2.01	1.89	2.16	1.51	1–4	2.05	1.89	2.34	1.51
5–9	0.36	0.41	0.94	0.24	5–9	0.32	0.40	0.88	0.22
10–19	0.24	0.27	0.57	0.16	10–19	0.19	0.28	0.45	0.16
20–99	0.19	0.23	0.23	0.09	20–99	0.20	0.25	0.17	0.11

Notes: BDS data for 2015-2019, conversion factors by class size for entry and exit derived for Retail Trade (NAICS 44-45), Education and Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72) and Other Services (NAICS 81).

and exit rates line up closely for Retail Trade and Leisure & Hospitality, but less so for Education & Health Services (NAICS 61-62) and Other Services (NAICS 81) where BED rates are several percent above BDS rates. While some of these differences are due to data source (Business Register for the BDS and QCEW for the BED), the main reason for the larger BED rates in NAICS 61-62 and NAICS 81 are industry definitions / reclassifications.[20] Investigating the details behind these differences is beyond the scope of this paper. Besides, they should be innocuous for our estimates since we adjust both entries and exits to match the BED rates, and the discrepancy between the BED and the BDS in respectively NAICS 61-62 and NAICS 81 seems to affect both rates by the same order of magnitude.

We use the BDS rates to construct conversion factors by industry and size class that we then apply to BED industry rates to benchmark our HB data. Figure D3 shows that differences in entry and exit rates across class sizes within an industry are very stable over time. We take the average over the past 5 years of available data to construct the conversion factors reported in Table D5. As the table shows, entry and exit rates decrease with establishment size.

BED death rates for 2020. For Figure 5 of the paper comparing the (untargeted) BED birth and deaths rates for 2020 with their HB counterparts, we need BED death rates for the four quarters of 2020. However, as of this writing, the BED death rates are available only up to 2020Q3. We compute death rates for the fourth quarter of 2020 using an approximation proposed by Ryan Decker: $\text{deaths}_{i,q} = \text{closures}_{i,q} - \text{repoenings}_{i,q+1}$, where $\text{repoenings}_{i,q} = \text{openings}_{i,q} - \text{births}_{i,q}$. We can compute business reopenings for 2020Q4 because the available BED data covers openings and births for 2021Q1. We verify

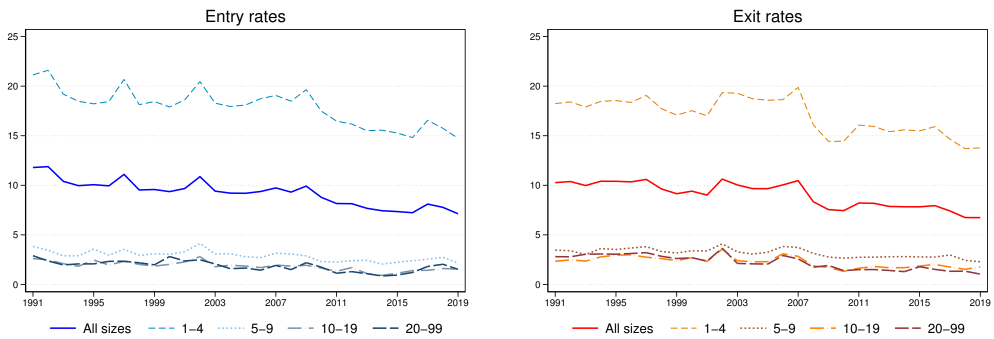
longer present (or have zero employment) in the third month of quarter q

We use these annual BED openings and closing rates to compare to the BDS annual entry and exit rates in Figure D2.

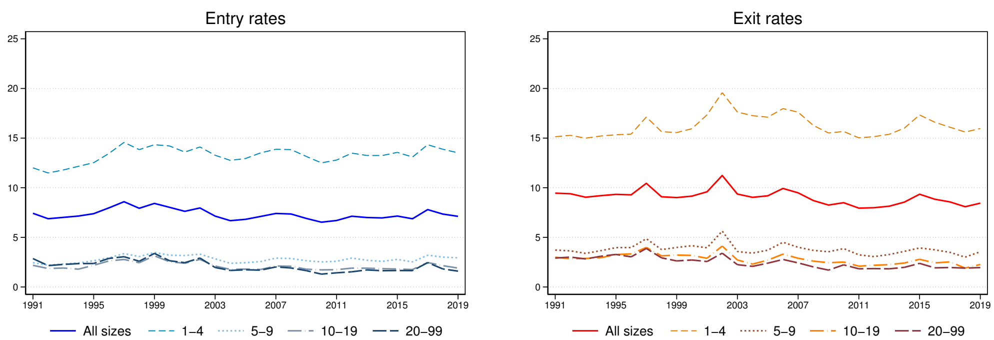
[20]For NAICS 61-62, the rates do not line up as closely before 2013, primarily because the definition of this sector in the BED was different from the definition of the (new) BDS. In particular, in 2013:Q1 the QCEW program reviewed establishments that provide non-medical, home-based services for the elderly and persons with disabilities and classified these establishments into services for the elderly and persons with disabilities (NAICS 624120). Many of these establishments were previously classified in the private households industry. (BDS industry rates are typically volatile around Economic Census years, which occur in 1997, 2002, 2007, 2012, when most of the reclassification of businesses occurs).

Figure D3: BDS annual entry and exit rates by size classes

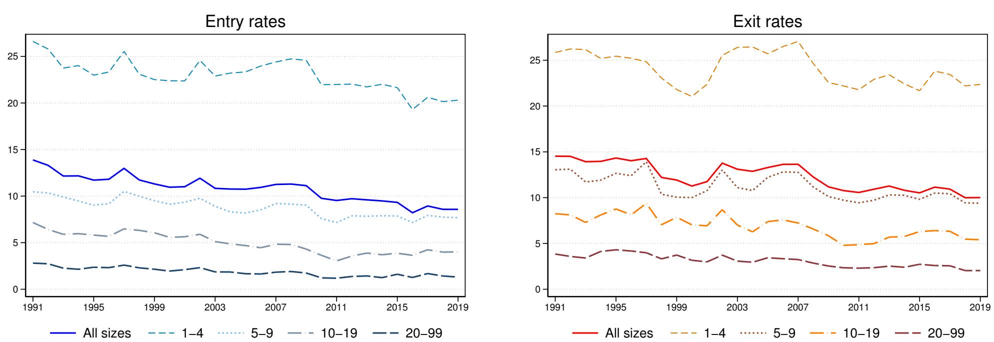
(a) Retail Trade



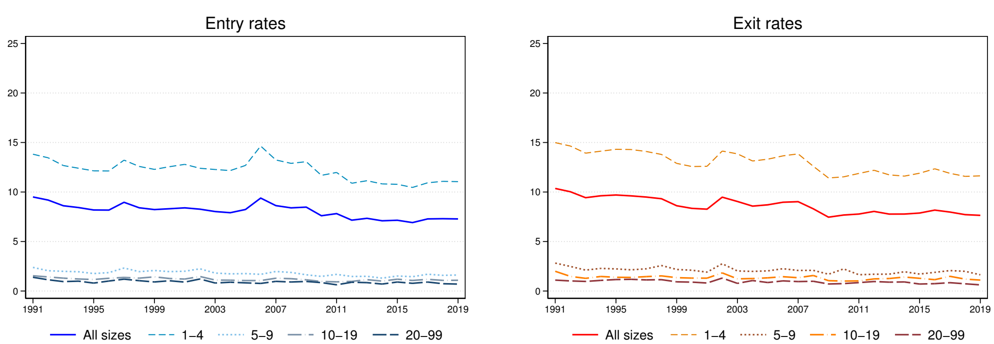
(b) Education & Health Services



(c) Leisure & Hospitality



(d) Other Services



Notes: BDS annual entry and exit rates, 1991-2019. The solid line in each plot reports the BDS rates for all class sizes, which include establishments with more than 100 employees. The other lines report BDS rates for establishment sizes that match our HB data.

that Decker's formula works well at the quarterly frequency for the four sectors of our analysis, even during the pandemic crisis where it predicts death rates for the second and third quarters of 2020 that come close to the actual rates.

Comparison between BED and HB rates. To compare BED rates to HB rates, we aggregate weekly tracked hours for each HB establishment ℓ in the base sample to the monthly level and then define establishment ℓ as having positive employment in quarter q if its tracked hours in the third month of that quarter are positive. We then define

- HB total entry for quarter q as HB establishments with positive employment in the third month of quarter q and no employment in the third month of the preceding four quarters $(q-4, q-3, q-2, q-1)$;[21]
- HB total exit for quarter q as HB establishments with positive employment in the third month of quarter $q-1$ and no employment in the third month of the subsequent four quarters $(q, q+1, q+2, q+3)$.

Because of sample churn, the HB total entry and exit rates are higher than the BED counterparts, as shown by Figure 1 in the paper. On the other hand, the adjusted HB birth and death rates (also shown in Figure 1) come very close to their BED counterparts. The HB birth rate in quarter q is the sum of $\hat{p}\left(\text{birth}_{i,t}\right)$ over the weeks of quarter q. The HB death rate in quarter q is the subset of HB establishments with positive employment in the third month of quarter $q-1$, no employment as of quarter q and that do not continue to operate according to our Google/Facebook approach. Since the Google/Facebook approach is intended to identify permanent closings, we effectively estimate that the establishment has no employment in the third month of the subsequent quarters.[22,23]

[21]As in the BED, no employment means either zero reported employment or no reported employment because the HB establishment appears for the first time in that quarter.

[22]If we strictly followed BED's definition of death, we would include HB locations that return to activity in $q+4$ or later. We exclude them here since these may simply be businesses that stopped using HB for a while and later on return as clients. In a previous version, we analyzed HB hazard rates defined as the ratio between the number of establishments that return to activity after t weeks of inactivity to the total number of establishments returning to activity. These hazard rates suggest returns to activity in $q+4$ is an extremely unlikely event.

[23]One caveat is that although Google Places distinguishes "temporary closed" and "permanently closed", we treat both indicators the same and tag the establishment as closed. Google's "permanently closed" accounts for more than 80% of all the exiting establishments from HB that are found to be Google-closed. Moreover, an establishment flagged as "temporary closed" at a given point in time might later on be flagged as "permanently closed" in Google Places.

E Representativeness and additional checks on closings/openings

E.1 Using visits data to assess representativeness of the Homebase sample

We use Safegraph's Points of Interest (POI) visits data to check for issues of representativeness and potential selection of small businesses into usage of the Homebase software. The basic idea is to take advantage of the much larger sample size of the Safegraph data to compare it to the Homebase data. The result of this exercise, which we explain in detail in this section, is readily summarized. Along the dimensions that can be compared, we find that Homebase establishments are not different in any significant manner from establishments from the larger Safegraph sample that covers over 20 percent of the universe of small businesses in the four sectors of our analysis. Thus, there is little evidence of sample selection into Homebase, at least among small businesses in service sectors that require in-person interaction.

Data preliminaries. We extract a sample of Safegraph POIs that meet two requirements: (1) they can be characterized as small establishments, and (2) have weekly foot traffic data available. (1) is challenging because the Safegraph data does not directly include a measure of establishment size. To address this issue, we take advantage of an extra data product called the NetWise dataset. NetWise is a data company that specializes in identifying business persons associated to company datapoints for sales, advertising and marketing purposes, using a variety of data aggregation techniques of online information; see `https://www.netwisedata.com/our-data/` for details. In September 2021, Safegraph released a cross-sectional NetWise dataset of counts of business persons that can be linked to the universe of POIs that were tracked by Safegraph at this point. POIs in this dataset are identified by a `placekey` identifier, which in turn allows us to match them to (a subset of) the Safegraph Core Places data (where NAICS codes can be obtained) and Safegraph Weekly Pattern (containing visits data that are relevant to characterize local economic activity).[24] The second data requirement slightly reduces the sample size because not all Safegraph POIs have visits (see Section B.1 for details on Safegraph)

Table E1 describes the samples of the analysis presented in the next paragraphs. After using the NetWise employment information to extract the set of POIs with fewer than 50 workers, we restrict the

[24] Between March 2020 and September 2020, POIs in the Safegraph Core Places data were identified using a `safegraph_place_id` (see Section B.1 for details). From the November 2020 release of the Core Places data to June 2021, they are identified using both a `safegraph_place_id` and `placekey` identifier. As of July 2021, the Core Places data only use the `placekey` identifier. We use the November 2020 through June 2021 release to create a crosswalk between `safegraph_place_id` and `placekey` identifiers. For the Safegraph Weekly Pattern, we rely on the latest release of these data, which, by construction of generating visits data based on the recent Core Places data, use `placekey` to identify POIs.

Table E1: Sample sizes of Safegraph and Homebase small establishments with available visits data

	Retail Trade	Education & Health	Leisure & Hospitality	Other Services	Total
Safegraph sample of small estabs:					
All	324,561	200,910	291,088	151,830	968,389
Without brand	188,622	187,023	167,980	146,139	689,764
Homebase sample	14,092	5,448	35,683	2,285	57,508

Notes: The table reports sample sizes for Safegraph small POIs in Retail Trade (NAICS 44-45), Education & Health (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81) with available visits data from Safegraph weekly patterns. The last row of the table reports sample size for the Homebase establishments from either the mid-February 2020 base sample or 2020-2021 new entrants sample with available Safegraph visits data.

sample to those in the four sectors of interest and with available visits data. The overall sample contains almost 1 million establishments (upper panel). Compared to the 4.4 million of small businesses for the four sectors according to QCEW establishment counts, this means the Safegraph-NetWise extract covers 22 percent of its universe. In Table E1, we further distinguish between POIs that are associated with a brand and those that are not. A Safegraph brand corresponds to a chains of commercial POIs (McDonald's, Starbucks, etc.); see `https://docs.safegraph.com/docs/core-places#section-brands`. The table shows that brands are pervasive in Retail Trade and Leisure & Hospitality, and much less so in Education and Health and in Other Services. As will be shown below, there are differences in visits data between POIs that have no brand associated (which typically are single commercial locations) and branded POIs. However, the patterns of changes in visits over time are similar across the two sets of POIs.

The last row of Table E1 describes the sample of Homebase establishments that we use to run the comparative analysis of visits data. The overall sample size is about 58,000 establishments. This corresponds to the mid-February 2020 base sample and the 2020-2021 new entrants sample put together, and for which we have available visits data from Safegraph weekly patterns. Note that we also use our 2020 samples to check the accuracy of the NetWise employment information. Not all the establishments from the Homebase samples can be linked to NetWise, due to differences in identifiers across databases,[25] in addition to issues such as business closing between February 2020 and September 2021. Meanwhile, among the roughly one third of establishments from the 2020 base and new entrants samples that can be linked across datasets, we find almost 90 percent of them have fewer than 50 workers according to the NetWise employment dataset.[26]

[25] In the NetWise dataset (which covers data for September 2021), establishments are identified using the `placekey` identifier. In our Homebase data, we link establishments to a `safegraph_place_id`, and not all `safegraph_place_id`'s that have existed since the March 2020 Core Places of Safegraph can be linked to a `placekey` identifier of September 2021.

[26] Across all four sectors, NetWise attributes fewer than 50 workers to 88 percent of the Homebase establishments from

Comparisons of visits data. Having described the two data samples, we now compare them with respect to the various dimensions of foot traffic data measured by Safegraph. We analyze four such dimensions: weekly counts of visits, median dwell time and share of long visits (defined as daily visits longer than 240 minutes) among all visits, and weekly visits per unique visitor. For each samples, we construct time series covering the period from mid-February 2020 through the end of the sample period, and compare the Homebase sample denoted by green crosses with the Safegraph ones denoted in orange and red marks. Figures E1–E4 present the results.

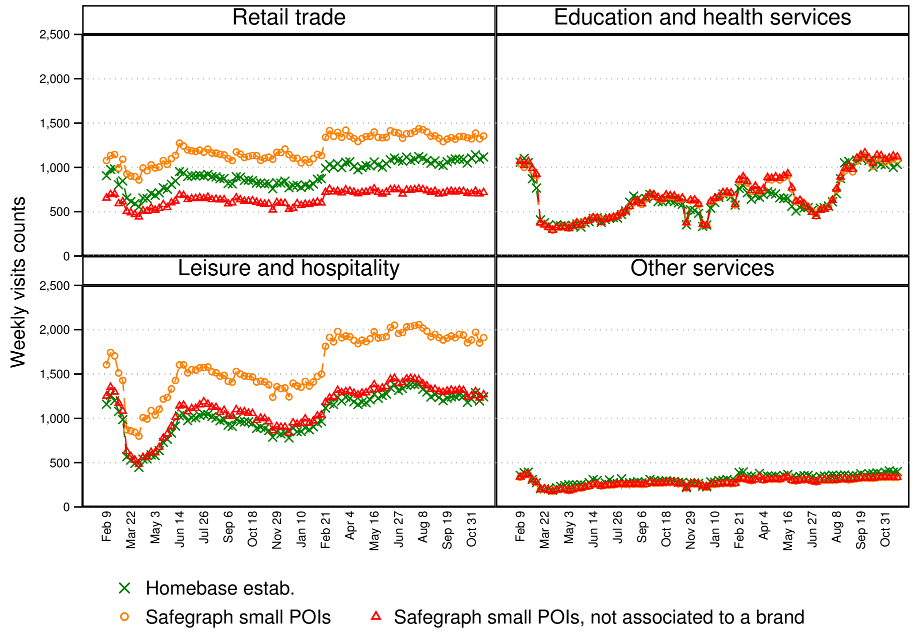
First, Figure E1 reports average weekly visits counts. Typically, establishments in Retail Trade and Leisure & Hospitality receive more visits than those in Education & Health Services and Other Services. Across all four sectors, we observe a large drop in visits at the beginning of the pandemic. In relative terms, the drop is larger in Education and health services (visits fall by almost 80 percent) and in Leisure & Hospitality (the decrease is by 60 percent in mid-April 2020). More importantly for our purposes, we observe a very similar behavior over time of the time series that correspond to the different samples. In Retail Trade, Homebase establishments are more similar (based on average weekly visits) to the sample of all Safegraph small POIs, while in Leisure & Hospitality it resembles that of Safegraph small POIs that have no brand associated. No matter these differences, in relative terms compared to mid-February 2020, the Homebase establishments do not behave differently from those of the larger Safegraph sample.

In Figure E2 and E3, we turn our attention to the duration of visits, by looking at median dwell times and the share of visits lasting longer than 4 hours among all visits. In Retail Trade and Leisure & Hospitality, visits are significantly shorter than in the other two sectors, with visits duration for the Homebase establishments between those of Safegraph establishments hat have no brand associated and those branded. In Other Services, we see no difference between the different samples for median dwell times and the share of long visits. In Education & Health Services, we observe an unclear pattern in what concerns long visits, which become relatively more important during the pandemic in the Homebase sample compared to the Safegraph samples. This discrepancy might be driven by outliers and leaves almost no discernible difference in median dwell times between the different samples in Education & Health Services.

Last, in Figure E4 we consider the number of visits per unique weekly visitor. Again, Retail Trade and Leisure & Hospitality differ from the other two sectors in that they have a lower share of returning

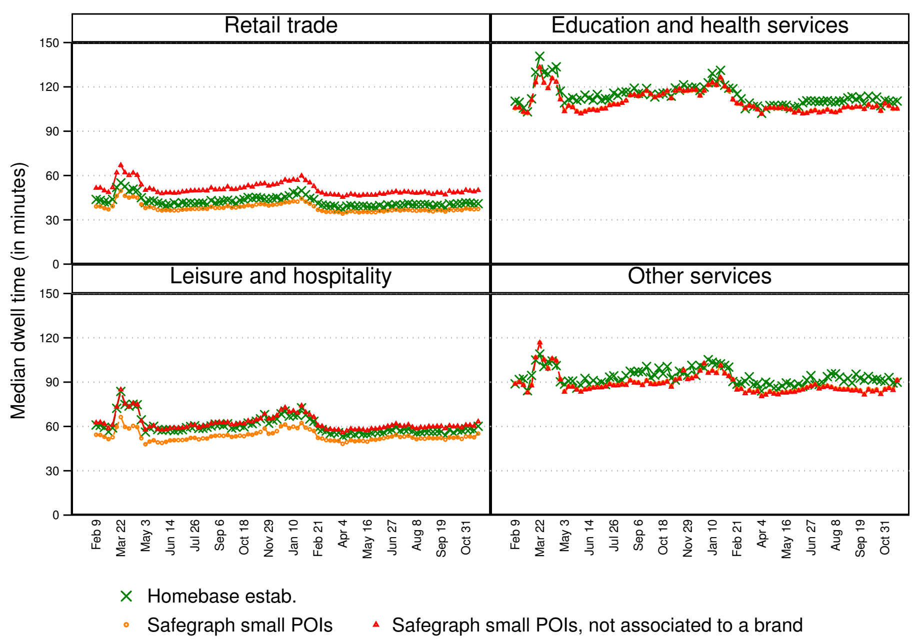
the 2020 base sample, and fewer than 100 workers to 95 percent of them. At the same time, there are differences across sectors. 92 percent of the Homebase establishments in Leisure & Hospitality have fewer than 50 workers according to Netwise employment, while the corresponding number is only 74 percent for Education & Health Services. In the latter sector, the distribution shifts towards small businesses if we exclude establishments in NAICS 611 ("Educational services") and 622 ("Hospitals"): Netwise employment then classifies 82 percent of the Homebases businesses as having fewer than 50 workers.

Figure E1: Safegraph small vs. Homebase establishments: Weekly counts of visits



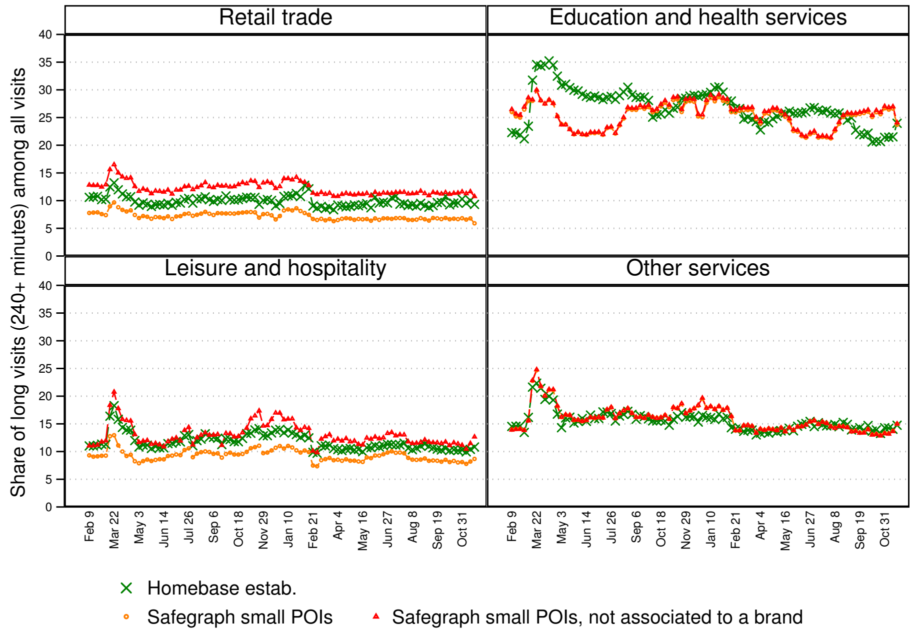
Notes: Safegraph visits data. The orange circles and red triangles denote Safegraph POIs with fewer than 50 workers according to NetWise employment data. The green crosses denote Homebase establishments matched to Safegraph visits data.

Figure E2: Safegraph small vs. Homebase establishments: Median dwell time



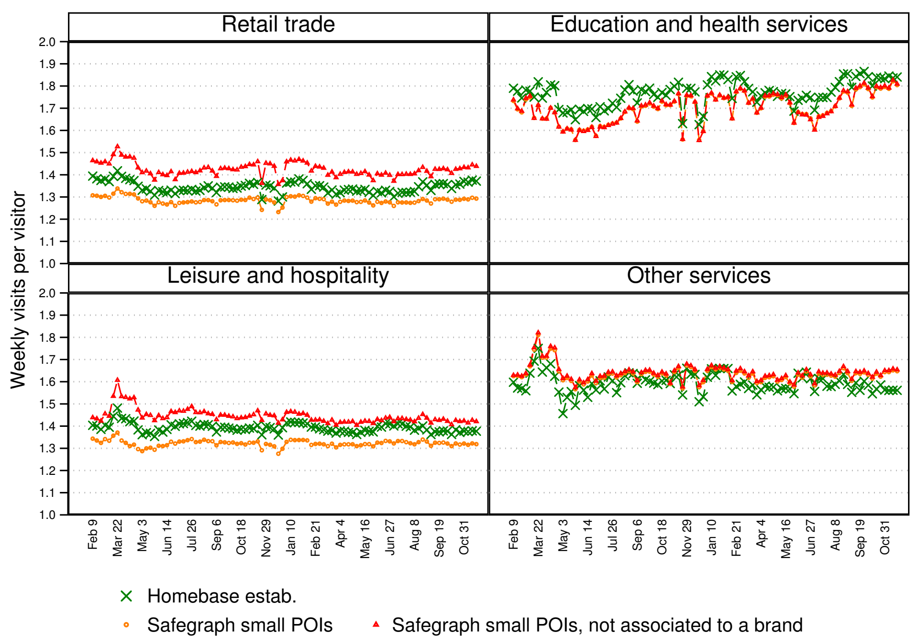
Notes: Safegraph visits data. The orange circles and red triangles denote Safegraph POIs with fewer than 50 workers according to NetWise employment data. The green crosses denote Homebase establishments matched to Safegraph visits data.

Figure E3: Safegraph small vs. Homebase establishments: Share of long visits



Notes: Safegraph visits data. The orange circles and red triangles denote Safegraph POIs with fewer than 50 workers according to NetWise employment data. The green crosses denote Homebase establishments matched to Safegraph visits data.

Figure E4: Safegraph small vs. Homebase establishments: Weekly visits per visitor



Notes: Safegraph visits data. The orange circles and red triangles denote Safegraph POIs with fewer than 50 workers according to NetWise employment data. The green crosses denote Homebase establishments matched to Safegraph visits data.

visitors within the week. What it is also remarkable in this figure is that we see little changes over time in the number of visits per visitor, including in Retail Trade and Leisure & Hospitality. The picture shown in Figure E4 is similar to the other figures, indicating that the Homebase establishments are not different in any significant manner from establishments from the larger Safegraph samples.

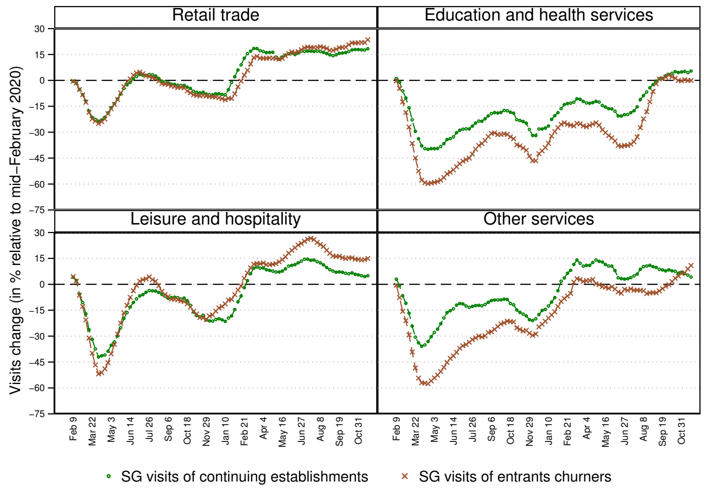
E.2 Using visits data to compare closings and openings against sample churn

We use the Safegraph visits to run additional checks on our procedure to distinguish new openings and closings from sample churn (see Section D.1). Recall that our Google/Facebook approach identifies establishments that were already operating before entry into HB and establishments that continue to operate outside of HB after disappearing from the data. In principle, these "churn" establishments should behave similarly to those that are included in the base sample of our analysis. This is the basic test that we perform in Figure E5.

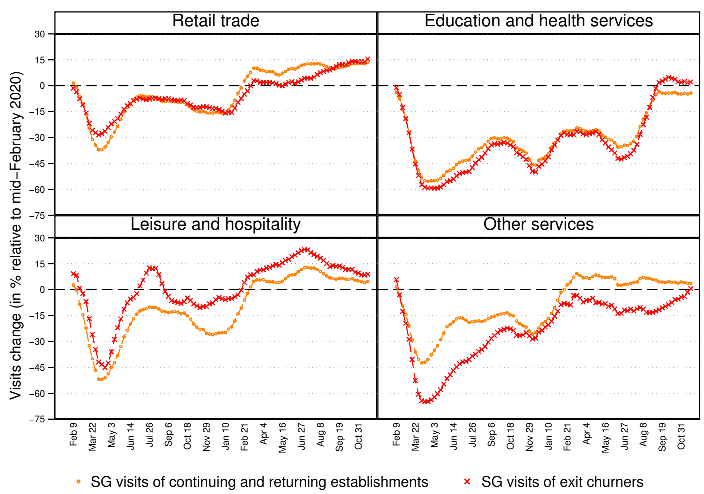
The upper panel of Figure E5 compares changes in visits among entrant churners (they enter HB after the mid-February base period and we identify that they were already operating before) with the continuously active establishments from the base sample. The latter is the relevant comparison group as it consists of the most stable establishments among those already active in the base period. Changes in visits among the two sets of establishments are extremely similar in Retail Trade and Leisure & Hospitality. In Education & Health Services and Other Services, there is a discrepancy coming from the larger decline in visits at the beginning of the pandemic among churners, but the subsequent dynamics is similar. In the lower panel of Figure E5, we compare exit churners to a larger set of establishments that includes returning establishments in addition to the continuing establishments from the base sample. The reason we include returning establishments is that they are the most relevant counterpart to establishments that exit but continue to operate outside of HB: if exit churners were to use HB software again, we would classify them as returning establishments. The plots show a great deal of overlap between the different series. In Retail Trade and Education & Health Services, they are virtually identical to each other. In Leisure & Hospitality, visits at exit churners seem to recover faster than among continuing and returning establishments, but by the end of the sample period the series are almost on top of each other. The opposite happens in Other Services. Overall, entry and exit churners, as identified by our Google / Facebook approach, seem to behave in a very similar manner to their counterparts from the base sample.

Figure E5: Safegraph visits among Homebase "churn" establishments

(a) Entrant churners



(b) Exit churners



Notes: Visit changes by continuing and returning businesses (dotted lines) vs businesses that we identify as churners among HB entries and exits (crossed lines) in percent of respective visits level during the week of Feb 9 - Feb 15, 2020 for Retail Trade (NAICS 44-45), Education and Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81).

F Employment decompositions

F.1 Decomposition by establishment status

As described in the main text, our employment estimator is

$$\widehat{E}_t = \widehat{E}_{t-1} \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}} \right)}{\sum_i \omega_i \left(\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{C}_{i,t}} \right)} \tag{F.1}$$

where ω_i denotes the sampling weight for industry-size-region cell i, constructed as the ratio of QCEW establishment counts in 2020:Q1 to HB establishment counts in that industry-size-region cell; $\hat{e}_{i,t}^{\mathcal{A}_{i,t}}$ denotes week t employment of the set of establishments $\mathcal{A}_{i,t}$ that are active in HB in both week t and $t-1$; $\hat{e}_{i,t}^{\mathcal{O}_{i,t}}$ denotes week t employment of the set of establishments $\mathcal{O}_{i,t}$ that are either newly opening or reopening in week t; and $\hat{e}_{i,t-1}^{\mathcal{C}_{i,t}}$ denotes week $t-1$ employment of the set of establishments $\mathcal{C}_{i,t}$ that are closing either temporarily or permanently in week t.

To motivate our decomposition, suppose that there is no sample churn; i.e. all exits from the HB sample are temporary or permanent closings and all entrants in the HB sample are new openings or reopenings. Under this scenario, the following equality holds

$$\hat{e}_{i,t-1} = \hat{e}_{i,t-1}^{\mathcal{A}_{i,t-1}} + \hat{e}_{i,t-1}^{\mathcal{O}_{i,t-1}} = \hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t-1}^{\mathcal{C}_{i,t}} \tag{F.2}$$

for every industry-size-region cell i. Intuitively, $\mathcal{A}_{i,t-1}$ is the set of establishments active in week $t-1$ and $t-2$ and $\mathcal{O}_{i,t-1}$ the set of establishments active in week $t-1$ but not $t-2$. Together they account for the set of all establishments active in week $t-1$. Without sample churn, this set is the same as $\mathcal{A}_{i,t}$, the set of establishments active in week $t-1$ that continue to be active in t plus $\mathcal{C}_{i,t}$, the set of establishments active in week $t-1$ but not in t. With sample churn, this equality would not hold since $\mathcal{A}_{i,t-1}$ would also contain establishments that exit HB in t but continue to operate (i.e. not closings) and $\mathcal{A}_{i,t}$ would also contain establishments that enter HB in $t-1$ but operate already beforehand (i.e. not openings).

Given (F.2), we can iterate Equation (F.1) backward to week 0 and obtain

$$\widehat{E}_t = E_0 \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}} \right)}{\sum_i \omega_i \left(\hat{e}_{i,0}^{\mathcal{A}_{i,1}} + \hat{e}_{i,0}^{\mathcal{C}_{i,1}} \right)} = E_0 \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}} \right)}{\sum_i \omega_i \hat{e}_{i,0}}. \tag{F.3}$$

where E_0 is CES employment in reference week 0, and $\hat{e}_{i,0}$ is HB employment of all establishments belonging to cell i in reference week 0. Subtracting E_0 from both sides, we can therefore express the change in employment relative to the reference week 0 as

$$\widehat{E}_t - E_0 = E_0 \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}} - \hat{e}_{i,0} \right)}{\sum_i \omega_i \hat{e}_{i,0}}. \tag{F.4}$$

Now, we split $\hat{e}_{i,t}^{\mathcal{O}_{i,t}}$ into employment $\hat{e}_{i,t}^{\mathcal{B}_{i,t}}$ from new openings (births) in week t and $\hat{e}_{i,t}^{\mathcal{R}_{i,t}}$ employment from establishments that were active in reference week 0, temporarily closed at some point, and reopen in week t. Hence,

$$\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}} = \hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{R}_{i,t}} + \hat{e}_{i,t}^{\mathcal{B}_{i,t}}. \tag{F.5}$$

Next and still supposing no sample churn, week t employment of establishments active in week $t-1$ and t, $\hat{e}_{i,t}^{\mathcal{A}_{i,t}}$, can be decomposed as

$$\hat{e}_{i,t}^{\mathcal{A}_{i,t}} = \hat{e}_{i,t}^{\mathcal{A}_{i,t}^{(2)}} + \hat{e}_{i,t}^{\mathcal{R}_{i,t-1}} + \hat{e}_{i,t}^{\mathcal{B}_{i,t-1}}, \tag{F.6}$$

where, with some abuse of notation, $\hat{e}_{i,t}^{\mathcal{A}_{i,t}^{(2)}}$ denotes week-t employment of establishments continuously open from week $t-2$ to t, $\hat{e}_{i,t}^{\mathcal{R}_{i,t-1}}$ denotes week-t employment of establishments reopening in $t-1$, and $\hat{e}_{i,t}^{\mathcal{B}_{i,t-1}}$ denotes week-t employment of new establishments opening in $t-1$. Combining Equations (F.5) and (F.6) and iterating back to reference week 0, we obtain

$$\hat{e}_{i,t}^{\mathcal{A}_{i,t}} + \hat{e}_{i,t}^{\mathcal{O}_{i,t}} = \hat{e}_{i,t}^{\mathcal{A}_{i,t}^{(t)}} + \sum_{s=1}^{t} \hat{e}_{i,t}^{\mathcal{R}_{i,s}} + \sum_{s=1}^{t} \hat{e}_{i,t}^{\mathcal{B}_{i,s}}, \tag{F.7}$$

where $\hat{e}_{i,t}^{\mathcal{A}_{i,t}^{(t)}}$ denotes week-t employment of establishments that stayed continuously open from reference week 0 to t, $\sum_{s=1}^{t} \hat{e}_{i,t}^{\mathcal{R}_{i,s}}$ denotes the sum of week t employment of all establishments that temporarily closed at some point after reference week 0 and reopened *by* week t, and $\sum_{s=1}^{t} \hat{e}_{i,t}^{\mathcal{B}_{i,s}}$ denotes the sum of week t employment of all establishments that newly opened after reference week 0. Similarly, we can decompose week-0 employment of all establishments belonging to cell i in reference week 0 as

$$\hat{e}_{i,0} = \hat{e}_{i,0}^{\mathcal{A}_{i,t}^{(t)}} + \sum_{s=1}^{t} \hat{e}_{i,0}^{\mathcal{R}_{i,s}} + \hat{e}_{i,0}^{\mathcal{C}_{i,t}} \tag{F.8}$$

where $\hat{e}_{i,0}^{\mathcal{C}_{i,t}}$ is added by definition of $\sum_{s=1}^{t} \hat{e}_{i,0}^{\mathcal{R}_{i,s}}$ not including establishment that are still closed in week t, and $\sum_{s=1}^{t} \hat{e}_{i,0}^{\mathcal{B}_{i,s}}$ is missing by definition of week-0 employment of establishments newly opening in week $s = 1, ..., t$ being zero.

Plugging (F.7) and (F.8) into Equation (F.4), we finally obtain

$$\widehat{E}_t - E_0 = \underbrace{E_0 \times \frac{\sum_i \omega_i \left(\hat{e}_{i,t}^{\mathcal{A}_{i,t}^{(t)}} - \hat{e}_{i,0}^{\mathcal{A}_{i,t}^{(t)}} \right)}{\sum_i \omega_i \hat{e}_{i,0}}}_{\text{Change from continously active estabs}} + \underbrace{E_0 \times \frac{\sum_i \omega_i \sum_{s=1}^{t} \left(\hat{e}_{i,t}^{\mathcal{R}_{i,s}} - \hat{e}_{i,0}^{\mathcal{R}_{i,s}} \right)}{\sum_i \omega_i \hat{e}_{i,0}}}_{\text{Change from reopenings}}$$
$$+ \underbrace{E_0 \times \frac{\sum_i \omega_i \sum_{s=1}^{t} \hat{e}_{i,t}^{\mathcal{B}_{i,s}}}{\sum_i \omega_i \hat{e}_{i,0}}}_{\text{Change from new openings}} - \underbrace{E_0 \times \frac{\sum_i \omega_i \hat{e}_{i,0}^{\mathcal{C}_{i,t}}}{\sum_i \omega_i \hat{e}_{i,0}}}_{\text{Change from closings}} \tag{F.9}$$

This decomposition holds exactly under no sample churn. With sample churn, the decomposition holds approximately under the assumption that while being active in HB, employment growth of entering establishments that operated prior to entry and exiting establishments that continue to operate after exit is about equal to employment growth of continuously active establishments. We verify that this indeed the case by comparing the change from continuously active establishments (the first term on the right-hand side above) with the residual obtained from subtracting the change from reopenings, the change from new openings and the change from closings (the last three terms on the right-hand side above) from $\widehat{E}_t - E_0$ (computed with Equation (F.1)).

F.2 Decomposition into hiring and separation flows

For a given establishment ℓ, we can decompose employment growth into hiring and separations

$$\hat{e}_{\ell,t} - \hat{e}_{\ell,t-1} = \hat{h}_{\ell,t} - \hat{s}_{\ell,t} \tag{F.10}$$

where $\hat{h}_{\ell,t}$ are all the employees in establishment ℓ who work in week t but not in $t-1$, and $\hat{s}_{\ell,t}$ are all the employees in establishment ℓ who work in week $t-1$ but not in t (for firms with several establishments, we define hiring and separations at the firm level; i.e. if an employee works at one establishment in one week but another establishment of the same firm in another week, we do not count it as a separation / hire). Hence,

$$\widehat{E}_t - \widehat{E}_{t-1} = \underbrace{\widehat{E}_{t-1} \times \frac{\sum_i \omega_i \sum_{\ell \in i} \hat{h}_{\ell,t}}{\sum_i \omega_i \hat{e}_{i,t-1}}}_{\text{Change from hiring}} - \underbrace{\widehat{E}_{t-1} \times \frac{\sum_i \omega_i \sum_{\ell \in i} \hat{s}_{\ell,t}}{\sum_i \omega_i \hat{e}_{i,t-1}}}_{\text{Change from job separation}}. \tag{F.11}$$

We let

$$\text{hiring rate}_t = \frac{\sum_i \omega_i \sum_{\ell \in i} \hat{h}_{\ell,t}}{\sum_i \omega_i \hat{e}_{i,t-1}} \quad \text{and} \quad \text{separation rate}_t = \frac{\sum_i \omega_i \sum_{\ell \in i} \hat{s}_{\ell,t}}{\sum_i \omega_i \hat{e}_{i,t-1}}$$

denote, respectively, the hiring rate and separation rate in week t. hiring rate_t and separation rate_t are plotted in Figure 8 of the paper, where separation rate_t is split into the hiring rate of new workers and that of recalled employees. The turnover rate (also in Figure 8 of the paper) is defined as

$$\text{turnover rate}_t = (\text{hiring rate}_t + \text{separation rate}_t) - \frac{|\sum_i \omega_i \hat{e}_{i,t} - \hat{e}_{i,t-1}|}{\sum_i \omega_i \hat{e}_{i,t-1}}.$$

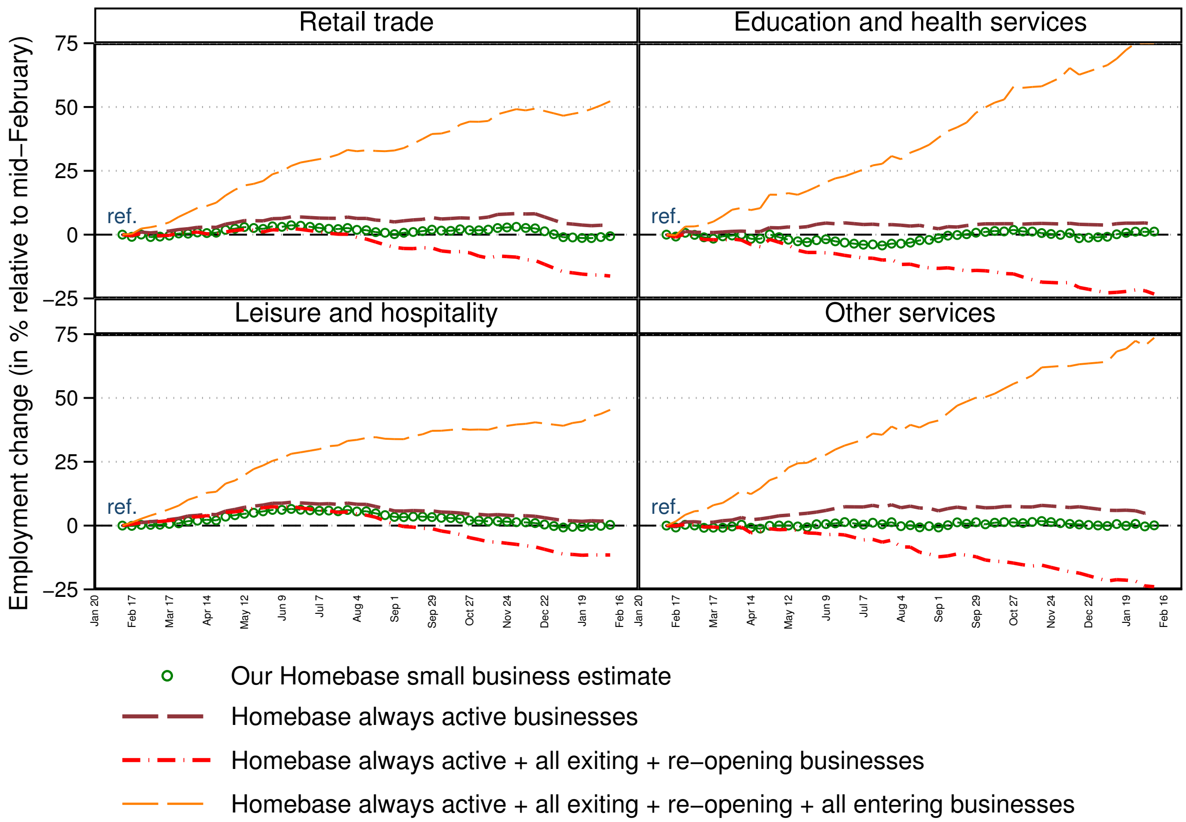
In Section 5 of the paper, we also discuss recall rates in our data. The recall rate in week t is defined as the ratio between recalled employees and the sum of recalled employees and new hires in week t.

G Additional figures and tables

G.1 Counterfactual employment estimates for the pre-pandemic period

Figure G1 is the counterpart of Figure 8 in the main text. The figure reports different counterfactual employment estimates: the brown short-dashed line uses only the set of establishments that are continuously active in HB; the red dashed-dotted line treats all exits as either temporary or permanent closings; the orange dashed line, finally, adds all entries and treats them as new openings. The green circled line corresponds to our baseline small business estimates. The latter shows that employment in mid-February 2020 was roughly similar to employment in mid-February 2019 in the four sectors considered. This is in line with the QCEW year-on-year employment growth rates reported in Figure 2 of the paper. The counterfactual employment estimates, on the other hand, would predict very large changes in employment between mid-February 2019 and 2020. For example, the orange dashed line yields year-on-year employment growth rates between 50 and 75 percent, depending on the sector considered. In sum, Figure G1 demonstrates that distinguishing closings and openings from sample churn is important even during the pre-pandemic period.

Figure G1: Comparison with counterfactual employment estimators



Notes: Estimated employment change in % relative to mid-February 2020 of small businesses with less than 50 employees in Retail Trade (NAICS 44-45), Education & Health Services (NAICS 61-62), Leisure & Hospitality (NAICS 71-72), and Other Services (NAICS 81) according to different estimation methods (see text). The estimates are constructed based on February 2020 CES employment estimates (week of Feb 9 – Feb 15) and QCEW shares of small business employment for the first quarter of 2020. The estimates for the weeks of Thanksgiving, Christmas, and New Year are smoothed by using the estimates of adjacent weeks.

G.2 Regression variables

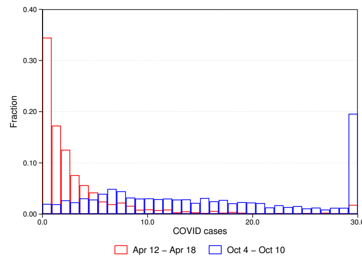
Table G1 describes the main variables used in the regressions reported in the paper. Table G2 provides descriptive statistics for the main variables. Figure G2 complements the description of time-varying control variables by showing distributions for the weeks of April 12 – April 18 (when small business employment is at its lowest) and October 4 – October 10 (when large parts of the economy had reopened). In Figure G2, the plots are based on data for the counties used in the regressions (1,957 counties), but the distributions are virtually identical if we use data for all counties. Figure G3 compares the distribution of delays in PPP loan among counties included in the regressions vs. all counties. Again, the distributions are very similar when comparing counties included in our regressions and all counties.

Table G1: Description of the control variables used in Section 6

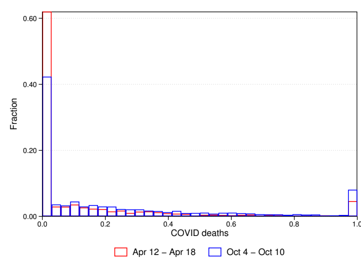
Variable	Description	Source
COVID_cases	Number of COVID cases per 1,000 county inhabitants	COVID Act Now
COVID_deaths	Number of COVID deaths per 1,000 county inhabitants	https://covidactnow.org/
containment_index	Weighted measure of restrictions for all containment measures (school closure, workplace closure, cancel public events, gathering restriction, public transportation restriction, stay-at-home order, internal movement control) at the state level	Oxford COVID-19 Government Response Tracker
NPI1	0: No restriction 1: 50% or more of all industries within a county having a closure restriction	Atalay et al. [2020] https://reopeningdata.github.io/
NPI1_sector	0: No restriction 1: businesses in a specific 2-digit NAICS within a county is required to be closed	
NPI2	0: No restriction/Advisory 1: Mandate stay-at-home policy for high-risk people 2: Mandate stay-at-home restriction for all people	Centers for Diseases Control and Prevention
NPI3	0: No restriction 1: Ban gathering above certain sizes 2: Ban gathering of all sizes	Centers for Diseases Control and Prevention
SG_school_visits	Safegraph school visits in log difference relative to same week of the previous year	Safegraph Weekly Patterns visits to places associated with NAICS code 611110 ("Elementary and Secondary Schools")
weather_max_temp	Weekly average of maximum daily temperature (in °F)	Climatology Lab GRIDMET climatologylab.org/gridmet.html
hh_income	County-level household income	2016-2019 American Community Survey 5-year Estimates (5-year ACS)
PPPdelay	PPP loans (at the county level) received during the week of Apr 26 - May 2 divided by the sum of PPP loans during theweeks of Apr 12 - Apr 18, Apr 19 - Apr 25, and Apr 26 - May 2	Small Business Administration and Doniger and Kay [2021]
PPPdelay_sector	PPP loans (county × 2-digit NAICS level) received during the week of Apr 26 - May 2 divided by the sum of PPP loans during theweeks of Apr 12 - Apr 18, Apr 19 - Apr 25, and Apr 26 - May 2	

Figure G2: Distribution of control variables in Apr 12 – Apr 18 vs. Oct 4 – Oct 10

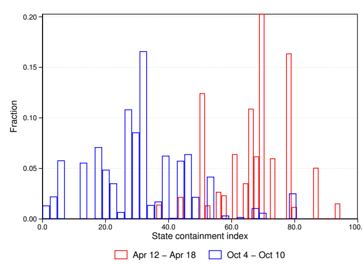
(a) COVID cases



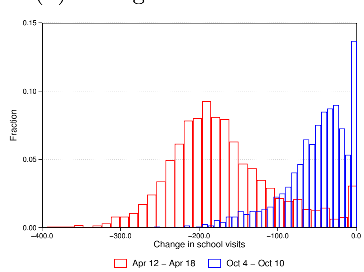
(b) COVID deaths



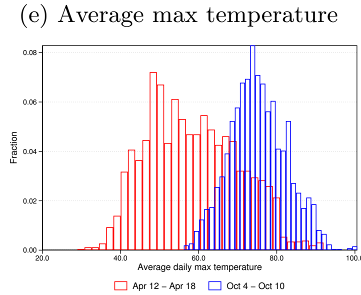
(c) Containment index



(d) Change in school visits



(e) Average max temperature



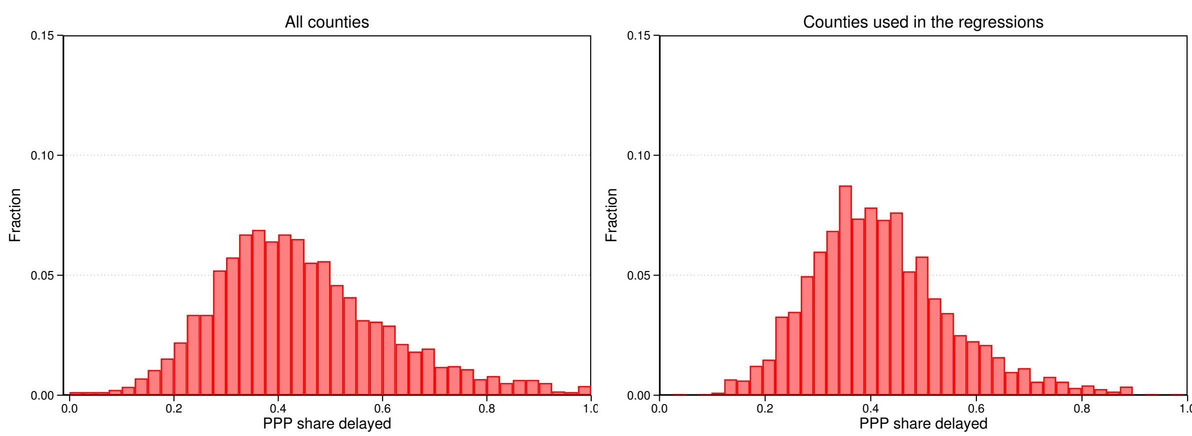
Notes: Distribution of control variables for the weeks of Apr 12 - Apr 18 and Oct 4 - Oct 10.

Table G2: Descriptive statistics of regression variables

	Mean	St. Dev.	p25	p50	p75	Min.	Max.
Controls: Week of Apr 12 – Apr 18							
COVID cases	4.04	8.44	0.52	1.66	3.98	0.00	132.90
COVID deaths	0.20	0.60	0.00	0.00	0.16	0.00	9.44
State containment index	66.82	11.22	61.11	69.44	72.22	36.11	94.44
Change in school visits	-172.25	68.03	-215.39	-181.03	-137.73	-393.81	144.71
Average daily max temperature	58.54	11.64	49.26	57.45	67.29	28.92	91.96
Controls: Week of Oct 4 – Oct 10							
COVID cases	19.71	19.68	7.23	14.50	24.88	0.00	249.25
COVID deaths	0.32	0.65	0.00	0.09	0.37	0.00	10.17
State containment index	31.15	16.04	20.83	30.56	40.28	0.00	80.56
Change in school visits	-46.72	51.98	-71.52	-42.45	-20.93	-257.42	279.70
Average daily max temperature	75.23	7.26	70.25	74.76	80.03	56.20	100.60

Notes: The table reports the mean, standard deviation ("St. Dev."), 25th, 50th, 75th percentiles (respectively "p25", "p50", "p75"), minimum and maximum values (respectively "Min.", "'Max.") of the time-varying control variables for the weeks of Apr 12 – Apr 18 (upper panel) and Oct 4 – Oct 10 (lower panel).

Figure G3: Distribution of delays in PPP, overall and for counties included in the regressions



Notes: Distribution of delays in PPP loans for all counties (left) and for counties included in the regressions (right).

G.3 Additional regression results

G.3.1 Coefficient estimates for different control variables. Table G3 reports regression estimates for the different time varying control variables. Generally, the estimates for the different variables have the correct sign and provide interesting additional information. Counties with higher rates of COVID new deaths are associated with lower small business employment and more business closings. Higher rates of COVID new cases, however, do not show up significantly. Turning to NPIs, the state containment index has a negative relation with county employment effect. County business restrictions (NPI1), stay-at-home orders (NPI2), and gathering bans (NPI3) also affect small business employment but only during the first three months of the pandemic. These variables do not, however, explain a lot of the variation in county employment, which echoes earlier findings by Bartik et al. [2020], Chetty et al. [2020] or Goolsbee and Syverson [2020] that NPIs were in and of themselves not a major factor for the decline in employment in the beginning of the pandemic.

Changes in school visits, which serve as a measure of school closings, also exert a negative effect on small business activity (this variable is scaled inversely; so the negative coefficient estimates implies that a larger decline in school visits is associated with less small business employment). This result is interesting and suggest that counties with more school closures experienced a more modest recovery in small business activity.

Finally, weather conditions as measured by average maximum daily temperature also exerts a significant negative effect on small business activity. This result is driven in large part by the Leisure & Hospitality sector.

G.3.2 Coefficient estimates for establishment-level regressions. Figure G4 plots the coefficient estimates for the establishment-level regressions

$$y_{i,t} = \sum_{t=0}^{57} \alpha_t \left(\mathbb{1}\left\{\text{week} = t\right\} \times \text{share PPP delayed}_{c(i)} \right) + \mathbf{X'_{c(i),t}}\boldsymbol{\gamma} + \phi_t + \mu_i + \varepsilon_{c,t} \tag{G.1}$$

where $\text{share PPP delayed}_{c(i)}$ is the share of delayed PPP loans in the county in which establishment i is located. All the controls are the same as in the county-level regressions in the main text, except that the fixed effect μ_i is at the establishment level instead of the county level. This fixed effect takes into account systematic differences in productivity and other unobservables across establishments.

Since this regression is at the establishment level, we do not have a county-level estimate of small business employment. But the three other regressions for employment of always active businesses, business

Table G3: Estimates for different control variables

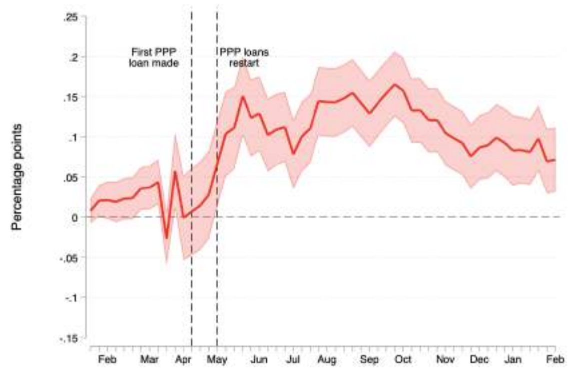
	A. County employment (percent change relative to mid-February)	B. Employment of always active businesses (percent change relative to mid-February)	C. Business closings (percent of active businesses in mid-February)	D. New business openings (percent of active businesses in mid-February)
Covid new cases per 100k	-0.01	-0.00	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Covid new deaths per 100k	-1.54***	-0.28	1.37***	0.03
	(0.26)	(0.21)	(0.19)	(0.11)
Containment index	-0.06***	-0.01	0.03***	0.00
	(0.02)	(0.01)	(0.01)	(0.00)
NPI1 × Feb-June 2020	-7.66***	-2.90***	6.01***	0.03
	(0.92)	(0.62)	(0.56)	(0.11)
NPI1 × July-Dec 2020	-3.61***	-2.41***	1.69**	-0.29*
	(0.89)	(0.74)	(0.54)	(0.15)
NPI1 × Jan-Feb 2021	-1.91	-1.53	1.29*	-0.37
	(1.23)	(1.06)	(0.68)	(0.27)
NPI2 × Feb-June 2020	-0.96***	-0.40*	1.26***	-0.03
	(0.38)	(0.25)	(0.24)	(0.05)
NPI2 × July-Dec 2020	-1.25	-1.59	0.74	-0.13
	(1.37)	(1.23)	(0.64)	(0.13)
NPI2 × Jan-Feb 2021	-1.50	-1.50	1.04*	0.26
	(1.01)	(0.97)	(0.53)	(0.22)
NPI3 × Feb-June 2020	-1.88***	-0.69**	1.45***	-0.16***
	(0.38)	(0.27)	(0.25)	(0.05)
NPI3 × July-Dec 2020	0.15	0.57*	-0.39**	-0.10
	(0.40)	(0.31)	(0.18)	(0.07)
NPI3 × Jan-Feb 2020	-2.18***	-0.71	0.39	-0.26**
	(0.52)	(0.45)	(0.30)	(0.12)
Log school visit change	0.00***	-0.00***	-0.00*	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Avg daily max temperature	0.15***	0.07***	-0.05***	-0.00
	(0.02)	(0.01)	(0.01)	(0.00)
R-squared	0.50	0.22	0.62	0.27
N	110,364	101,348	131,840	138,622
Controls:				
Relative county income × Week	✔	✔	✔	✔
County FE	✔	✔	✔	✔
Week FE	✔	✔	✔	✔

Notes: Standard errors are clustered at the county level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated over all weeks beween February 9-15, 2020 and January 31 - February 6, 2021. Table only shows coefficient estimates for COVID health NPI, school visit and maximum temperature regressors. Percent employment change relative to mid-February (Feb 9-15, 2020) in Column A is computed for all county-weeks for which HB sample contains positive employment observations. Percent employment change relative to mid-February (Feb 9-15, 2020) in Column B is computed for all county-weeks with continously active businesses. Percent of closed businesses in Column C is computed as the count of businesses closed (either temporarily or permanently) in week t relative to the count of businesses in the reference week. Percent of new business openings in Column D is computed as the cumulative count of new businesses as of week t relative to the count of businesses in the reference week and businesses that newly open after the reference week. NPI1 equals one if 50% or more of all industries within a county had a closing restriction in that week. NPI2 equals 1 (2) if the county imposed a stay-at-home restriction for high risk people (for all people) in that week. NPI3 equals 1 (2) if the county imposed a ban on gatherings of certain sizes (of all sizes) in that week.

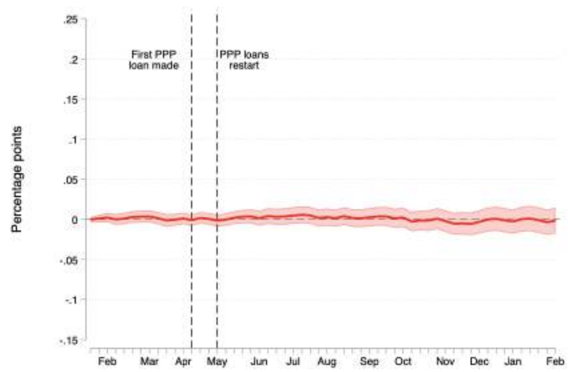
Figure G4: Effect of delayed PPP loans on small business activity

(a) Employment of always active businesses

(b) Business closings



(c) Newly opened businesses



Notes: Coefficient estimates of share PPP delayed$_c$ interacted with weekly fixed effects. Shaded areas show 95% confidence bands. All regressions are estimated over all weeks between January 5-11, 2020 and January 31 - February 6, 2021. share PPP delayed$_c$ is constructed as the amount of PPP loans issued in county c during the week of April 26 relative to the total amount of PPP loans issued per county during the weeks of April 12, April 19, and April 26. Employment of always active businesses in Panel (a) is the percent deviation relative to mid-February 2020 employment for all establishments that are continuously active throughout the entire sample. Business closings in Panel (b) is the probability that an establishment active in the reference period is closed in week t. Newly opened businesses in Panel (c) is the probability that an establishments not active in the reference period is a new opening as of week t. All regressions control for county-specific time-varying controls as described in the text as well as week- and establishment fixed effects. Standard errors are clustered at the establishment level.

closings, and new business openings are directly comparable to the county-level regressions. As can be seen, the estimates are very similar, confirming the robustness of the results reported in the main text.

References

Enghin Atalay, Shigeru Fujita, Sreyas Mahadevan, Ryan Michaels, and Tal Roded. Reopening the economy: What are the risks, and what have states done? *Research Brief*, 2020.

Keith Barnatchez, Leland Dod Crane, and Ryan Decker. An assessment of the national establishment time series (NETS) database. *FEDS Working Paper 2017-110*, 2017.

Alexander W. Bartik, Marianne Bertrand, Feng Lin, Jesse Rothstein, and Matthew Unrath. Measuring the labor market at the onset of the COVID-19 crisis. *Brookings Papers on Economic Activity*, forthcoming, 2020.

Raj Chetty, John N. Friedman, Nathaniel Hendren, Michael Stepner, and the Opportunity Insights Team. The Economic Impacts of COVID-19: Evidence from a New Public Database Built from Private Sector Data. Working paper 26463, National Bureau of Economic Research, 2020.

Cynthia Doniger and Benjamin Kay. Ten days late and billions of dollars short: The employment effects of delays in paycheck protection program financing. *FEDS Working Paper 2021-003*, 2021.

Austan Goolsbee and Chad Syverson. Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020. Working paper, Becker Friedman Institute, 2020.