Supplementary material for: Measuring Small Business Dynamics and Employment with Private-Sector Real-Time Data

André Kurmann — Drexel University

Etienne Lalé — York University

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.

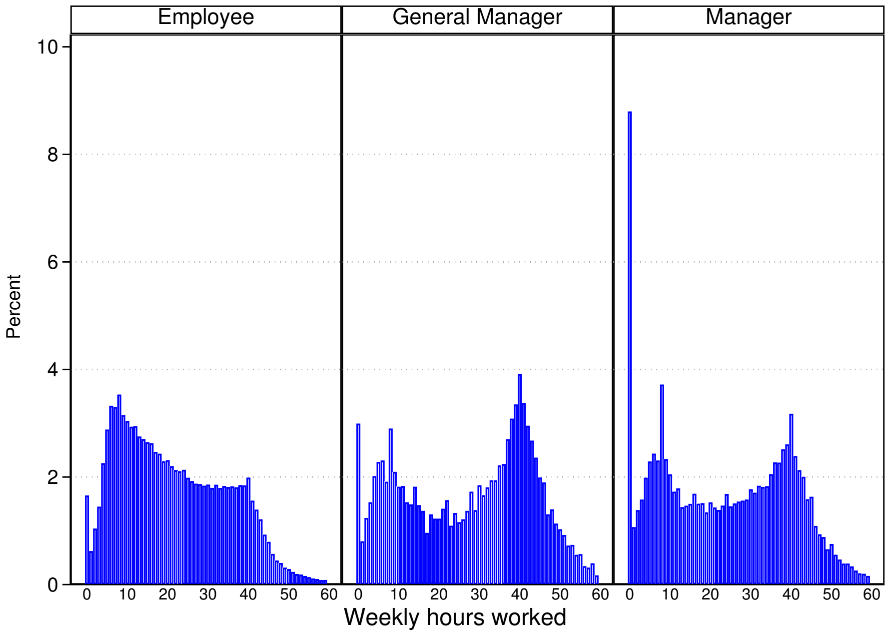
- Figure A2 shows the proportion of location_id's in 2020 with tracked hours only, untracked hours only, and both tracked and untracked hours. At the onset of the pandemic (mid-March 2020), location_id's with untracked hours make up 10% of all location_id's with some activity. The corresponding figure 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.

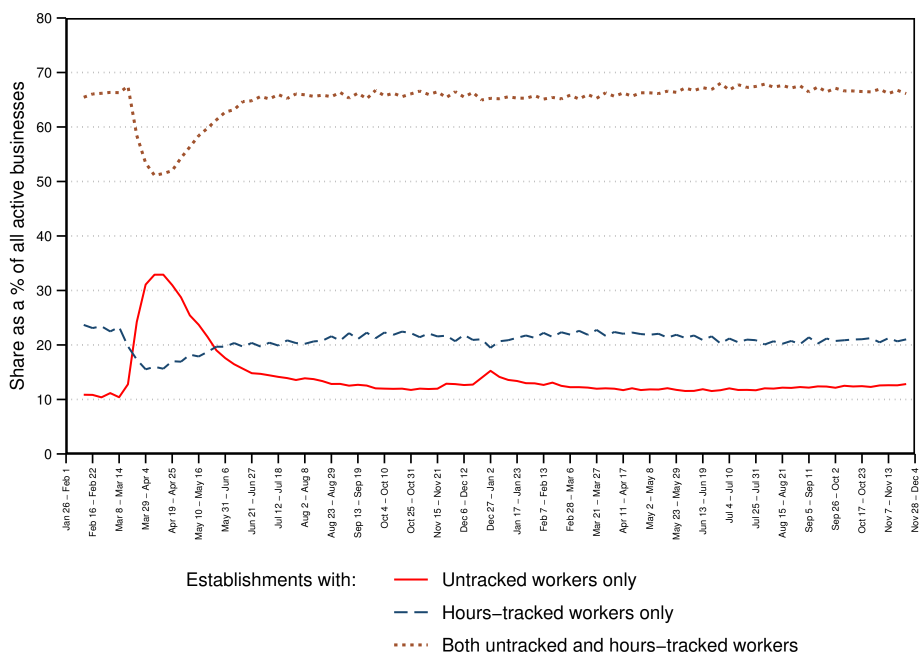
[4]In Figure A1, the sample is restricted to weekly hours worked (strictly) under 60 hours. Less than 1 percent of workers have weekly hours worked (averaged over this 10-week window of time) above the 60 hours cutoff.

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.

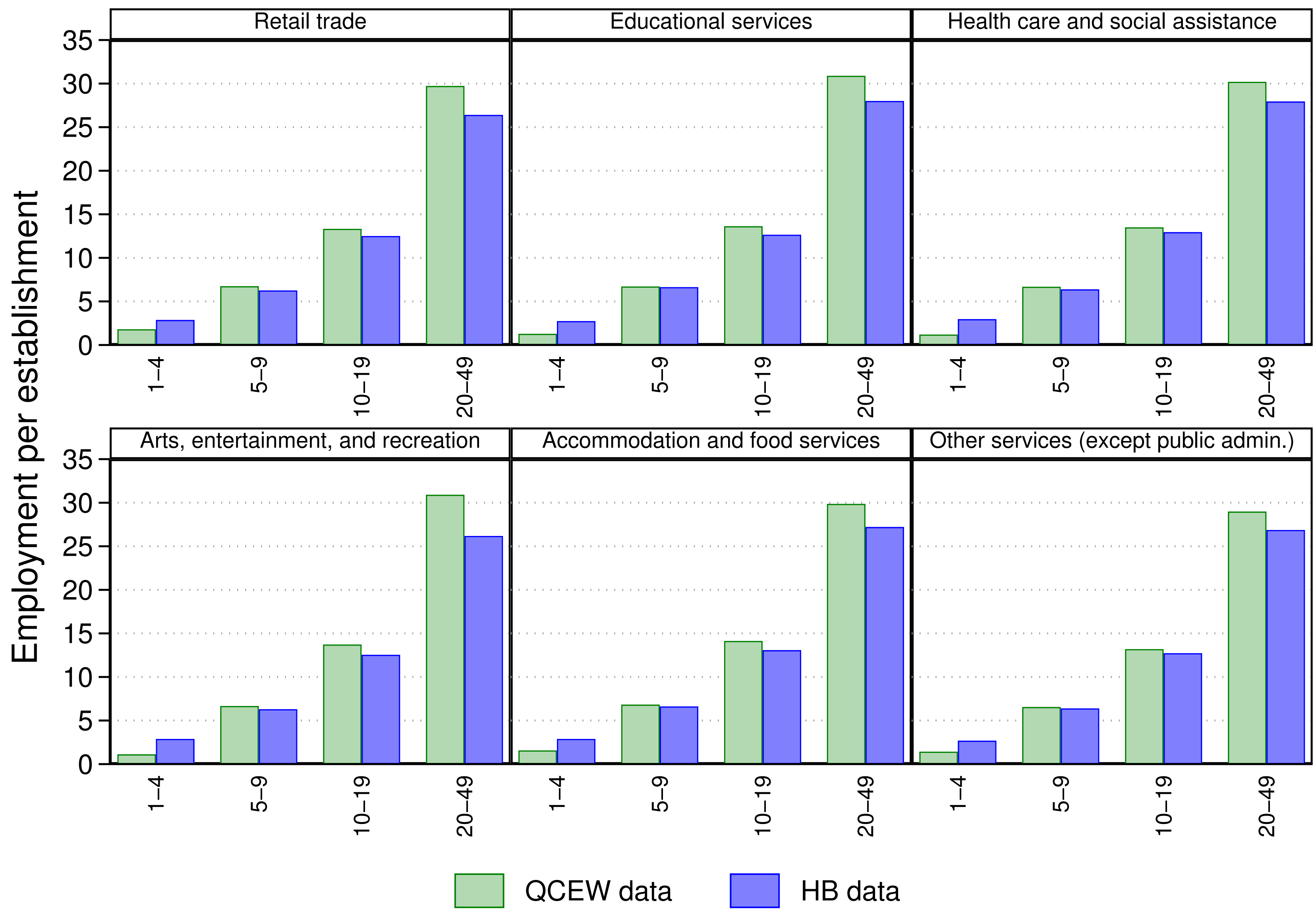
for 15% of establishments with fewer than 50 workers in the sectors of our analysis. Within NAICS 72, the distribution across size classes in HB is, again, less concentrated towards small businesses compared to QCEW data. In terms of employment, NAICS 72 accounts for an even larger share of the HB sample, 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 $\times$ size $\times$ 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.1.3 Sample selection issues. A potential concern with the Google/Facebook approach described in this section is that the matched establishments differ from the non-matched establishments in systematic ways. It is conceivable, for instance, that new entrants that can be matched to Facebook are more active in general (including on social media, leading to a successful match from HB to FB) than non-matched establishments. In turn, this may bias upward the probability of operating already prior to entering HB which is applied to new entrants not matched to Facebook in Step 2 of our Google/Facebook approach. As explained in the next section, this problem would jeopardize estimation only under specific conditions: it would need to be the case that the bias coming from the matched establishments changes over time. In Section E.3, we provide evidence that leads us to discard these concerns. We show that:

- For respectively permanent closings and new entrants, the sector and size characteristics of establishments that can be matched to Google or Facebook are very close to those of non-matched establishments;

- Throughout the sample period, the behavior of matched establishments in terms of their visits data is similar to that of non-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 $\theta_{\ell,t}^{\mathcal{D}}$ 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 $\theta_{\ell,t}^{\mathcal{D}}$ 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

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

where $\hat{N}_{i,q}^{\mathcal{D}}$ 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 $\hat{N}_{i,q}^{\mathcal{A}}$ 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, $N_{i,q}^{\mathcal{A}}$ is the count of all establishments with positive employment in the third month of quarter q; and $\hat{N}_{i,q}^{\mathcal{D}}$ 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

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

where $\hat{p}_{\mathrm{BED/BDS}}(\mathrm{death}_{i,q})$ denotes the BED/BDS quarterly death rate for industry-size cell i in quarter q.[15] Finally, we set $\theta_{\ell,t}^{\mathcal{D}} = \theta_i^{\mathcal{D}}$. Thus, by construction of the $\theta_{\ell,t}^{\mathcal{D}}$, the adjusted HB death rates (i.e., $\sum_{\ell\in i}\theta_{\ell,t}^{\mathcal{D}}\times\hat{p}(\mathcal{D}_{\ell,t}|\mathrm{exit}_{\ell,t})$) 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 $\theta_i^{\mathcal{D}}$ 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 $\theta_i^{\mathcal{D}}$ separately for all size classes, both because our sample is larger and because death rates vary more substantially across 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.

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

where $\hat{e}_{\ell,t}$ denotes employment of entering establishments during the week of entry into HB, $\hat{p}(\mathcal{B}_{\ell,t}|\mathrm{entry}_{\ell,t})$ 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, $\hat{p}(\mathcal{B}_{\ell,t}|\mathrm{entry}_{\ell,t})$.

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 $\hat{p}(\mathcal{B}_{\ell,t}|\mathrm{entry}_{\ell,t}) = p(\mathcal{B}_{\ell,t})$; 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

[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.

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 $t = 0$), $\frac{\hat{n}^{\text{entry}}_{i,t}/\hat{n}^{\mathcal{A}}_{i,t}}{\hat{n}^{\text{entry}}_{i,0}/\hat{n}^{\mathcal{A}}_{i,0}}$. Then, employment from new openings would be calculated as $\hat{e}^{\mathcal{B}_{i,t}}_{i,t} = \sum_{\ell\in i}(\frac{\hat{n}^{\text{entry}}_{i,t}/\hat{n}^{\mathcal{A}}_{i,t}}{\hat{n}^{\text{entry}}_{i,0}/\hat{n}^{\mathcal{A}}_{i,0}})^{-1} \times \hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t}) \times \hat{e}_{\ell,t}$, or equivalently $\hat{e}^{\mathcal{B}_{i,t}}_{i,t} = \sum_{\ell\in i} \theta^{\mathcal{B}}_i \times p(\mathcal{B}_{i,t}) \times \hat{n}^{\mathcal{A}}_{i,t} \times \hat{e}_{\ell,t}/\hat{n}^{\text{entry}}_{i,t}$, with $\theta^{\mathcal{B}}_i = \hat{n}^{\text{entry}}_{i,0}/\hat{n}^{\mathcal{A}}_{i,0}$ Thus $\hat{e}^{\mathcal{B}_{i,t}}_{i,t}$ would be the estimated number of new births ($p(\mathcal{B}_{i,t}) \times \hat{n}^{\mathcal{A}}_{i,t}$) times average employment of newly entering establishments ($\sum_{\ell\in i} \hat{e}_{\ell,t}/\hat{n}^{\text{entry}}_{i,t}$).

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 $\theta^{\mathcal{B}}_{\ell,t} = \theta^{\mathcal{B}}_i \times (\frac{\hat{n}^{\text{entry}}_{i,t}/\hat{n}^{\mathcal{A}}_{i,t}}{\hat{n}^{\text{entry}}_{i,0}/\hat{n}^{\mathcal{A}}_{i,0}})^{-1}$ and calculating the adjustment factor $\theta^{\mathcal{B}}_i$ 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 $\theta^{\mathcal{D}}_i$: establishments with 5 to 49 employees are pooled together for Retail Trade, Education & Health Services, and Other Services, 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

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

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 $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t})$. 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, $\hat{p}_{\text{BED/BDS}}(\text{birth}_{i,q})$, on the $\hat{p}_{\text{HB}}(\text{birth}_{i,q})$'s. We find that this approach does a better job at controlling for changes in $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t})$ 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 D4 in Appendix D.4 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 $\hat{p}(\mathcal{D}_{\ell,t}|\text{exit}_{\ell,t})$

[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.

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:

ℓ active in both t and $t-1$ → $\hat{e}_{i,t-1}^{\mathcal{A}_{i,t}} = \sum_{\ell \in i} \hat{e}_{\ell,t-1}^{\mathcal{A}}$, $\hat{e}_{i,t}^{\mathcal{A}_{i,t}} = \sum_{\ell \in i} \hat{e}_{\ell,t}^{\mathcal{A}}$

Approach:

ℓ active in $t-1$ but not in t
- → returns in $t+n$ → $\hat{e}_{i,t-1}^{\mathcal{T}_{i,t}} = \sum_{\ell \in i} \hat{e}_{\ell,t-1}^{\mathcal{T}}$
- → without return before the end of sample
  - → Google-closed or matched to FB → identifies **sample churn** among **exits** from HB; → $\hat{e}_{i,t-1}^{\mathcal{D}_{i,t}} = \sum_{\ell \in i} \theta_{\ell,t}^{\mathcal{D}} \times \hat{p}(\mathcal{D}_{\ell,t} | \text{exit}_{\ell,t}) \times e_{\ell,t-1}^{\text{exit}}$
  - → neither Google-closed nor matched to FB → $\hat{e}_{i,t-1}^{\mathcal{D}_{i,t}} = \sum_{\ell \in i} \theta_{\ell,t}^{\mathcal{D}} \times \hat{p}(\mathcal{D}_{\ell,t} | \text{exit}_{\ell,t}) \times e_{\ell,t-1}^{\text{exit}}$

Properties:	Varying over time t?	Specific to estab. ℓ?	Specific to cell i?
$\hat{p}(\mathcal{D}_{\ell,t} \mid \text{exit}_{\ell,t})$	**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:

ℓ active in t but not in $t-1$
- → active in $t-n$ → $\hat{e}_{i,t}^{\mathcal{R}_{i,t}} = \sum_{\ell \in i} \hat{e}_{\ell,t}^{\mathcal{R}}$
- → in HB data for the first time in t
  - → matched to FB → identifies **sample churn** among **entries** in HB; → $\hat{e}_{i,t}^{\mathcal{B}_{i,t}} = \sum_{\ell \in i} \theta_{\ell,t}^{\mathcal{B}} \times \hat{p}(\mathcal{B}_{\ell,t} | \text{entry}_{\ell,t}) \times \hat{e}_{\ell,t}^{\text{entry}}$
  - → not matched to FB → $\hat{e}_{i,t}^{\mathcal{B}_{i,t}} = \sum_{\ell \in i} \theta_{\ell,t}^{\mathcal{B}} \times \hat{p}(\mathcal{B}_{\ell,t} | \text{entry}_{\ell,t}) \times \hat{e}_{\ell,t}^{\text{entry}}$

Properties:	Varying over time t?	Specific to estab. ℓ?	Specific to cell i?
$\hat{p}(\mathcal{B}_{\ell,t} \mid \text{entry}_{\ell,t})$	**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.

from our Google / Facebook approach, and multiplied by the industry-size adjustment factor for closing $\theta_i^{\mathcal{D}}$. The adjustment factor is calibrated to match average BED death rates in 2019. For any subsequent period, since $\theta_i^{\mathcal{D}}$ 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 $\theta_i^{\mathcal{D}}$'s. For **new openings**, we multiply new entrants in HB in a given week t by the Google / Facebook probability $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t})$, weight them by the inverse of $\frac{\hat{n}_{i,t}^{\text{entry}}/\hat{n}_{i,t}^{\mathcal{A}}}{\hat{n}_{i,0}^{\text{entry}}/\hat{n}_{i,0}^{\mathcal{A}}}$ 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 $\theta_i^{\mathcal{B}}$. $\theta_i^{\mathcal{B}}$ 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, $\hat{n}_{i,t}^{\text{entry}}$, relative to $\hat{n}_{i,t-1}^{\mathcal{A}}$.

We now present implementation results on our approach.

For **permanent closings**, the probability $\hat{p}(\mathcal{D}_{\ell,t}|\text{exit}_{\ell,t})$ derived from Google and Facebook is described in Table D1. The left panel of Table D4 presents the adjustment factors $\theta_{\ell,t}^{\mathcal{D}} = \theta_i^{\mathcal{D}}$ 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.

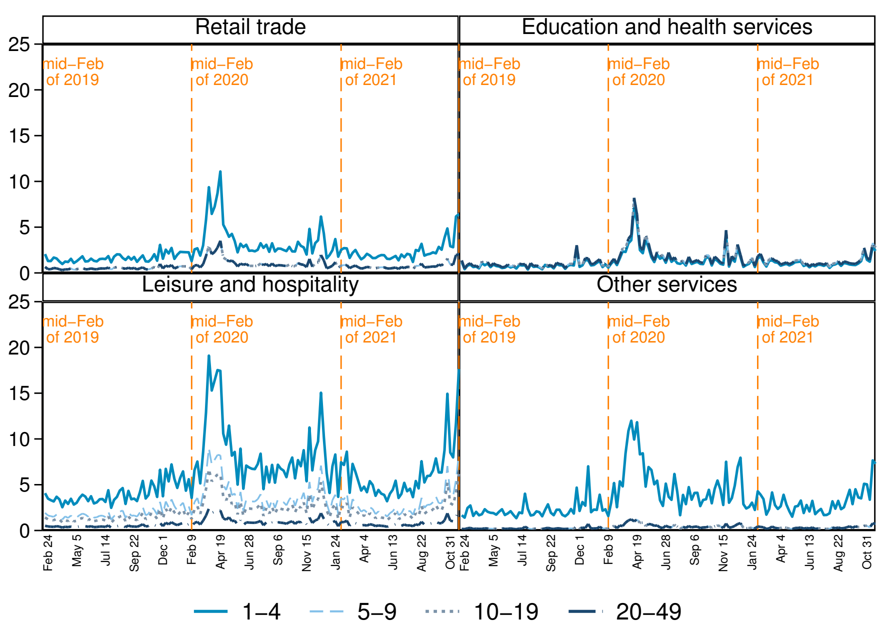
For **new openings**, the probability $\hat{p}(\mathcal{B}_{\ell,t}|\text{entry}_{\ell,t})$ is described in Table D2 and the adjustment factors $\theta_{\ell,t}^{\mathcal{B}}$, averaged over the weeks of 2019, is described in Table D4. Similarly as for closings, the adjustment factors tend to be larger 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 $\theta_{\ell,t}^{\mathcal{B}}$'s for each sector and size class over time. Recall that the time variation is coming from $(\frac{\hat{n}_{i,t}^{\text{entry}}/\hat{n}_{i,t}^{\mathcal{A}}}{\hat{n}_{i,0}^{\text{entry}}/\hat{n}_{i,0}^{\mathcal{A}}})^{-1}$. 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 $\theta_{\ell,t}^{\mathcal{B}}$. 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 $\theta_{\ell,t}^{\mathcal{B}}$ 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/BDS 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 $(q, q+1, q+2, q+3)$.

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. BDS exit and entry rates are computed somewhat differently than BED entry and exit rates.[19] As Figure D2 shows, BDS and BED annual entry and exit rates line up closely for Retail Trade and Leisure & Hospitality, but less so for Education & Health Services (NAICS 61-62) and Other

[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).

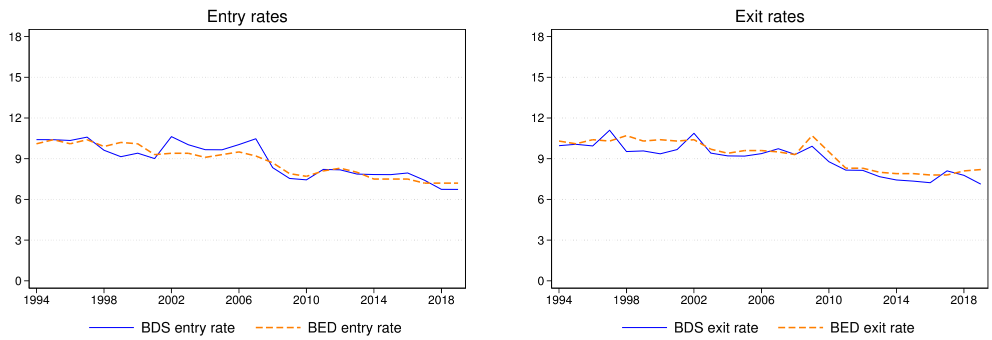
[19] BDS establishment entry and exits are defined as rates in a way similar to the BED's *annualized* opening and closing rates, defined as

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

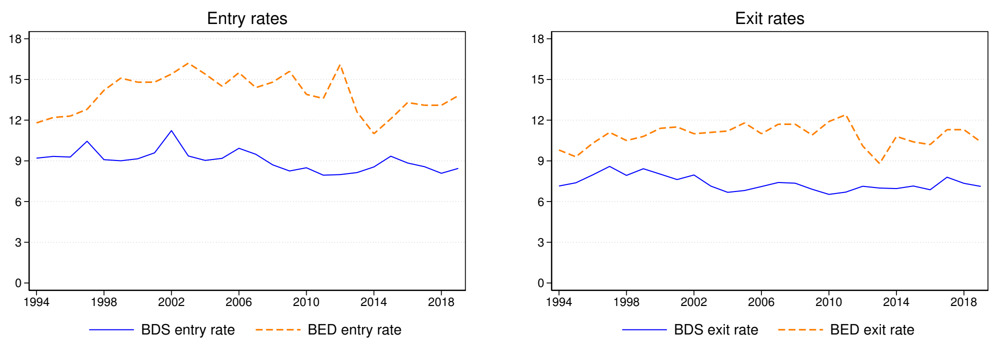
We use these annual BED openings and closing rates to compare to the BDS annual entry and exit rates in Figure D2.

Figure D2: BDS and BED annual entry and exit rates

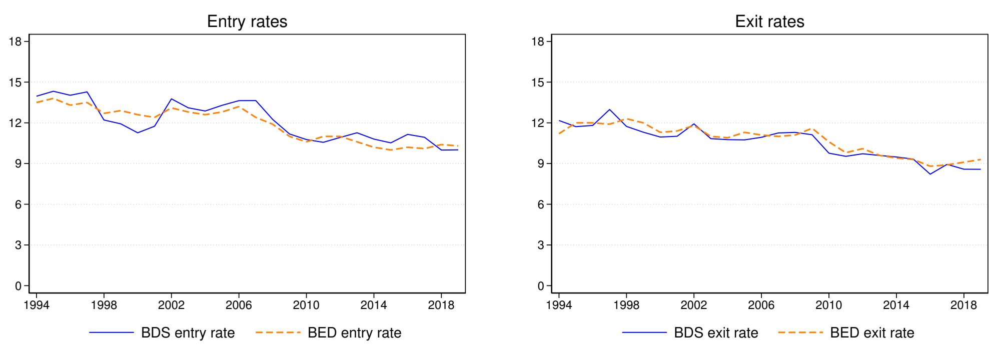
(a) Retail Trade



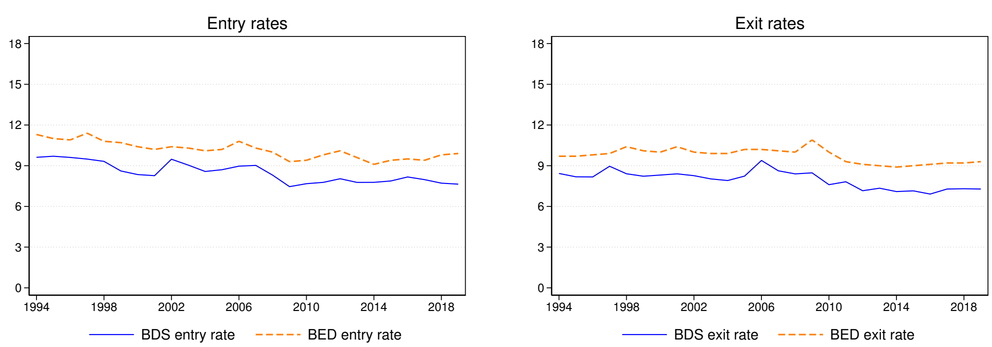
(b) Education & Health Services



(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).

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 5 years of data available before the pandemic to construct the conversion factors reported in Table D5. As the table shows, entry and exit rates decrease with establishment size.

Construction of HB counterparts to the BED/BDS birth and death rates. To compare HB to BED/BDS birth and death 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

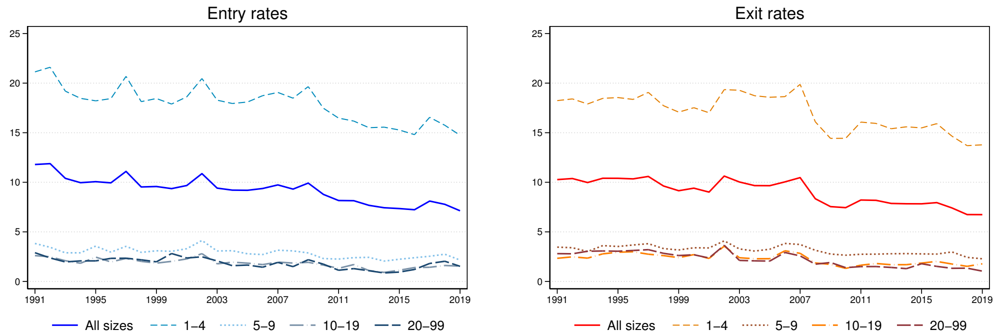
- HB total entry for quarter q as HB establishments with positive employment in the third month of quarter q and no employment in the third month of the preceding four quarters $(q-4, q-3, q-2, q-1)$;[21]

[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 2013Q1 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).

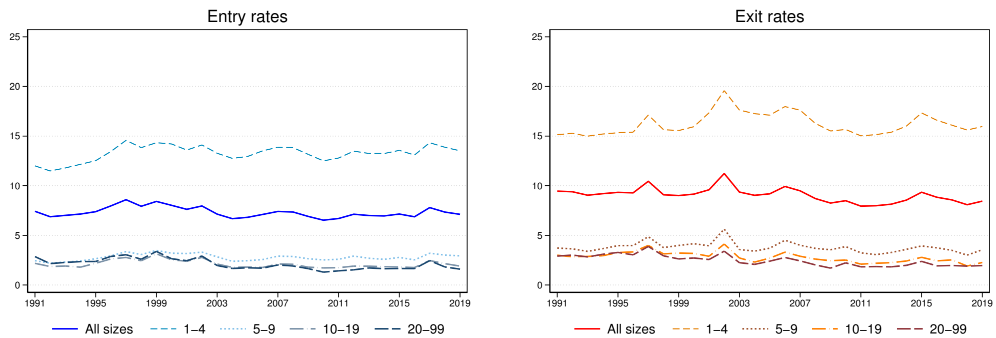
[21]As in the BED, no employment means either zero reported employment or no reported employment because the HB establishment appears for the first time in that quarter.

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

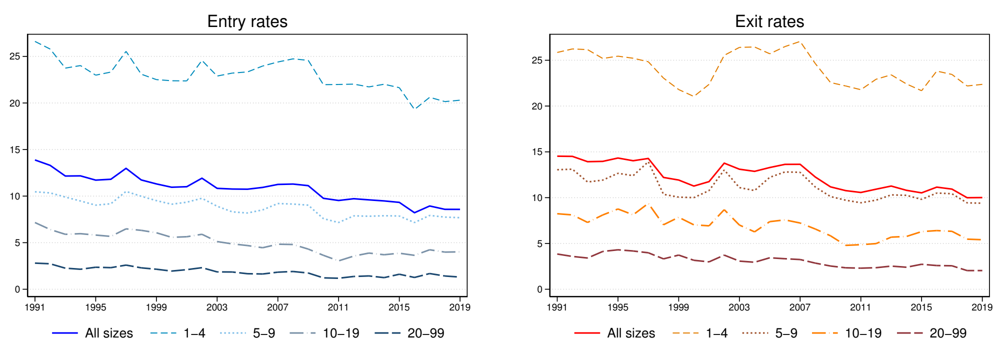
(a) Retail Trade



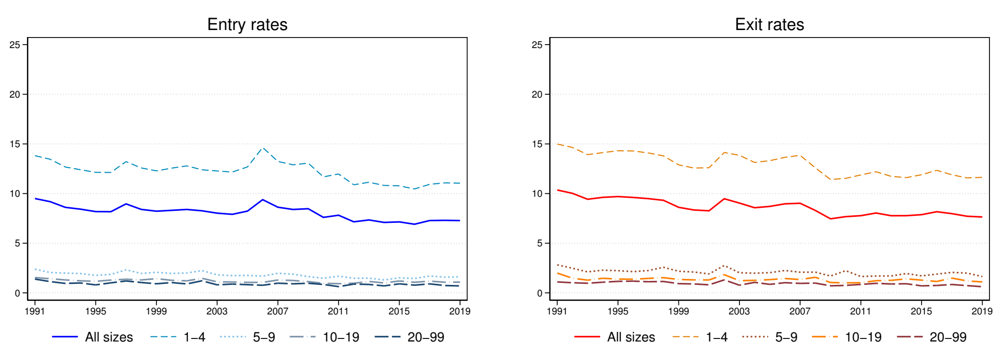
(b) Education & Health Services



(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.

- 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/BDS counterparts (see Figure D4 below and Figure 1 in the paper. On the other hand, the adjusted HB birth and death rates come very close to their BED/BDS 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]

Benchmarking against 2019 BED/BDS birth and death rates. Figure D4 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.

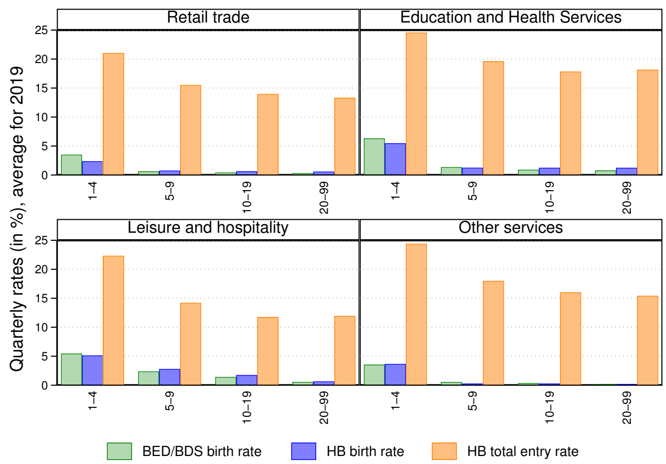
By construction, the HB birth and death rates implied by our estimation fit the BED/BDS benchmarks very closely (the fit is not perfect because, as described above and in the main text, we pool over some of the sector-size classes). In comparison, total new entry and permanent exit rates are much larger, 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.

[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.

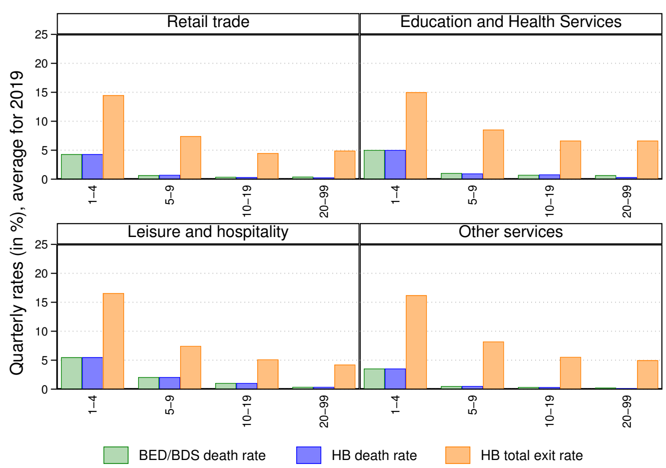
[23] One caveat is that although Google Places distinguishes "temporary closed" and "permanently closed", we treat both indicators the same and tag the establishment as closed. Google's "permanently closed" accounts for more than 80% of all the exiting establishments from HB that are found to be Google-closed.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

Figure D4: Benchmarking against 2019 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.

E Using Safegraph visits data to assess quality and representativeness

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

E.1 Representativeness of Homebase sample

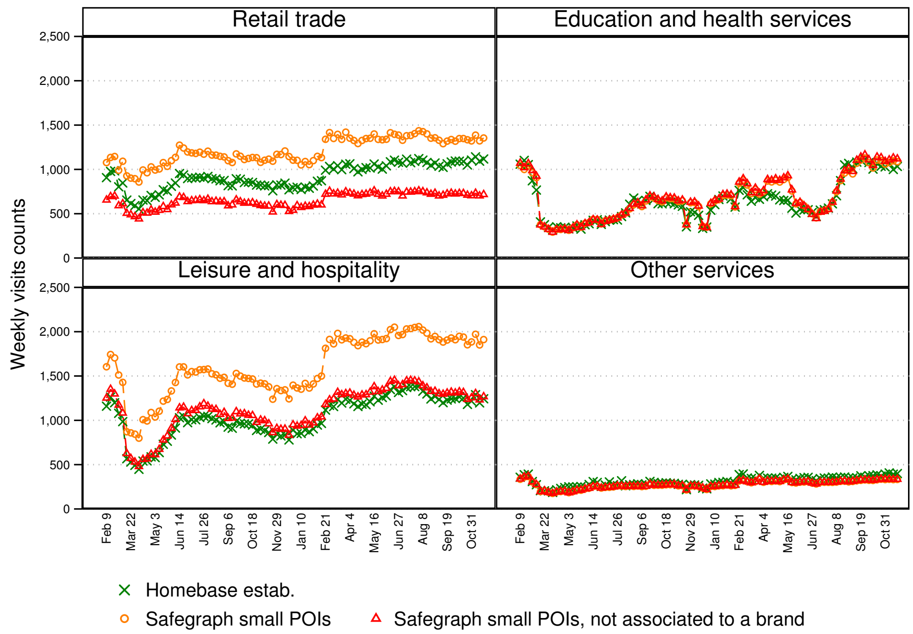
Having described the two data samples, we now compare them with respect to the various dimensions of foot traffic data measured by Safegraph. We analyze four such dimensions: weekly counts of visits, median dwell time and share of long visits (defined as daily visits longer than 240 minutes) among all 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 that 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.

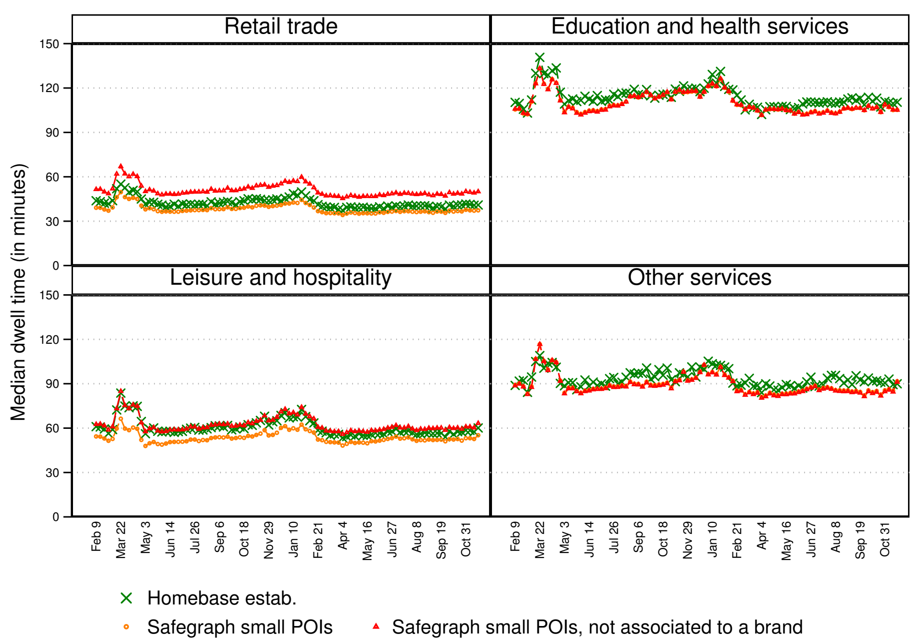
the 2020 base sample, and fewer than 100 workers to 95 percent of them. At the same time, there are differences across sectors. 92 percent of the Homebase establishments in Leisure & Hospitality have fewer than 50 workers according to Netwise employment, while the corresponding number is only 74 percent 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



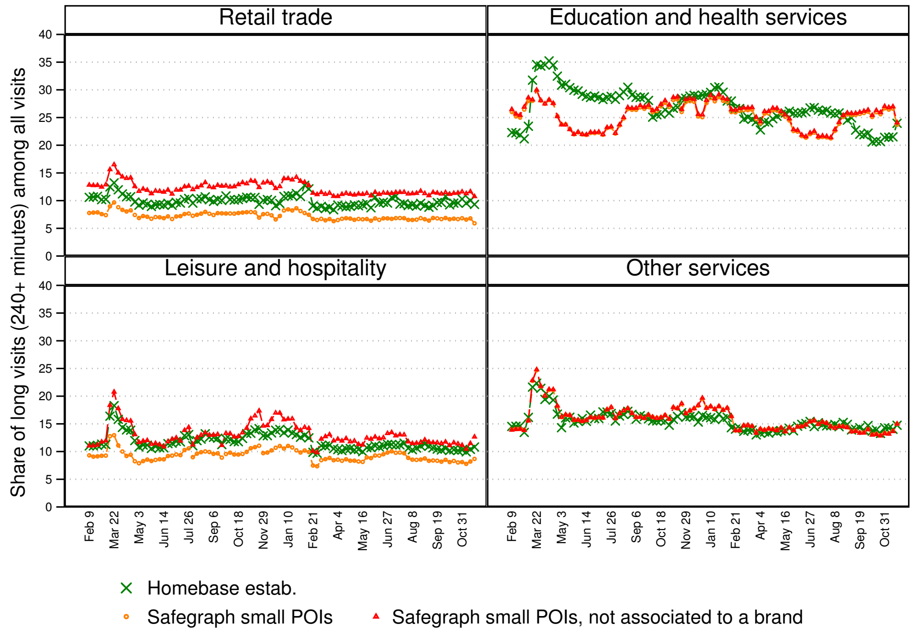
Notes: Safegraph visits data. The orange circles and red triangles denote Safegraph POIs with fewer than 50 workers according to NetWise employment data. The green crosses denote Homebase establishments matched to Safegraph visits data.

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



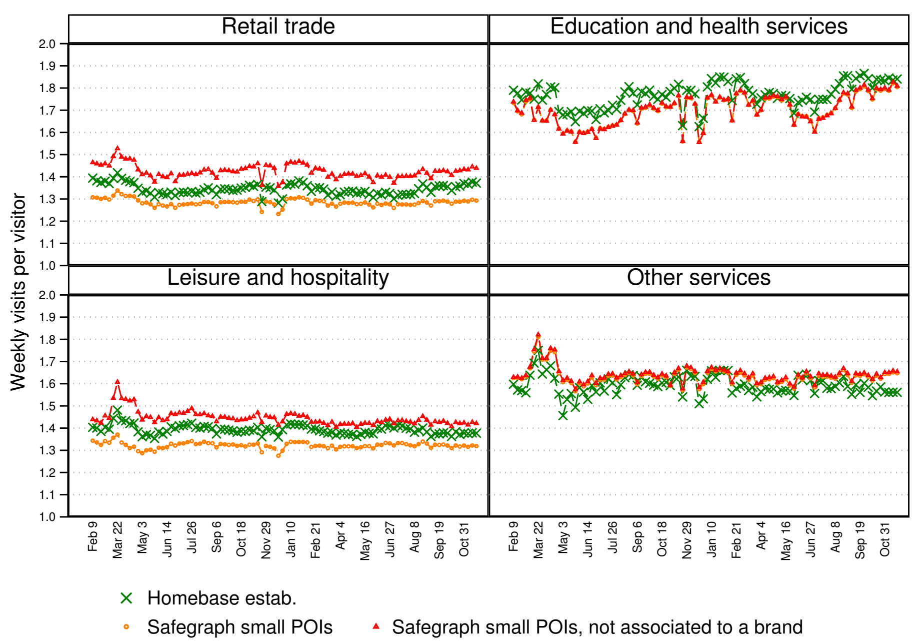
Notes: Safegraph visits data. The orange circles and red triangles denote Safegraph POIs with fewer than 50 workers according to NetWise employment data. The green crosses denote Homebase establishments matched to Safegraph visits data.

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.

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 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.

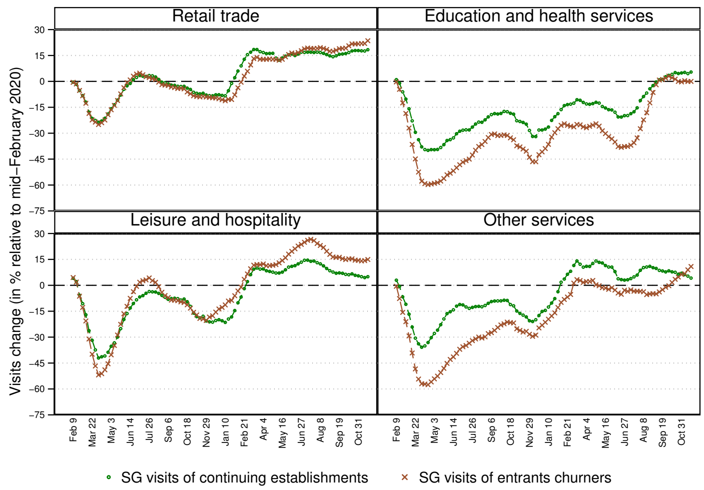
E.2 Ability to identify sample churn

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.

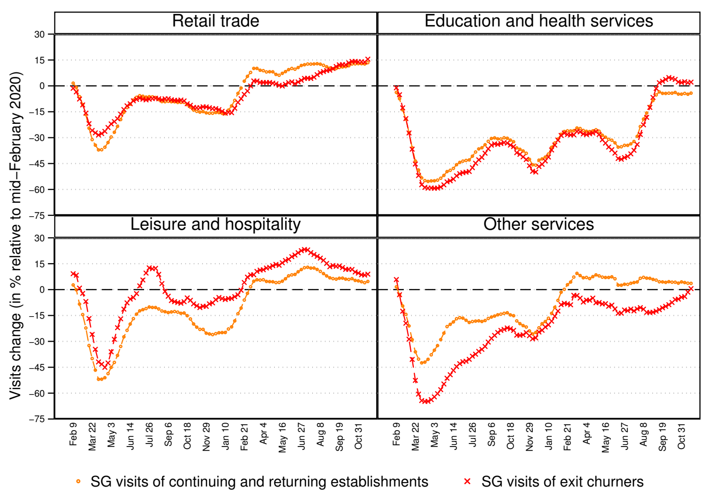
The upper panel of Figure E5 compares changes in visits among entrant churners (they enter HB after the mid-February base period and we identify that they were already operating before) with the continuously active establishments from the base sample. The latter is the relevant comparison group as it 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).

E.3 Representativeness of entrants and exits matched to Google/Facebook

Recall from Section D.1 that for permanent closings in HB we attempt to obtain information from Google and Facebook to assess whether these establishments continue to operate outside of HB. For exiting establishments that cannot be matched to either Google Places or Facebook, we identify them as closed with a probability estimated out of the matched establishments. Likewise, for new entries we attempt to match them to Facebook to estimate whether they operated already prior to entering HB, and we attribute a probability of new opening to the non-matched establishments based on the sample of matched establishments. The underlying assumption is that matching to Google or Facebook does not induce any selection among permanently closed, respectively newly entering establishments. In this section, we assess the plausibility of this assumption.

Table E2: Distribution of permanent closings and new openings

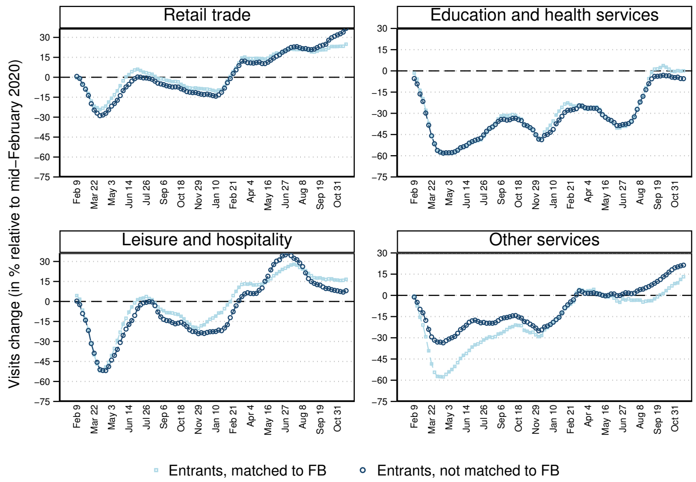
New openings									
Matched to FB					**Not matched to FB**				
Class size	Retail Trade	Education & Health	Leisure & Hospitality	Other Services	Class size	Retail Trade	Education & Health	Leisure & Hospitality	Other Services
1–4	9.2	3.6	7.5	2.7	1–4	9.2	0.9	8.3	2.2
5–9	11.8	6.3	16.4	3.9	5–9	11.0	5.2	18.4	3.4
10–19	5.6	4.0	16.2	1.8	10–19	6.0	3.6	17.6	1.5
20–99	1.7	1.7	7.1	0.5	20–99	2.2	1.7	8.3	0.5
Permanent closings									
Google-closed or matched to FB					**Neither Google-closed nor matched to FB**				
Class size	Retail Trade	Education & Health	Leisure & Hospitality	Other Services	Class size	Retail Trade	Education & Health	Leisure & Hospitality	Other Services
1–4	11.0	3.4	10.7	3.6	1–4	8.1	2.1	10.8	2.4
5–9	10.2	5.9	18.2	3.8	5–9	8.5	4.2	23.1	2.7
10–19	4.5	3.6	14.5	1.5	10–19	4.1	2.9	19.5	1.3
20–99	1.4	1.6	5.7	0.4	20–99	1.3	1.3	7.1	0.4

Notes: Distribution of the 2020 samples of new openings (upper panel) and permanent closings (lower panel) by sector and establishment size, conditional on matching to Google or Facebook (FB)

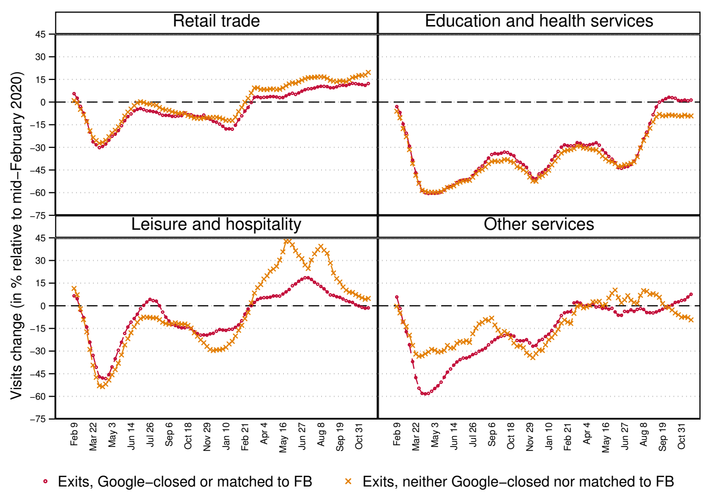
We begin in Table E2 by reporting the sample composition of new openings (upper panel) and permanent closings (lower panel) in terms of sector and establishment size. As can be seen by comparing the left-hand side of the table, referring to establishments that can be matched to Google or Facebook, with the right-hand side, there is no evidence that matching systematically selects establishments in terms of their sector or size. This holds true for both new entries and exits. Next, in Figure E6, we use visits data to compare the different samples. Note the relation to Figure E5: entrant churners in Figure E5 are establishments that can be matched to Facebook and for which our procedure establishes that they operated already prior to entering HB, hence they are a subset of the establishments denoted as "matched to FB" in Figure E6, Likewise, exit churners are a subset of those denoted by the dotted

Figure E6: Safegraph visits among Google- or Facebook-matched vs. not-matched establishments

(a) New openings



(b) Permanent closings



Notes: Visit changes by new openings that are matched or not to FB (upper panel), and permanent closings that are Google-closed or matched to FB or are neither (lower panel) 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).

lines in the lower panel of Figure E6. We find that the matched establishments behave fairly similarly to establishments that cannot be matched to Google or Facebook.

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

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

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 7 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

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

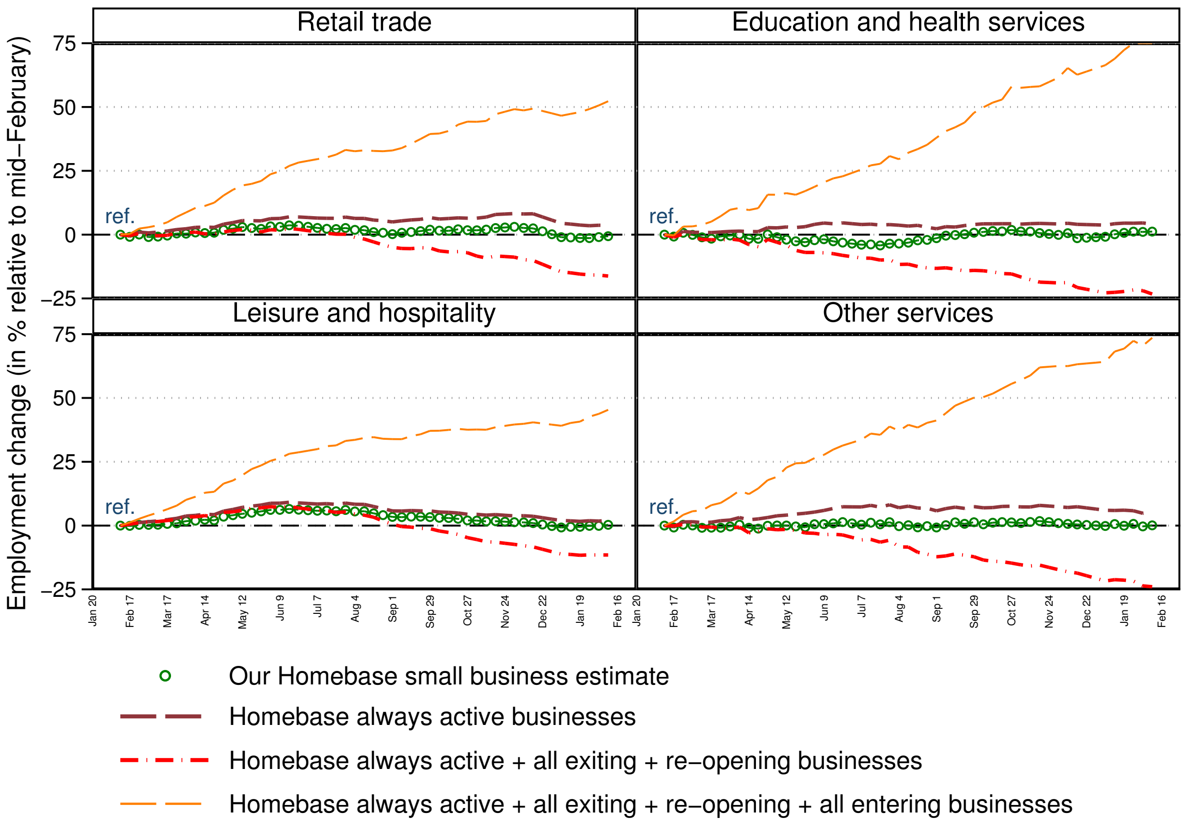
In Section 5 of the paper, we also discuss recall rates in our data. The recall rate in week t is defined as the ratio between recalled employees and the sum of recalled employees and new hires in week t.

G Additional figures and tables

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

Figure G1 is the counterpart of Figure 4 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

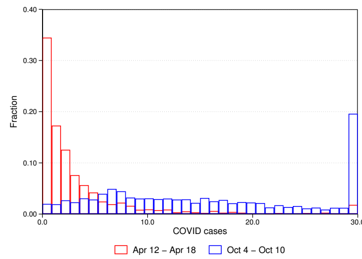
Table G1 describes the main variables used in the regressions reported in the paper. Table G2 provides descriptive statistics for the main variables. Figure G2 complements the description of time-varying control variables by showing distributions for the weeks of April 12 – April 18 (when small business 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

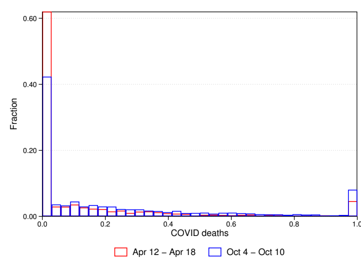
Variable	Description	Source
COVID_cases	Number of COVID cases per 1,000 county inhabitants	COVID Act Now
COVID_deaths	Number of COVID deaths per 1,000 county inhabitants	https://covidactnow.org/
containment_index	Weighted measure of restrictions for all containment measures (school closure, workplace 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

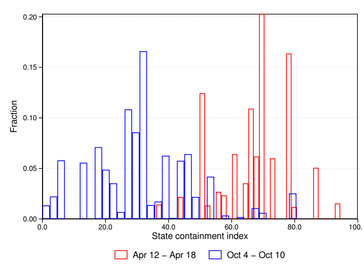
(a) COVID cases



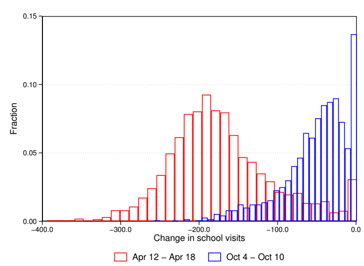
(b) COVID deaths



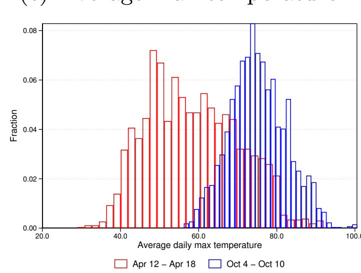
(c) Containment index



(d) Change in school visits



(e) Average max temperature



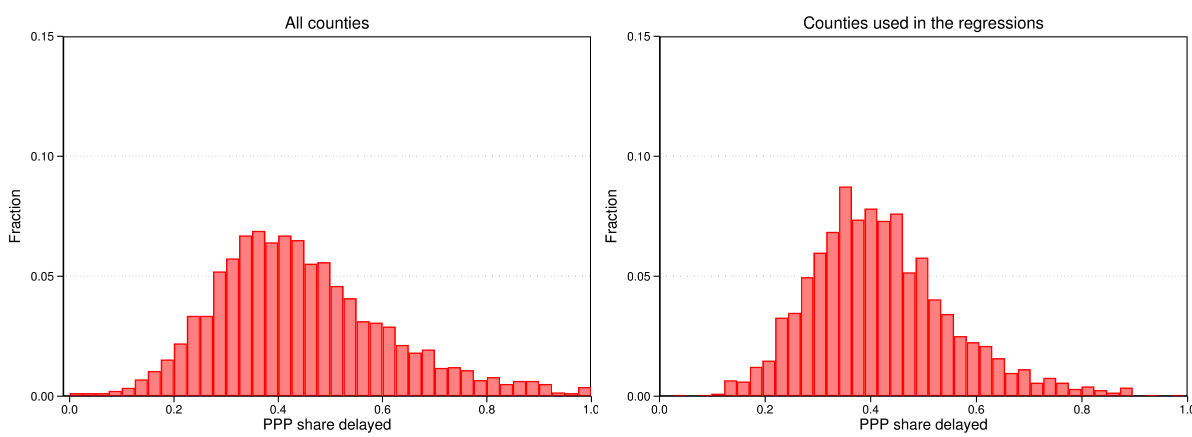
Notes: Distribution of control variables for the weeks of Apr 12 - Apr 18 and Oct 4 - Oct 10.

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. [2023] or Goolsbee and Syverson [2021] 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

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

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

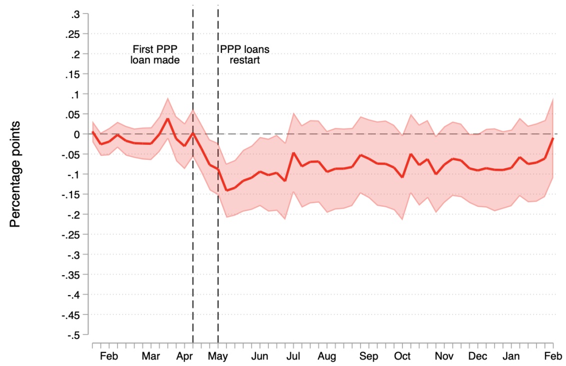
Table G3: Estimates for different control variables

	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	✔	✔	✔	✔

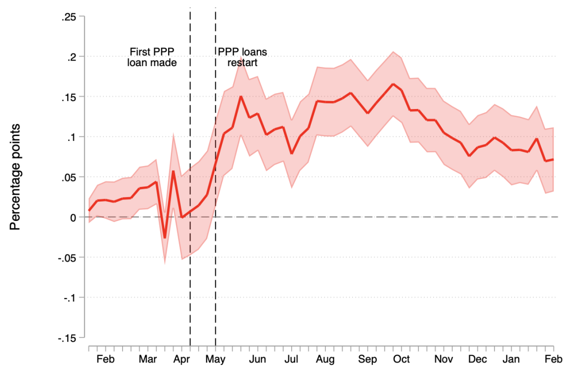
Notes: Standard errors are clustered at the county level; * p < 0.10, ** p < 0.05, *** p < 0.01. All regressions are estimated over all weeks beween February 9-15, 2020 and January 31 - February 6, 2021. Table only shows coefficient estimates for COVID health NPI, school visit and maximum temperature 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

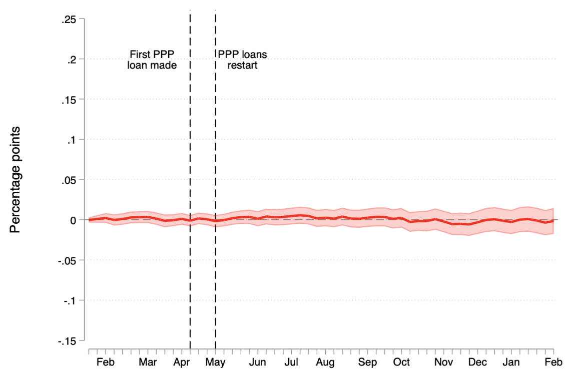
(a) Employment of always active businesses



(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: Evidence from traditional and non-traditional data. *Brookings Papers on Economic Activity - Special Edition: COVID-19 and the Economy*, pages 239–268, July 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. *Quarterly Journal of Economics*, forthcoming, 2023.

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. *Journal of Public Economics*, 193:104311, 2021.