NBER WORKING PAPER SERIES

LOCAL VARIATION IN ONSITE WORK DURING THE PANDEMIC AND ITS AFTERMATH

Katharine G. Abraham
Mohammad Ashoori
Aref Darzi
Nathalie Gonzalez-Prieto
John C. Haltiwanger
Aliakbar Kabiri
Erkut Y. Ozbay

Working Paper 32042
http://www.nber.org/papers/w32042

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2024

Abraham, Haltiwanger and Ozbay are members of the faculty in the Department of Economics, University of Maryland; Ashoori, Darzi, and Kabiri are affiliated with the Center for Advanced Transportation Technology Laboratory (CATT Lab), University of Maryland; and Gonzalez is a Research Economist at the World Bank. We thank the National Science Foundation for generous support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Local Variation in Onsite Work during the Pandemic and its Aftermath
Katharine G. Abraham, Mohammad Ashoori, Aref Darzi, Nathalie Gonzalez-Prieto, John C. Haltiwanger, Aliakbar Kabiri, and Erkut Y. Ozbay
NBER Working Paper No. 32042
January 2024
JEL No. J21,L23,R23,R40

ABSTRACT

Using longitudinal data on the location of mobile devices, we provide new evidence on the evolution of onsite work (OSW) over the course of the pandemic and its aftermath. We start with a large sample of individuals who, based on their mobile device activity, had a job at which they worked onsite in February 2020. We track the evolution of these individuals' onsite work activity over the following thirteen to fourteen months, observing them in May 2020, August 2020, November 2020 and March/April 2021. Consistent with other evidence, we find a dramatic decline in OSW in May 2020 followed by a substantial rebound by the spring of 2021, albeit to a lower level than in February 2020. We document considerable cross-state, cross-city and cross-county variation in OSW. We also find, however, that the tract-level variation in OSW within states, cities and even counties far exceeds the variation across larger geographic areas. Observable characteristics such as industry, occupation, education and income account for much of the variation in OSW across large geographic areas since the pandemic. These same variables account for much of the enormous cross-tract variation in OSW that remains after controlling for state or county, but more than half of the cross-tract variation is accounted for by residual factors. These findings imply considerable heterogeneity in how the pandemic has affected where the resident populations of U.S. neighborhoods spend their days, a finding that has significant implications for businesses, workers, and policymakers.

Katharine G. Abraham
Department of Economics and
Joint Program in Survey Methodology
University of Maryland
1218 LeFrak Hall
College Park, MD 20742
and NBER
kabraham@umd.edu

Mohammad Ashoori
University of Maryland
mashoori@umd.edu

Aref Darzi
Center for Advanced Transportation
Technology Laboratory
University of Maryland
College Park, Mary 20742
adarzi@umd.edu

Nathalie Gonzalez-Prieto
The World Bank
ngonzalezprieto@worldbank.org

John C. Haltiwanger
Department of Economics
University of Maryland
College Park, MD 20742
and NBER
haltiwan@econ.umd.edu

Aliakbar Kabiri
Center for Advanced Transportation
Technology Laboratory University
of Maryland
College Park, MD 20742
kabiri@umd.edu

Erkut Y. Ozbay
Department of Economics
University of Maryland
College Park, MD 20742
ozbay@umd.edu

I. Introduction

The enormous shift away from onsite work at offices and other business locations during the early stages of the pandemic has received considerable attention (see, e.g., Barrero et al. 2021; Bick et al., 2021). For many workers, continued work from home appears to be the new normal (Barrero et al. 2021; Aksoy et al. 2022; Brynjolfsson et al. 2023; Hansen et al. 2023). Research has documented significant differences in the shift towards remote work in the post-pandemic period by industry, occupation, educational attainment and income, but has had less to say about how this shift has varied across geographic areas.

In this paper, we examine spatial variation in the probability that individuals working at an onsite location continued to work onsite over the year following the onset of the pandemic. Our analysis makes use of mobile device location data that we use to identify individuals' home and work locations. We begin with a sample of about 4.2 million people with identified February 2020 home and work locations. Then, we track these same individuals forward in time with snapshots in May 2020, August 2020, November 2020 and March-April 2021 (the latest time period for which our longitudinal sample supports estimates of acceptable quality), algorithmically identifying the home location and, if there was one, the onsite work location for each individual remaining in the sample in these subsequent months. To help with the interpretation of changes in the prevalence of onsite work, we also conduct a similar baseline analysis that starts with a sample of individuals for whom we can identify home and work locations as of February 2019, then track those individuals forward to May 2019 and August 2019.

Our analysis makes use of mobile device activity data from a repository created by researchers at the Maryland Transportation Institute (MTI) and the Center for Advanced

Transportation Technology Laboratory (CATT Lab). Based on location information collected through numerous common smart device apps, this infrastructure allows us to track the locations of millions of smart devices. The primary application of these data has been to track and model transportation activity (see, for example, Zhang et al., 2022, Zhang et al., 2023). At the start of the pandemic, MTI researchers used the mobile device location data to create a real-time dashboard that included geographically disaggregated measures of social distancing (see https://data.covid.umd.edu/about/index.html). Our analysis builds on the MTI/CATT Lab data infrastructure.

In our sample of devices, we designate the location where we observe the device most frequently during a month as the home location, provided we observe it in that location during at least 60 unique hours and on at least 14 unique days during the month. We designate the second most frequently observed location as the work location, provided it meets a minimum threshold number of observed hours in that location and is observed there during at least two unique weeks during the month. The logic of this approach to identifying home and work locations has been discussed in more detail in previous research (Pan et al., 2023). Our thresholds for onsite work are such that we should capture a work location for an individual working onsite either full-time or on a regular and substantial part-time schedule.

For privacy and confidentiality reasons, we aggregate the data to the Census tract level based on place of residence rather than reporting individual-level results. Most of the results we report are based on tracts for which we observe at least ten devices with a home location in each of the periods represented in our analysis, though our main findings are not sensitive to being either less restrictive or more restrictive with respect to which tracts we include. The primary outcome of interest in our analysis is the percentage of people who had been working onsite in

February 2020 who have onsite work activity in subsequent months. Based on our measurement criteria, in our sample of tracts with ten or more usable devices in each observation period, just 53 percent of those who had been working onsite in February 2020 were working onsite in May 2020, two months into the pandemic. This share increased to about 63 percent by August 2020 and to about 67 percent in March-April 2021, a little more than a year later.

Although these average effects are of interest in their own right, we also find enormous spatial variation in the prevalence of post-pandemic onsite work. Again, restricting attention to tracts that met our ten-device cutoff, in the 90^{th} percentile (employment weighted) tract (measured by place of residence), 67 percent of February onsite workers were working onsite in May 2020; in the 10^{th} percentile tract, only 38 percent were doing so, a gap of 29 percentage points. This dispersion appears to have widened slightly over time. As of March-April 2021, in the 90^{th} percentile tract, 83 percent of February 2020 onsite workers were working onsite; in the 10^{th} percentile tract, this was true of only 49 percent of February 2020 onsite workers, a gap of 34 percentage points. Some of this dispersion reflects variation across states or counties, but even within counties there is substantial variation across neighborhoods. A unique feature of our analysis is our ability to document and provide insights regarding this neighborhood level variation.

With a rich set of tract level characteristics from a variety of sources, we are able to account for about 58 percent of the May 2020 across-tract spatial variation in onsite work. Industry and occupation mix account for about half of the explained variation, but the education and household income of tract residents as well as the severity of the initial COVID wave and the persistence of initial lockdown restrictions in the county where the tract is located also have notable explanatory power. Even after accounting for these other factors, political preferences appear to matter—the share of voters in the tract who voted for Donald Trump in 2016 accounts for about 7 percent of the cross-tract variation. By March-April 2021, this same rich set of covariates accounts for only about 34 percent of the spatial variation. Industry and occupation mix continue to be important, as do political preferences, which are slightly more important with respect to the share of cross-tract variation they explain in March-April 2021 than in May 2020. No other factor explains more than a few percentage points of the cross-tract variation.

One possible explanation for these patterns is that they reflect differences in normal employment dynamics across tracts. To explore this possibility, we consider first-difference specifications that examine the relationship between the *change* in tract-level onsite work and the same tract-level characteristics used in the regressions just described. Perhaps not surprisingly, observable covariates account for less of the cross-sectional variation in the change in onsite work between May 2019 and May 2020, measured in both cases for a sample of people who had been engaged in onsite work in February of the same year, than of the cross-sectional variation in May 2020 (45 percent versus 58 percent). Industry and occupation mix are both important, as is average household income in the tract. COVID deaths and restrictions also play a small role. Interestingly, political preferences no longer appear to matter, suggesting that what at first blush appeared to be evidence of a strong partisan divide in the response to the pandemic may instead reflect other unobserved differences in employment persistence or the likelihood of continuing employment at a fixed place of work that are correlated with political preferences. The pattern of cross-sectional variation in the change in onsite work between August 2019 and August 2020 is qualitatively similar to that in the May 2019-May 2020 changes.

The factors that matter in the models just described translate into distinct differences in the prevalence of onsite work across large geographic areas such as states and metropolitan areas. For example, through early 2021, Texas and Florida had higher average rates of onsite work among people who had worked onsite in February 2020 than either New York or California. Similarly, Houston and Miami had substantially higher rates of onsite work than either New York City or, especially, San Francisco over the same period. The differences in the average levels of onsite work observed in these examples largely can be attributed to differences in the observable factors mentioned above.

Although we observe differences across states and metropolitan areas in the subsequent level of onsite work among individuals who had a February 2020 onsite work location, even within a state or metropolitan area, there is significant variation in its prevalence. In other exercises, we examine the extent to which spatial differences in subsequent onsite work following the onset of the pandemic reflect state effects or county effects as opposed to differences across smaller geographic areas. State effects on their own explain only about 22 percent of the cross-tract variation as of May 2020 and less in subsequent months. After controlling for the observable characteristics already discussed, they add almost nothing to our models' explanatory power. County fixed effects on their own can explain considerably more of the cross-tract-level variation, but again, adding them to the regressions containing other explanatory variables contributes relatively little to our models' explanatory power and much of the cross-tract variation remains unexplained. A key take-away from this portion of the analysis is that the pandemic led to large neighborhood-level differences in the evolution of onsite work.

The paper proceeds as follows. Section II provides a more extensive review of the literature to which our paper contributes. Section III describes the data and measurement. Section IV presents basic facts about the spatial variation in onsite work. We analyze the factors that account for this spatial variation in Section V. Section VI concludes.

II. Literature Review

The COVID-19 pandemic precipitated significant changes in behavior. Prompted by fear of the disease together with government-mandated lockdowns, many people began to avoid activities that involved close contact with others. Social distancing was a means of slowing the spread of a disease for which, at that point, there was no vaccine and no proven effective treatment. Although social distancing was widespread, it was not universal (see, for example, Allcott et al. 2020).

An important dimension of social distancing during the early months of the pandemic was a sharp increase in the prevalence of work from home (WFH) and a corresponding decline in the prevalence of onsite work (OSW). As with social distancing more generally, WFH was not distributed evenly across the population. Nor was it clear whether WFH was a temporary phenomenon or one that would continue after the pandemic abated. Although widespread adoption of WFH was forced on employers, even early on, some scholars argued that positive experiences with remote work and the development of new tools to facilitate it would cause WFH to persist (Barrero et al. 2021). Researchers studying WFH and its evolution over the post-pandemic period have taken several different approaches.

One approach has been to assess the possibilities for remote work using detailed job descriptions to characterize the tasks associated with different occupations. The most widely cited estimate is that about 37 percent of U.S. jobs could be performed remotely (Dingel and Neiman 2020). Of course, the fact that a job *could* be performed remotely does not mean that it *will* be performed remotely. This sort of analysis nonetheless has proven to be helpful for understanding differences across occupations in remote work prevalence.

Second, several different surveys have provided information on changes in WFH during the post-pandemic period.[2] From May 2020 to September 2022, the Current Population Survey (CPS) asked, "At any time in the LAST 4 WEEKS, did you telework or work at home for pay BECAUSE OF THE CORONAVIRUS PANDEMIC?" (Dey et al. 2021). This was a sensible question to ask early in the pandemic. Over time, however, workers may have come to view WFH as normal rather than as due to the pandemic, creating uncertainty about exactly what the responses to the question were capturing. The Real-Time Population Survey (RPS), administered using the Qualtrics online panel from May 2020 through June 2021, also included questions about WFH. The survey was designed to be representative of the U.S. population along a variety of dimensions. The RPS identified WFH by asking the number of days respondents worked in the previous week and how many of those days they commuted. The survey also asked respondents in each wave of the survey the same questions about their work, if any, in February 2020, allowing the authors to produce estimates of both the level and the change in WFH over their study period (Bick et al. 2022). The Survey of Working Arrangements and Attitudes (SWAA), an online survey conducted monthly since May 2020, is another source of data on trends in WFH. Prior to June 2022, the survey sample was restricted to individuals with significant prior year work attachment. The remote work questions on the SWAA also have evolved over time, but since November 2020 have included questions asking respondents how many full days they worked in the previous week and, of those days, how many they worked from home (Barrero et al. 2021).

[2] Reflecting differences in the questions asked and the methods used to collect the data, the surveys have produced different estimates of the prevalence of WFH. Brynjolfsson et al. (2023) provide a useful discussion of the factors that have contributed to these differences.

Surveys have the advantage that, in addition to asking whether an individual worked remotely, they can collect information about the characteristics of the person and their employment. WFH rose substantially more (in the CPS) and has been substantially more prevalent (in the RPS and SWAA) among workers with higher levels of education. Estimates from the RPS and SWAA also show higher WFH rates among higher-income individuals. Analyses of both the CPS and the RPS have found WFH to be substantially more common in occupations identified by Dingel and Neiman (2020) as compatible with remote work. While informative about broad patterns with respect to WFH, because of sample size limitations, surveys are not well suited for the production of geographically disaggregated statistics.

A third approach to measuring WFH is to use online job postings data to trace changes in the share of postings for jobs that permit remote work. Some of these studies have used the presence of pre-specified keywords to classify jobs as remote versus onsite (see, for example, Adrjan et al. 2021). Hansen et al. (2023) apply machine-learning methods to accomplish the classification task. Because the number of job postings available for analysis is large, the data can be disaggregated temporally, geographically and by occupation. Hansen et al. (2023) have constructed monthly county-level estimates of the share of job postings offering remote work for the period from January 2019 through April 2023 that cover most larger counties. While interesting, since postings relate to the flow rather than the stock of jobs, these estimates are not directly comparable to other estimates of WFH prevalence.

Finally, mobility data have been used to track changing patterns of workplace activity. The Google Community Mobility reports, available from early in the pandemic through October 2022, are one source of such information (see, e.g., Sehra et al. 2020, Jacobsen and Jacobsen 2020, Mendolia et al. 2021, and Rafiq et al. 2022). Based on mobile device location data, the creators of these reports constructed indexes of the number of visits to different types of locations (points of interest or POIs) relative to the number of visits during the five-week period from January 3 to February 6, 2020. Workplaces are one of the six location types; the others are groceries and pharmacies; retail and recreation; transit stations; parks; and residential. Daily data are available at the county level. The publicly available information about the methodology used to produce the indexes is sparse. One limitation of these series for measuring trends in travel to work is that all of the location types could be workplaces for the people employed there. Similar comments would apply to analyses of Safegraph, PlaceIQ and other similar data sources that use POI information to identify trip purposes. Jay et al. (2020) work with Safegraph mobile device location data using a different strategy to identify the prevalence of travel to work. Their measure is the share of devices that, on a given day, stopped at a location for three or more hours between 8:00 am and 6:00 pm (taken as an indication of onsite work) or visited four or more locations during the day for less than 20 minutes each (taken as an indication of delivery or similar work). Limitations of this measure are that visiting a friend, running errands or other non-work trips might show up as travel to work, while work that occurs outside of daytime hours might be missed.

Our approach is closest in spirit to the studies that have used mobility data to track changes in travel to work. We identify travel to work by identifying locations other than their home at which people regularly spend significant amounts of time. In contrast to existing studies, we take advantage of the fact that we observe mobile devices over a period of as long as 14 months. Our analysis identifies devices that are associated with people engaged in OSW as of February 2020 and then follows those devices longitudinally. This allows us to see how the probability of OSW is changing for individuals with a pre-existing work attachment. Another distinguishing feature of our work is that we consider estimates disaggregated to the Census tract level based on the home location of the device owners. As we will show, there is substantial cross-tract variation following the onset of the pandemic in the likelihood that an onsite worker continues in OSW. Although there is interesting variation in OSW at the state and county level, looking only at data aggregated to those levels does not tell the whole story about how the pandemic has affected travel to work patterns.

III. Data and Measurement

Our analysis makes use of the repository of mobile device location data created by MTI and the CATT Lab at the University of Maryland. The dataset utilized in this analysis was gathered by one of the leading U.S. location-based services data aggregators. The research team conducted essential data cleaning procedures that included the removal of entries with invalid values and deletion of duplicated observation from the dataset (see Zhang et al. 2023 for additional details). To give a sense of the scale of the data, for February 2020, the repository contains information for more than 150 million devices observed on one or more occasions during the month. Only a fraction of these devices, however, were observed with sufficient frequency to be usable for our purposes.

Using the information in the repository, we construct a sample of mobile devices for which we are able to identify a home location and an onsite work location as of February 2020. Because mobile devices typically are not shared, we refer to these as individual home and work locations. We begin by dropping devices that were not observed at least 100 times during the month. The February 2020 home location is defined as the location where the device is observed most frequently, provided it is observed for a minimum of 60 distinct hours and on at least 14 unique days. The work location is the second most frequently observed location, provided it is observed for a minimum of 60 distinct hours spread across at least two different weeks. Note that hybrid work—working a few days a week onsite and the rest of the time at home or in another location—should be captured along with full-time onsite work provided the person works a sufficient number of onsite days. We identify both the home and the work location at geohash level 8, which corresponds to an area of no more than 38.2 meters by 19.0 meters.[3] Additional steps were taken to ensure that the home and work locations do not overlap. Both the home and the work location then were assigned to a Census tract. Census tracts are geographic areas that generally include between 1,200 and 8,000 residents, with an optimum size of about 4,000 residents, and can be thought of as a neighborhood.

Using the sample of devices for which we were able to identify both a home and a work location in February 2020, we search for records for those same devices in May 2020, August 2020, November 2020 and March-April 2021.[4] In each of those months, we first attempt to identify a home location. Then, for devices observed a sufficient number of times to identify a home location, we attempt to identify a work location. To set a baseline for interpreting the results for the sample of workers who were working onsite in February 2020, we have carried out a similar analysis based on the sample of people who were working onsite in February 2019, searching for records for the same devices in May 2019 and August 2019.

---

[3] The listed dimensions are an upper bound, representing the size of a geohash level 8 at the equator.

[4] We opted not to use February 2021 as the final month for our panel because we wanted to avoid the large COVID spike that was in progress during that month. May 2021 would have been a natural final month, but Apple's introduction of new privacy protection measures in late April 2021 would have adversely impacted our ability to track the devices in our sample. The month-long period from mid-March to mid-April 2021 is the latest for which we could observe an adequately large number of devices.

In February 2020, we could identify a home location for about 11.7 million devices. Of these, the 4.2 million that also had an identifiable onsite work location constitute the initial sample for our main analysis. By March-April 2021, we were able to identify home locations for about 1.0 million devices or about 23 percent of our original sample. The usable 2019 sample is noticeably smaller. In February 2019, we could identify a home location for about 9.4 million devices of which about 2.1 million had an identifiable onsite work location.[5]

One impediment to identifying work locations in the months following February 2020 is that the average number of observations per device drops off substantially. This likely reflects people having become less mobile during the pandemic and thus making less use of some of the apps that generate location data. In May 2020, August 2020, November 2020 and March/April 2021, the sample consists of devices with a home and work location in February 2020 and a home location in the later month. The average number of observations per device in the March-April 2021 sample was similar in that period to what it had been for the same devices in February 2020. In the May 2020, August 2020 and November 2020 samples, however, the average number of observations in the end month was much lower than for the same sample of devices in February 2020. For this reason, we lower the thresholds used to define the work location in the later months. For example, if we had only 300 observed hours in a later month for a device for which there had been 600 observed hours in February 2020, the hours threshold for identifying the device's work location in the end month also was cut in half. [6] Although the average number of observed hours in the end month for the May 2019 and August 2019 samples

[5] See Appendix Table B1 for descriptive statistics on the raw number of devices available for different periods.

[6] Couture et al. (2022) note a similar decline in the number of observations in their mobile device data. They address this issue by extending the observation window used to construct their outcome measures. See the appendix for more details on the modifications to the hours thresholds we used to identify work locations.

is more similar to the average number in February 2019, for consistency we applied the same procedures to those samples.[7]

In addition to the decline in sample size implied by sample attrition, the rate of attrition we observe in our main sample varies across geography. We construct separate weights for the May 2020, August 2020, November 2020 and March-April 2021 samples to adjust for this differential attrition. To construct these weights, we first regress a dummy variable that equals one if a device for which we observed a home and work location in February 2020 continues to have an identifiable home location in the later month on block-group characteristics of the device's February 2020 residence location. These block-group characteristics are measured using data from the 2015-2019 American Community Survey (ACS).[8] Then, we compute the device's attrition weight as the inverse of its predicted continuation probability based on this regression. We use the same procedure to construct attrition weights for the May 2019 and August 2019 samples based on the residence location as of February 2019.

We also would like our sample to match the actual pattern of workers' residences and job locations. Starting with device counts that reflect the attrition weighting, we use an iterative proportional fitting (IPF) algorithm to reweight the data. For each of the samples for later months based on the set of devices with a February 2020 work location, the weighted pattern of February 2020 joint home and work locations matches the county-level information contained in the 2019 LEHD Origin-Destination Employment Statistics (LODES) data set.[9] The data appendix

[7] See Appendix Table B2 for descriptive statistics on mean observed hours and observed hours at the 10th, 50th, and 90th tract percentiles.

[8] These characteristics include the shares of block-group residents age 25-54, age 65 and older, White non-Hispanic, and having a college education, and the logarithm of average household income in the block group.

[9] One limitation of this weighting scheme is that the LODES data do not include self-employed individuals. To the extent that the prevalence of self-employment is similar across counties and cross-county commuting patterns are

provides details of this procedure. All of the within-tract measures based on the device data used in our analyses are constructed using a composite weight that is the product of the attrition weight and the weight derived from this IPF procedure. We use the same procedures to reweight the data based on the samples created from the set of devices that had both a home and work location in February 2019.

In addition to producing descriptive statistics based on the device-level onsite work measures, we are interested in explaining the variation in the prevalence of OSW across the tracts where people live. To do this, we have constructed variables that capture community characteristics that have been found in other research to be associated with social distancing and/or the prevalence of remote work. Several of our explanatory variables come from the 2015-2019 ACS. In addition to the variables used for our attrition analysis (now measured at the tract level rather than the block group level), these include the share of workers using public transportation to commute to work, the share of workers commuting more than 30 minutes to their jobs, and a set of 17 occupation dummies. We measure the shares of employment for each of 20 two-digit NAICS industries among workers living in the Census tract using 2019 LODES data. The Department of Agriculture is our source for a dummy variable indicating whether a tract is rural.[10] We use Donald Trump's share of the 2016 presidential vote to measure political preferences among tract residents. The source for this measure is the 2016 Precinct-Level

similar for the self-employed and wage-and-salary workforces, however, all of the estimated weights will be affected in the same proportion and our findings should be unaltered. Our work assignment algorithm will not count workers who drive for work (e.g., truck drivers) or travel from place to place for work (e.g., utilities meter readers) as onsite workers, but occupations where mobile work is the norm account for only a small percentage of total employment (Pan et al. 2023).

[10] We also experimented with measures of population density as opposed to a simple rural dummy variable, but found they added little to the explanatory power of our models.

Election Results database created by the Voting and Elections Science Team, which reports the Trump vote share at the precinct level; we assign devices to a precinct based on their home location and then take the weighted average of the Trump vote share across devices to create a tract-level measure. Goolsbee et al. (2020) is our source for information on COVID restrictions at the state and (if applicable) county level. We compute the share of days in the month of May 2020 during which a tract was under either state or county lockdown restrictions; tracts in locations where these restrictions were lifted later have higher values for this variable. Finally, our measure of COVID's impact is deaths per 100,000 people through May 31, 2020 (at the county level), as reported by the New York Times. The descriptive statistics and regression analyses based on our tract-level measures are weighted using tract-level employment as measured in the ACS. For comparison purposes, we use the same explanatory variables in our analyses of the variation in OSW in May 2019 and August 2019 for the February 2019 sample and of the *change* in the prevalence of OSW from May 2019 to May 2020 and from August 2019 to August 2020.

As mentioned, most of our analysis is based on a sample of tracts with a minimum of 10 devices for which we could identify a home location in each of the months we examine. The reasoning behind this restriction is that the OSW percentage for tracts with too few devices would be measured with considerable noise. We also have produced results for samples that are both more inclusive (including all tracts with one or more devices in every month) and less inclusive (requiring a minimum of 20 or 30 devices in every month). Although the variation in OSW percentage falls slightly with the stringency of the inclusion threshold, our main results are not sensitive to the threshold choice.

Another potential measurement issue is that some device owners we identify as traveling to work could be attending school rather than working. Using information from SafeGraph on the locations of points of interest (POI), we have repeated all of our main analyses dropping places identified as work locations that are within 100 meters of a school. This also has very little effect on our findings.

IV. Basic Facts

We begin our analysis by reporting some basic facts about the distribution of OSW and how it has changed. Figure 1 shows the distribution of the share of February 2020 onsite workers who were engaged in OSW in May 2020. To make the figure more readable, we have constructed these estimates at the county level rather than the tract level. The underlying numbers are employment-weighted averages of the OSW share across Census tracts within each county, calculated with each device assigned to its February 2020 residential location. To produce the best possible *county-level* estimates, rather than exclude the data for tracts with fewer than ten devices in one or more of our analysis months, we use data for all tracts that had at least one assigned device in each of those months. Darker shading in the figure is associated with lower OSW rates. Some 40.5 percent of counties, accounting for 82.4 percent of employment, have OSW shares in May 2020 for those who had been working onsite in February 2020 that are below 60 percent. Further analysis of the underlying data shows that average OSW rates are highest in counties that are not part of an urban area and, among urban areas, fall monotonically with CBSA size. The OSW shares are noticeably lower in the Northeast and along the West coast than elsewhere in the country. The shading for New York and California, for example, is distinctly darker than the shading for Florida and Texas; the lightest shading can be seen in the Midwest and Deep South.

This contrasts sharply with the pattern in Figure 2, which displays the distribution of the share of February 2019 onsite workers who were doing OSW in May 2019. More than 90 percent of counties, accounting for more than 99 percent of employment, had May 2019 OSW rates for those who had been working onsite in February 2019 of at least 70 percent. Indeed, more than 60 percent of counties accounting for about two-thirds of employment had OSW rates among those who had been working onsite in February 2019 of 80 percent or more.

Even in normal times, there is turnover in the labor market, as people move in and out of employment. In addition, people may move between working onsite at one primary location and other work arrangements. This suggests that, rather than focusing just on the OSW percentage among workers who previously had been working onsite, we also should be interested in the *changes* in this percentage associated with the pandemic. Using the same underlying data as were used to create the previous two figures, Figure 3 displays the change in the OSW share between May 2019 and May 2020 among those with identified onsite work locations in February of the corresponding year. Most counties experienced a decline in OSW persistence over this twelve-month period and the regional patterns are broadly similar to those in Figure 1, with the parts of the country having the smallest OSW shares in May 2020 also tending to be places where the OSW share fell the most between May 2019 and May 2020.

To further illustrate the very different early impact of the pandemic on onsite work in different parts of the country, Figure 4 shows the change in the OSW percentages between May 2019 (for the February 2019 OSW sample) and May 2020 (for the February 2020 OSW sample) for four cities. These are Houston and Miami, both large cities with relatively small changes in the OSW percentage, and New York and San Francisco, two large cities with relatively large OSW percentage changes. Only one of the nine counties in the Houston-The Woodlands-Sugar Land, TX Core-Based Statistical Area (CBSA) had a decline in OSW among workers who had previously been onsite between May 2019 and May 2020 of more than 30 percent. The Miami-Fort Lauderdale-Lucie, FL CBSA has fewer counties but only one of the three had a decline in this OSW rate of more than 30 percent. In contrast, in the New York-Newark-Jersey City, NY-NJ-PA CBSA, all but one county exhibited a decline in OSW for previously onsite workers of more than 30 percent and 14 out of 23 counties exhibited a decline of more than 40 percent. All five counties in the San Francisco-Oakland-Berkeley, CA CBSA experienced OSW declines in excess of 30 percent and two of the five counties experienced declines in excess of 40 percent.

These county-level displays mask significant tract-level variation. In addition, we are interested in how OSW evolved in later months. Figure 5 summarizes the cross-tract variation in OSW percentages in the form of a histogram of the tract-level OSW for May 2020 together with histograms for August 2020, November 2020 and March-April 2021. As in the earlier figures, these percentages are the shares of individuals who were working onsite in February 2020 who also had an identifiable onsite work location in the indicated month. Because we now are looking at tract-level data, the sample for this figure is restricted to tracts with 10 or more devices in every month. Table 1 reports summary statistics from the Figure 5 histograms.

Among February 2020 onsite workers, the mean tract-level OSW share in May 2020, constructed based on person-weighted OSW data by tract of residence using our 10-plus-device sample, is 53 percent. At the 90th percentile, 67 percent of February onsite workers were working onsite in May 2020; at the 10th percentile, only 38 percent were doing so, a difference of 29 percentage points. The mean OSW share rises after May 2020 but is still only 67 percent by

March-April 2021. The dispersion in the OSW share widens slightly over time; as of March-April 2021, the difference between the OSW percentage at the 90th percentile (83 percent) and that at the 10th percentile (49 percent) was 34 percentage points.

For comparison purposes, Figure 6 shows the analogous tract-level histogram of OSW for May 2019, along with the histogram for May 2020 and a histogram for the difference in OSW in the two months. The May 2019 histogram is for individuals who were working onsite in February 2019 and the May 2020 histogram is for individuals who were working onsite in February 2020. Of those who were working onsite in February 2019, 83 percent were still doing so in May 2019. Although normal labor market dynamics mean this percentage is not 100 percent, it is still considerably higher than the 53 percent of people working onsite in February 2020 who were still doing so in May 2020.

Figure 7 further explores the distribution of OSW in later months among those who were working onsite in February 2020. The histograms in panels A through E of the figure show how this varies across tracts that differ with respect to the share of their employment in Food and Accommodations; share of employment in Finance and Insurance; share of the population with a college degree; mean household income; and Donald Trump's share of votes in the 2016 presidential election. Each figure shows distributions for tracts grouped by quartile of the relevant characteristics. Some interesting patterns emerge. Despite the sharp initial downturn in employment in the sector, workers in tracts where a larger share of February 2020 employment was in Food and Accommodations were more likely to be doing onsite work in later months. The opposite is true in tracts where a larger share of February 2020 employment was in Finance and Insurance. The share of February 2020 onsite workers who were working onsite in later months also is lower in tracts with a higher fraction of workers who were college educated or had higher average household income. Rates of onsite work are higher in tracts where Donald Trump received a larger share of votes in the 2016 presidential election. Even among tracts in a given quartile of the distribution for a particular characteristic variable, however, there is considerable cross-tract variation in the OSW percentage.

As a check on these numbers before we proceed to further analysis of the spatial variation in OSW, it is useful to compare our estimates of onsite work to estimates from the Real-time Population Survey (RPS) described earlier in the paper. Among all of the sources of information on how the prevalence of onsite work changed following the onset of the pandemic, the RPS provides the estimates that conceptually are most comparable to our estimates. The RPS asks respondents about OSW activity in a given month and also about their OSW work activity in February 2020. This means that the RPS data can be used to construct estimates of the share of people working onsite in February 2020 who also were working onsite in various later months. For this purpose, we define OSW in the RPS data as working onsite three days per week, a threshold that we believe roughly approximates our threshold based on observed device hours for OSW in the MTI/CATT based tabulations.

Table 2 compares our estimates of the share of February 2020 onsite workers who were working onsite in May 2020, August 2020, November 2020 and March/April 2021 with those based on the RPS data. The two sets of estimates are very similar. The OSW share estimated in the RPS for November 2020 is somewhat higher than our estimate, but by March/April 2021 the RPS estimate has fallen while our estimate has risen, bringing the two estimates back into closer alignment. The similarity of the two sets of estimates bolsters our confidence in our measurement approach. We turn next to further analysis of the spatial variation in OSW.[11]

V. Accounting for Spatial Variation in Onsite Work

We now turn to a more formal accounting of the factors that underlie the enormous and persistent variation in OSW from May 2020 to March-April 2021. To accomplish this, we fit a series of regressions that relate the OSW percentage for previously onsite workers at different points in time or the *change* in this percentage from one year to the next to observable characteristics of the tracts. Then, we consider the explanatory power of these regressions for understanding the patterns we observe.

A. OSW Regression Specification

We begin our accounting exercise by estimating the following tract-level regression for each of our "snapshot" months:

$$OSW_{it} = X_i'\beta + \varepsilon_{it} \quad (1)$$

In this regression, OSW_{it} is the share of previously onsite workers in tract *i* working onsite in month t and X_i is a set of tract characteristics. As described above, we measure OSW for the months May 2020, August 2020, November 2020 and March-April 2021 for the individuals who

[11] The RPS data also permit calculations at the Census Division and state levels. The RPS sample size is relatively small, ranging from 2,245 to 4,287 across the four months for which we report estimates, meaning that the subnational RPS estimates should be viewed with considerable caution. With that caveat, it also is reassuring that our estimates and the RPS estimates are reasonably highly correlated across both Census Divisions and states. For example, in March/April 2021, the Pearson correlation between the comparable OSW estimates in the RPS and MTI/CATT Lab data is 0.79 at the Census Division level and 0.55 at the state level.

were working onsite in February 2020. For May and August 2019, we measure OSW for individuals who were working onsite in February 2019. All of the regressions are weighted by tract employment.

The covariates included in the regressions are basic demographics plus other tract characteristics that previous research has suggested may be associated with onsite work. These include the share of the population 25-64; the share of the population age 65 and older; the share of the population that is White, non-Hispanic; the share of the population age 25 and older with a college education; the mean of log(household income); the shares of the workforce commuting via public transportation and with one-way commutes of 30 minutes or more; whether the tract is rural; the share of the presidential vote for Donald Trump in 2016; the fraction of days in May 2020 for which state or (if applicable) county COVID lockdown restrictions were in place; and cumulative COVID deaths per 100,000 people in the population through May 2020. The regressions also include variables capturing the shares of workers in the tract who work in different industries and occupations.

B. OSW Regression Results

The estimated coefficients for the covariates other than those capturing industry and occupation mix are reported in Table 3. There is a systematic relationship between the likelihood that a person working onsite in February 2020 was still working onsite in the later months of 2020 and many of these observable covariates. OSW is less likely in tracts with a larger share of residents who are college graduates, an effect that persists over time. OSW initially is less likely in higher-income tracts, though this relationship becomes weaker in successive months. Reliance on public transportation and long commutes also have the expected negative effect on the likelihood of OSW. Residents of rural tracts and tracts where a higher share of 2016 votes were for Trump are more likely to be working onsite in later months. COVID restrictions and COVID deaths have the expected negative effect on onsite work in May 2020; interestingly, the former persist even after the early lockdowns were lifted.

The pattern of coefficients on the industry and occupation mix variables, shown in Appendix Tables B3 and B4, also are consistent with what one might have expected. The omitted industry group in these regressions is Finance and Insurance and the omitted occupation group is Computer, Engineering and Science, both categories where work is especially amenable to being performed remotely. Relative to these groups, most industries and occupations have higher rates of OSW in later months. For example, the estimated May 2020 coefficient on the share of workers in Retail Trade is 0.686, indicating that if the share of workers in a tract who work in Retail Trade is 10 percentage points higher, the share of onsite workers is about 7 percentage points higher. That differential remains high through March-April 2021. Similarly, although occupational effects are not as large as the industry effects, having a higher share of workers in Installation, Maintenance and Repair or Production occupations where an in-person presence is necessary is associated with a higher rate of OSW in May 2020 that continues through March-April 2021.

Table 3 also reports coefficients for similar regressions where the dependent variables are working onsite in May or August 2019, conditional on working onsite in February 2019. In contrast to the results for 2020, the covariates account for very little of the variation in onsite work across tracts—just 13 percent in May 2019 and 9 percent in August 2019. In addition, the pattern of the coefficient estimates is quite different. For example, whereas the May 2020 coefficient on mean log (household income) in the tract was negative and statistically significant, the May 2019 coefficient on that variable is positive and statistically significant.

One way to frame these differences is to think of the May 2019 coefficients as reflecting patterns during a normal period and the May 2020 coefficients as reflecting patterns during the COVID period. The difference between the two estimated coefficients thus may be a better estimate of the COVID impact than the coefficients from the May 2020 regressions. In the final two columns of Table 3, we report estimates of regressions in which the dependent variable is the change in the OSW rate for previously-OSW workers between May 2019 and May 2020 or between August 2019 and August 2020. Many of the same factors that helped to explain differences in the level of OSW in May and August 2020 also are important in these change regressions, but there are some differences.

In the change regressions, the effect of log(household income) is larger than in the level regressions, reflecting that income had a positive effect in 2020 but a negative effect in 2019. Members of high-income households may have more stable jobs during normal times and thus be more likely to persist in onsite work, but during COVID they may have been more able to work remotely and thus not to be measured as having an onsite job. A similar comment would apply to the share of workers in the tract who commute more than 30 minutes to work (negative coefficients in the 2020 level regressions, positive coefficient in the 2019 level regressions).

In other cases, the COVID effect measured in the May change regression is smaller than in the May 2020 regression. The share of votes for Trump in the 2016 election in the tract, for example, is sizeable and positive both in the 2020 level regressions and the 2019 level regressions. Taken on their own, the 2020 coefficient would seem to imply that political attitudes were an important source of variation in the COVID effect on onsite work. The fact that the coefficient on the 2016 Trump vote share is smaller in the change regressions, however, suggests that the Trump 2016 vote at least partially may be standing in for something else.

C. Decomposing the Variance in OSW

Although the coefficients estimates reported in Table 3 provide insight into how various covariates are correlated with the prevalence of onsite work in a tract among people who had been onsite workers prior to the pandemic's onset, these do not translate in a simple fashion into the share of the variation in the onsite work percentages that each explains. To provide insights into the quantitative importance of the relationships shown in Table 3, we have carried out a variance decomposition that allows us to say how much of the variation each covariate or (in the case of industry and occupation) set of covariates explains. The methodology we use to do this is as described in Hottman et al. (2016) and Eslava, Haltiwanger, and Urdaneta (forthcoming). This decomposition methodology assigns to each covariate the combination of its direct contribution to the variance along with terms related to its covariance with each of the other covariates.[12] By construction, this method yields a decomposition in which the terms, including the residual, sum to one. Table 4 reports the results of this decomposition for the eight models shown in Table 3. In May 2020, the largest contributor to the 57.8 percent of the variance accounted for by the covariates is industry mix, which accounts for 19.5 percent. Occupation mix accounts for 8.2

[12] Appendix C provides details regarding the variance decomposition methodology. It is possible for the covariance terms in the decomposition to more than offset the own variance contribution so that a covariate has a negative contribution to explaining the overall variance in the dependent variable. This is rare but holds, for example, for the share of the population aged 25 to 64 in the models for May 2020, August 2020, November 2020 and March-April 2021. In each of these models, the age 25-64 covariate has an estimated positive coefficient that yields a positive but modest own variance contribution. This covariate, however, is positively correlated with the log of median household income, a variable that has a large negative coefficient in each of the same models. The opposite signs of the coefficient estimates combined with the positive correlation between the two covariates yields a negative covariance contribution that outweighs the positive own variance contribution for the age 25-64 share, yielding a small negative overall contribution.

percent. Other large contributors are log(mean household income) at 8.9 percent and Trump voters at 7.4 percent. Additional covariates with greater than 3 percent contributions include the share of residents who are college graduates, state COVID restrictions and COVID deaths. Other demographic characteristics, the commuting variables, the rural indicator and county COVID restrictions contribute little after taking into account the effects of the other covariates (including the parceling out of covariance effects on an equal basis). The sum of the effects of the included covariates equals 100 times the R-squared in the corresponding Table 3 regression and the residual variance shown near the bottom of the column is just 100 minus the sum of the covariate effects.

Consistent with the R-squareds reported in Table 3, the contribution of the covariates to explaining the variance in OSW falls in later months. Industry mix, occupation mix and the Trump share of the 2016 election vote continue to be important through March-April 2021.Other covariates generally become less important after May 2020. The covariates explain relatively little of the variance in OSW in either May 2019 or August 2019.

Many of the same covariates that account for significant portions of the variance in OSW in the 2020 regressions also are important in explaining the variance of the 2020 minus 2019 first differences. Focusing on the May 2020 minus May 2019 column, likely for the reasons discussed above, log(mean household income) actually explains a larger share of the first difference variance than of the May 2020 variance. Industry and occupation mix also are important, each explaining a substantial fraction of the first-difference variance. Reflecting the fact that it is positively related to the OSW share in both May 2020 and May 2019 however, the 2016 Trump vote share explains little of the variance in the first difference. The same patterns broadly hold for the August 2020 minus August 2019 OSW change, though the explained portion of the first-difference variance drops from 44.6 percent for the May change to 24.3 percent for the August change.

The estimated covariate effects as reported in Table 3 do a good job of accounting for the broad regional variation in OSW shown in the maps provided earlier. Consider, for example, the higher rate of OSW indicated by the lighter shading of the May 2020 map in Texas and Florida as compared to New York and California. As can be seen in the four sets of rows in the panel at the top of Table 5, on an employment-weighted basis, the mean OSW in May 2020 was 54.7 percent in Texas and 54.5 percent in Florida, compared to 46.7 percent in New York and 47.9 percent in California. This aligns well with the predictions from our tract-level model. Those estimated coefficients imply an OSW rate of 55.7 percent in Texas, 54.7 percent in Florida, 47.7 percent in New York, and 47.5 percent in California. The regressions also do a good job of capturing the OSW levels and the changes in OSW levels between 2019 and 2020 in these states.

By way of further illustration, we also have carried out a similar exercise comparing the Houston-The Woodlands-Sugar Land, TX CBSA, the Miami-Fort Lauderdale-Lucie, FL CBSA, the New York-Newark-Jersey City, NY-NJ-PA CBSA, and the San Francisco-Oakland-Berkeley, CA CBSA. As was apparent from Figure 4, the May 2020 OSW share was much higher in Houston and Miami than in either New York or San Francisco. The difference between Houston and New York shrinks in later months but remains sizable; the gap between Houston and San Francisco actually grows. Similar remarks apply to the Miami versus New York and Miami versus San Francisco differences.[13]

[13] Although both Houston and Miami have OSW means that are close to the national average in Table 1, large cities generally have lower OSW rates. Among the 25 largest CBSAs, all with 2020 populations of 2 ½ million or more, the six with the highest OSW rates in our data are Tampa, Houston, Riverside, San Antonio, Charlotte and Miami. Among the same 25 CBSAs, Washington, DC, San Francisco, and New York have the lowest May 2020 OSW rates.

As with the differences we observe across sites, these differences largely are accounted for by differences in the characteristics of the two cities. Looking at the bottom four panels of Table 5, the actual May 2020 OSW share in Houston was 54.0 percent; the share predicted from the coefficients estimated for the national model was 54.4 percent. For Miami, the corresponding figures were 51.1 percent (actual) and 51.8 percent (predicted); for New York, they were 40.1 percent (actual) and 40.4 percent (predicted); and for San Francisco, they were 40.0 percent (actual) and 39.0 percent (predicted). The regression coefficients do a similarly good job of capturing the OSW levels in these cities in other months as well as the changes in OSW levels between May and August 2019 and the same months in 2020. [14]

These patterns reflect the very different characteristics of the four cities. For example, the share of college graduates is only about 43 percent in Houston and 42 percent in Miami but substantially larger in New York City (51 percent) and San Francisco (53 percent). Relatedly, mean household income is 34 percent higher in New York City than in Houston and 62 percent higher than in Miami; mean household income in San Francisco is 51 percent higher than in Houston and 83 percent higher than in Miami. Both New York City and San Francisco had state or county shutdowns in place for a larger part of May 2020 than either Houston or Miami.

Relative to overall employment, there are fewer jobs in the construction and transportation and warehousing sectors, but more jobs in information, and professional, scientific

[14] As a point of comparison, Hansen et al. (2023) find that New York and especially San Francisco are cities with relatively large shares of vacancy postings offering remote work in 2022 while Miami and Houston have much lower remote vacancy posting shares. The cross-city patterns in the vacancy posting data are broadly consistent with the pattern of onsite work we find in the MTI/CATT data. Hansen et al. (2023) also note their cross-city patterns are similar to the patterns of change in work-from-home (WFH) that emerge in American Community Survey (ACS) data between 2019 and 2021. Decker and Haltiwanger (2023) find a much weaker correlation between the ACS data and the Hansen et al. data at the county level. They suggest this may be because the ACS data on WFH are based on place of residence while the Hansen et al. (2023) data on vacancy postings are based on the location of the business posting the vacancy.

and technical services in New York City and San Francisco compared to Houston and Miami. Miami also stands out as having an especially high share of employees in the food and accommodations sector. These differences are important not only because OSW rates tend to be lower in the industries for which New York City and San Francisco have higher employment shares but also because of the interactions between these shares and the changes in the industry effects on OSW rates over time. More specifically, the changes in the industry effects we estimate in our OSW models contribute a 1.3 to 1.5 percentage point widening in the gap between the OSW rates in Houston and Miami compared to that in San Francisco, largely because information represents such a large share of San Francisco employment and the information sector coefficients in our OSW models imply that, all else the same, OSW has fallen in information relative to for other sectors. New York City and San Francisco also have smaller shares of employment in blue collar occupations than either Houston or Miami.

Much of the variation in OSW across cities is accounted for by observable factors across cities, but there is considerable variation across tracts within these cities. Table 6 compares the cross-tract variation in OSW in Houston and San Francisco, two cities with notably different mean levels of OSW. In both of these cities, the standard deviation across tracts is large and grew over time. The within-city dispersion implies that a tract one standard deviation above the mean in San Franciso in the spring of 2021 has a significantly higher share of residents working onsite (63.5 percent) than a tract one standard deviation below the mean in Houston (56.4 percent). Observable differences across tracts account for a substantial fraction of this variation. By the spring of 2021, the leading measurable factors contributing to the cross-tract variance are the industry and occupation mix of the resident workers along with the share of college educated residents and median income. While measurable factors play an important role, there is also substantial residual variation across tracts within both of these cities. Table 6 highlights the importance of cross-tract variation within larger geographic areas such as cities, a topic we explore further in the next section.

D. What Does a Tract-Level Analysis Add?

To further explore the role of neighborhood differences (i.e., tract differences) versus what can be accounted for by systematic differences across broader geographic areas (state and county), we have estimated the specifications summarized in Tables 3 and 4 with state and county fixed effects added to the models. Table 7 reports variance decompositions for the models with state effects added; variance decompositions for the models with county effects added are reported in Table 8. For comparison, we also estimated models with only state effects or only county effects.

On their own, as shown in Table 7, state fixed effects explain only 22.4 percent of the May 2020 variance in tract-level OSW, as compared to the 57.8 percent explained by our list of covariates. Adding state fixed effects to the covariate model raises the explained variance by only 1.7 percentage points. In the full model, 9.5 percent of the cross-tract variance—or about 15 percent of the explained variance—is absorbed by the state fixed effects, though some of this is due to the effects of state COVID lockdown differences that we could not control for separately once the state fixed effects were introduced. Similar comments apply to the variance decompositions for the later months. State fixed effects on their own explain substantially less of the variance in tract-level OSW shares than our covariates; add little explanatory power when added to models that include the covariates; and, in models that include both covariates and the state fixed effects, absorb only a modest portion of the cross-tract OSW variance.

As can be seen in Table 8, county fixed effects capture considerably more of the OSW variance than the state level effects but also add little to the models' overall explanatory power. On their own, county fixed effects account for 50.3 percent of the variation in the May 2020 tract-level OSW, compared to 57.8 percent in the model including only our covariates. Adding the county fixed effects to the model that includes the covariates raises the explained variance by just 6.0 percentage points. In the combined model, county fixed effects absorb 20.5 percent of the variance—about a third of the total explained variance—but this includes the effects of state and county COVID lockdowns and county-level COVID deaths that we could not control for separately once the county fixed effects were added. By March-April 2021, county fixed effects alone account for 35.0 percent of the variation, essentially the same as the covariates on their own. Adding the county fixed effects to the model with covariates raises the explained variance by only 6.4 percentage points, with the county effects accounting for only 15.0 percent of the overall variance or a bit more than a third of the explained variance. Again, however, part of the variance captured by the county effects otherwise would have been attributable to the effects of early state and county lockdowns and county COVID deaths that we had to drop from the covariate list when the county effects were added. The clear implication of these results is that within-county cross-tract variation in OSW is a more important part of the overall picture than the cross-county variation in OSW.

E. *Robustness Checks*

One of the key take-aways from our analysis is the existence of substantial cross tract variation in OSW even within cities and counties. In our baseline analysis, in an effort to restrict attention to tracts in which the OSW percentage is measured with reasonable accuracy, we have focused on tracts for which we observe a minimum of 10 devices in every period. Still, even with a minimum of 10 devices per tract, there will be some sampling error in our estimates. To assess the sensitivity of our estimates to sampling error, we have looked at how the mean OSW rate and the cross-tract OSW distribution vary across several different tract selection rules—including all tracts with one or more usable devices in every period, tracts with at least 10 usable devices in every period (the baseline specification), tracts with at least 20 usable devices in every period and tracts with at least 30 devices in every period. These estimates are shown in Table 9. Not surprisingly, the larger the minimum number of usable devices required for inclusion in the sample, the smaller the number of tracts. The mean OSW share is relatively insensitive to the choice of sample restriction. The gaps between the 90th percentile and 10th percentile OSW rates are smaller when the sample of tracts is restricted to those with a larger number of usable devices, but the declines are modest in size.[15] Moreover, the time series variation in these measures is broadly robust across the different samples.

We also report the R-squared from our baseline regression specification for each of these samples. Consistent with sampling error being smaller when the underlying sample is larger, the R-squared is higher when the tract inclusion criterion is more restrictive. In the May 2020 regressions, for example, it rises from 0.453 when we include all tracts with one or more devices to 0.578 in our baseline sample with a minimum of 10 devices per tract to 0.704 for the sample of tracts with a minimum of 30 devices. Although the level of the R-squareds varies across the samples, their behavior over time is similar. Appendix Table B5.A reports the estimated coefficients for the sample of all tracts with one or more usable devices; Appendix Table B5.B

[15] The standard deviations of the OSW rate across tracts also decline only slightly as more stringent sample inclusion criteria are imposed.

reports the estimated coefficients for the sample of tracts with 20 or more usable devices; and Appendix Table B5.C reports the estimated coefficients for the sample of tracts with 30 or more usable devices. The corresponding variance decompositions are shown in Appendix Tables B6.A, B6.B and B6.C. The contributions of specific covariates to explaining the cross-tract variation in OSW are very robust across samples. For example, in all of the samples including the sample restricted to tracts with thirty or more usable devices, industry mix effects account for about a third of the explained May 2020 cross tract variation.[16]

Another possible concern is that the identification of work locations may be confounded with commuting to schools. In Appendix Table B7, we report estimates with the OSW percentages recalculated to exclude all identified work locations within 100 meters of a site identified in the SafeGraph data as a school. All of the estimated coefficients are very similar to our baseline estimates, as are the corresponding variance decompositions shown in Appendix Table B8.

VI. Conclusion

Over the course of the pandemic and its aftermath, Americans in different communities have had – and continue to have – very different experiences in terms of work from home (WFH) or, conversely, onsite work (OSW). Other research has found that cities such as New York and San Francisco exhibited larger and more persistent declines in OSW than cities such as Houston and Miami, but the existing literature has not systematically explored the variation in OSW at a more disaggregated level. We establish that, both within cities with relatively higher

[16] Regression results for the different samples are available from the authors on request.

rates of post-pandemic OSW and within cities with notably lower rates of post-pandemic OSW, there is substantial cross-neighborhood variation in OSW. More generally, across the United States, even after controlling for county of residence, there is substantial cross-tract variation in OSW.

The variation in OSW across neighborhoods is large in magnitude. For example, across the San Francisco metropolitan area, in the average tract, 49.4 percent of people who worked onsite in February 2020 also were working onsite in the spring of 2021, but this percentage varies considerably across tracts. In San Francisco tracts one standard deviation above the mean, 63.5 percent of previously onsite workers were working onsite in the spring of 2021; in tracts one standard deviation below the mean, this was true of just 35.3 percent of previously onsite workers. In the average tract in the Houston metropolitan area, 66.5 percent of people who worked onsite in February 2020 were doing so in the spring of 2021. In tracts one standard deviation below the mean, however, this was true of just 56.4 percent of previously employed workers, a figure that is substantially below the estimate for San Francisco tracts one standard deviation above the San Francisco mean. Put differently, there is substantial overlap in the tract-level distributions of OSW rates between higher-OSW cities such as Houston and Miami and lower-OSW cities such as New York and San Francisco. We observe substantial cross tract variation in OSW more generally, with much of the variation occurring within states and even within counties.

We identify a number of factors that help to explain this significant cross-tract variation. The most important contributors are the industry and occupation mix of the workers who live in the tract. Median household income and the share of the population with a college education play strong supporting roles. Factors such as COVID restrictions, COVID related deaths, commuting time, and the fraction of workers who commute by public transportation contributed more early in the pandemic but had become less important by the spring of 2021. The 2016 share of votes for Trump also helps to explain the cross-tract variation in the level of OSW, though it is substantially less important in first-difference specifications that relate the change in OSW between 2019 and 2020 to tract-level characteristics. Although we are able to account for much of the cross-tract variation in OSW, a significant fraction of that variation cannot be explained even by the rich set of tract characteristics that we consider. The variables in our model do a better job early in the pandemic, but by the spring of 2021, they account for only about a third of the cross-tract variation in onsite work.

These findings have important implications for analysts and policymakers. In neighborhoods with fewer residents who are working onsite, there will be more people at home during the day and, correspondingly, fewer people in the business locations they previously frequented. Among other potential effects, having more people at home during the day will affect the demand for local services and change the demands on the local transportation infrastructure. There is evidence that businesses are beginning to respond to these changes. For example, Decker and Haltiwanger (2023) find that the surge in business formation since the start of the pandemic exhibits spatial variation at the local level consistent with changing OSW/WFH patterns. Fully adapting to a situation in which there is persistent neighborhood-by-neighborhood variation in the share of workers who are at home during the day, however, is likely to take time.

References

Adrjan, Pawel, Gabriele Ciminelli, Alexandre Judes, Michael Koelle, Cyrille Schwellnus, and Tara Sinclair. 2021. "Will It Stay or Will It Go? Analysing Developments in Telework During COVID-19 Using Online Job Postings Data," OECD Productivity Working Papers, No. 30, OECD Publishing, Paris.

Aksoy, Cevat Giray, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Mathias Dolls and Pablo Zarate. 2022. "Working from Home Around the World," *Brookings Papers on Economic Activity* (Fall): 281–330.

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang. 2020. "Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic," *Journal of Public Economics*, 191, 104254.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2021. "Why Working from Home Will Stick," National Bureau of Economic Research, No. w28731.

Bick, Alexander, Adam Blandin, and Karel Mertens. 2023. "Work from Home Before and After the COVID-19 Outbreak," *American Economic Journal: Macroeconomics*, 15(4): 1-39.

Brynjolfsson, Erik, John J. Horton, Christos Makridis, Alexandre Mas, Adam Ozimek, Daniel Rock and Hong-Yi TuYe. 2023. "How Many Americans Work Remotely? A Survey of Surveys and Their Measurement Issues," National Bureau of Economic Research, No. w31193.

Couture, Victor, Jonathan I. Dingel, Allison Green, Jessie Handbury, and Kevin R. Williams. 2022. "JUE Insight: Measuring Movement and Social Contact with Smartphone Data: A Real-time Application to COVID-19," *Journal of Urban Economics*, 127, 103328.

Decker, Ryan, and John Haltiwanger. 2023. "Surging Business Formation in the Pandemic: Causes and Consequences?" *Brookings Papers on Economic Activity*, Fall (forthcoming).

Deole, Sumit S., Max Deter, and Yue Huang. 2023. "Home Sweet Home: Working from Home and Employee Performance During the COVID-19 Pandemic in the UK," *Labour Economics,* 80, 102295.

Dey, Matthew, Harley Frazis, David S. Piccone Jr, and Mark A. Loewenstein. 2021. "Teleworking and Lost Work during the Pandemic: New Evidence from the CPS," *Monthly Labor Review*, 144.

Dingel, Jonathan I., and Brent Neiman. 2020. "How Many Jobs Can Be Done at Home?," *Journal of Public Economics* 189, 104235.

Eslava, Marcela, Haltiwanger, John C., and Nicolas Urdaneta. 2023. "The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand and Wedges," *Review of Economic Studies*.

Goolsbee, Austan, Nicole Bei Luo, Roxanne Nesbitt, and Chad Syverson. 2020. "COVID-19 Lockdown Policies at the State and Local Level," University of Chicago, Becker Friedman Institute for Economics Working Paper 2020-116.

Hansen, Stephen, Peter John Lambert, Nicholas Bloom, Steven J. Davis, Raffaella Sadun, and Bledi Taska. 2023. "Remote Work Across Jobs, Companies, and Space." National Bureau of Economic Research, No. w31007.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity," *The Quarterly Journal of Economics*, 131(3): 1291-1364.

Jacobsen, Grant D. and Kathryn H. Jacobsen. 2020. "Statewide COVID-19 Stay-at-Home Orders and Population Mobility in the United States," *World Medical and Health Policy*, 12(4): 347-356.

Jay, Jonathan, Jacob Bor, Elaine O. Nsoesie, Sarah K. Lipson, David K. Jones, Sandro Galea, and Julia Raifman. 2020. "Neighbourhood Income and Physical Distancing During the COVID-19 Pandemic in the United States," *Nature Human Behavior*, 4: 1294–1302.

Mendolia, Silvia, Olena Stavrunova and Oleg Yerokhin. 2021. "Determinants of the Community Mobility during the COVID-19 Epidemic: The Role of Government Regulations and Information," *Journal of Economic Behavior and Organization*, 184: 199-231.

Pan, Yixuan, Qianqian Sun, Mofeng Yang, Aref Darzi, Guangchen Zhao, Aliakbar Kabiri, Chenfeng Xiong, and Lei Zhang. 2023. "Residency and Worker Status Identification Based on Mobile Device Location Data," *Transportation Research Part C: Emerging Technologies*, 146, 103956.

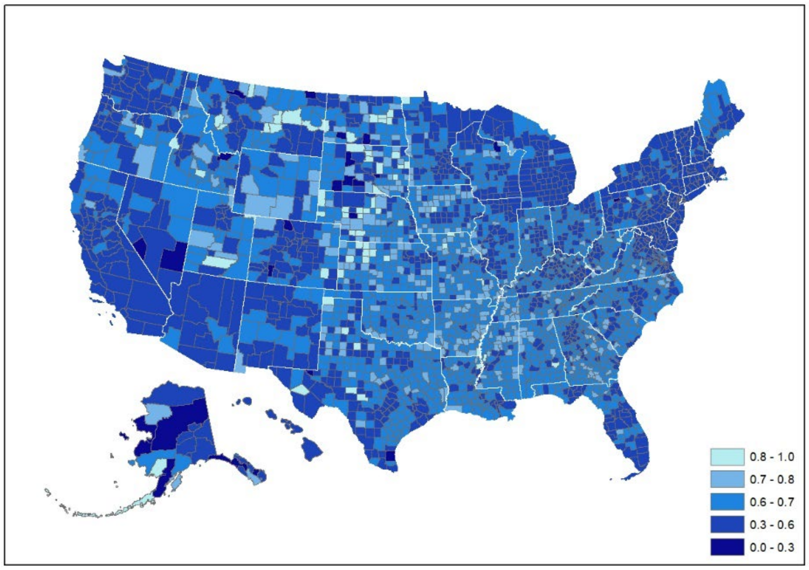
Rafiq, Rezwana, Michael G. McNally, Yusuf Sarwar Uddin, and Tanjeeb Ahmed. 2022. "Impact of Working from Home on Activity-Travel Behavior During the COVID-19 Pandemic: An Aggregate Structural Analysis." *Transportation Research Part A: Policy and Practice*, 159: 35-54.

Sehra, Shiv T., Michael George, Douglas J. Wiebe, Shelby Fundin, and Joshua F. Baker. 2020. "Cell Phone Activity in Categories of Places and Associations with Growth in Cases of COVID-19 in the US," *JAMA Internal Medicine*, 80(12):1614-1620.

Zhang, Lei, Aref Darzi, Sepehr Ghader, Michael L. Pack, Chenfeng Xiong, Mofeng Yang, Qianqian Sun, Aliakbar Kabiri, and Songhua Hu. "Interactive COVID-19 Mobility Impact and Social Distancing Analysis Platform." *Transportation Research Record*, 2677(4): 168-180.

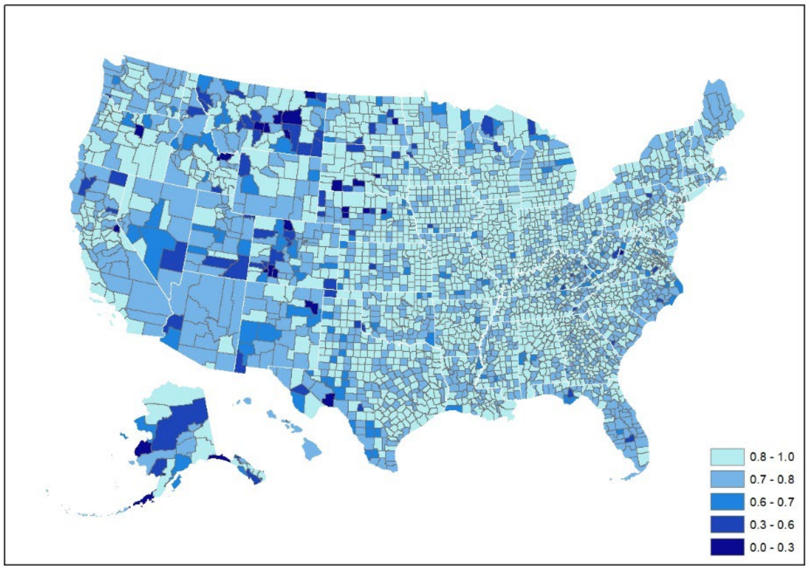
Zhang, L., A. Darzi, Y. Pan, M. Yang, Q. Sun, A. Kabiri, G. Zhao, and C. Xiong. 2022. "Next generation National Household Travel Survey National Origin Destination Data Passenger Origin-Destination Data Methodology Documentation," Unpublished.

Figure 1: Shares of Workers Who Were Onsite in February 2020
Who Also Were Onsite in May 2020



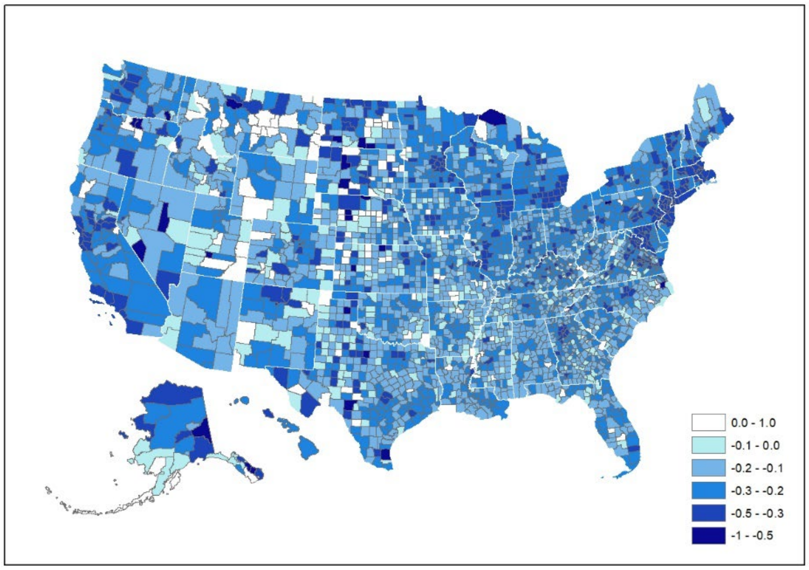
Source: Authors' tabulations from MTI/CATT Lab database.

Figure 2: Shares of Workers Who Were Onsite in February 2019
Who Also Were Onsite in May 2019



Source: Authors' tabulations from MTI/CATT Lab database.

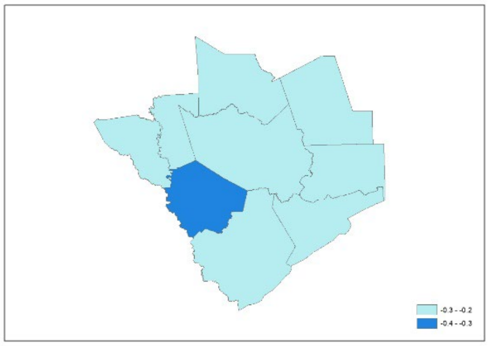
Figure 3: Difference in Share of Workers Persisting in Onsite Work from February, May 2020 minus May 2019



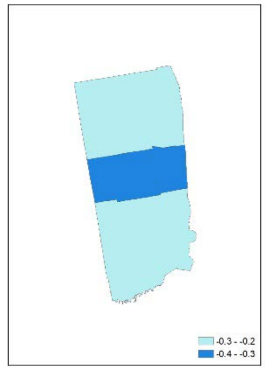
Source: Authors' tabulations from MTI/CATT Lab database.

Figure 4: Difference in Share of Workers Persisting in Onsite Work from February, May 2020 minus May 2019

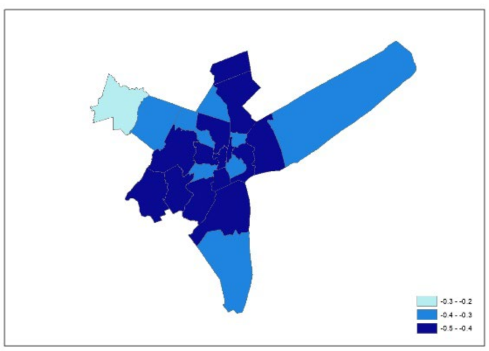
A. Houston



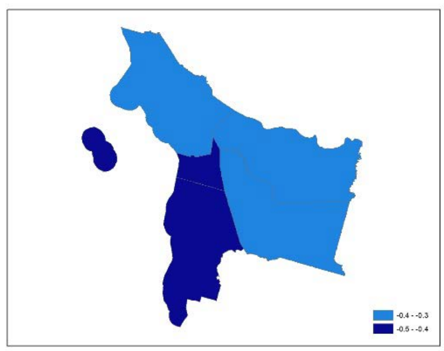
B. Miami



C. New York

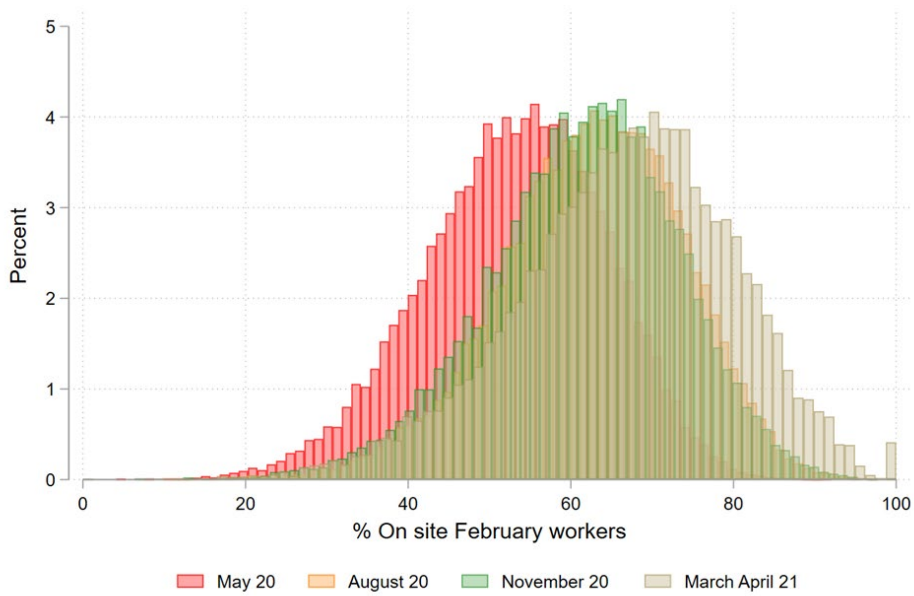


D. San Francisco



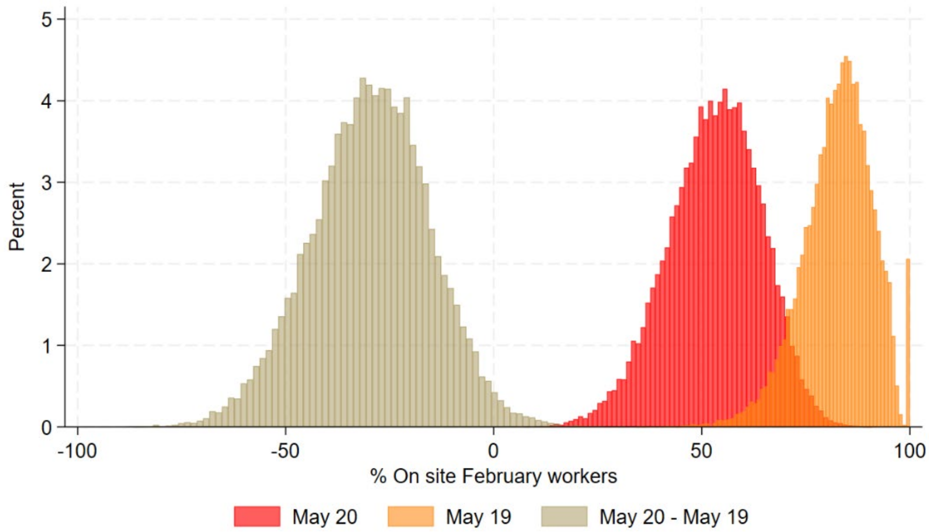
Source: Authors' tabulations from MTI/CATT Lab database.

Figure 5: Tract-Level Distribution of Onsite Work Percentages among February 2020 Onsite Workers, Selected Months, Sample of Tracts with 10 or More Devices



Source: Authors' tabulations from MTI/CATT Lab database.

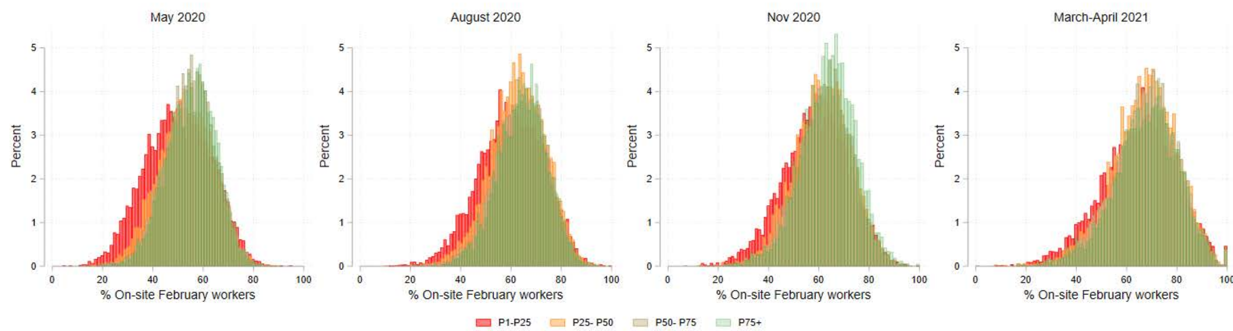
Figure 6: Tract-Level Distribution of Onsite Work among February 2020 and 2019 Onsite Workers in May 2020 and May 2019 and Difference across Two Years, Sample of Tracts with 10 or More Devices



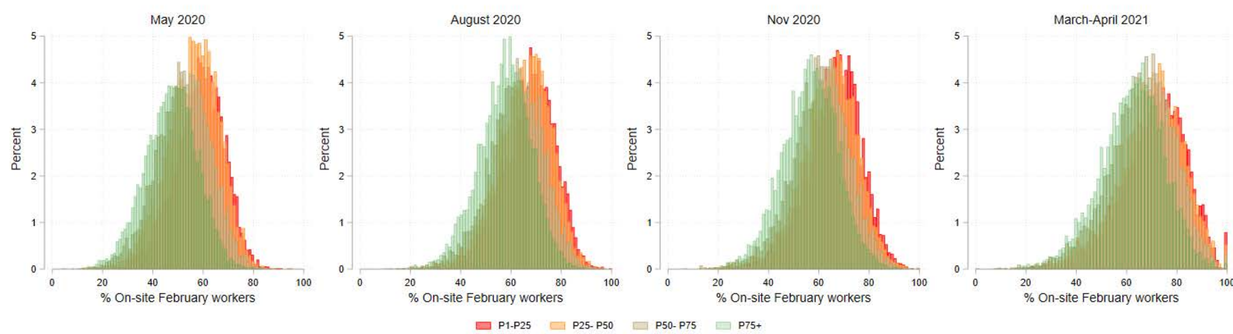
Source: Authors' tabulations from MTI/CATT Lab database.

Figure 7: Tract-Level Distribution of Onsite Work Percentages among February 2020 Onsite Workers by Quartiles of Selected Tract Characteristics, Selected Months, Sample of Tracts with 10 or More Devices

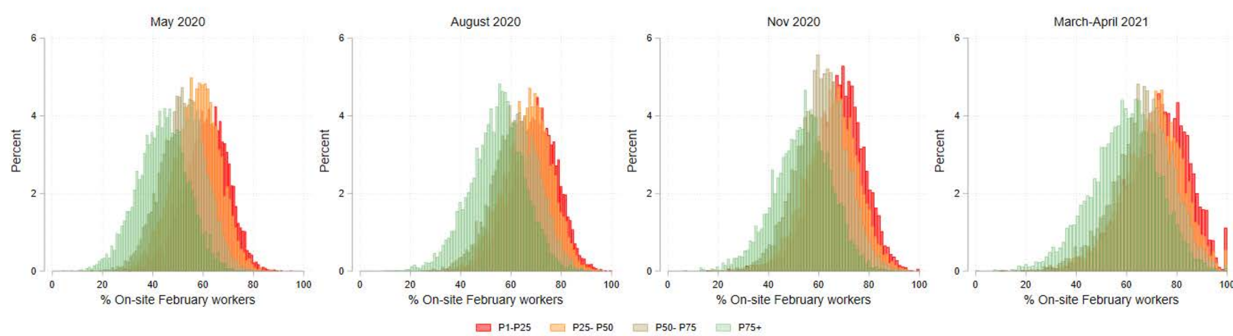
Panel A: Share of Workers in Accommodation and Food Services



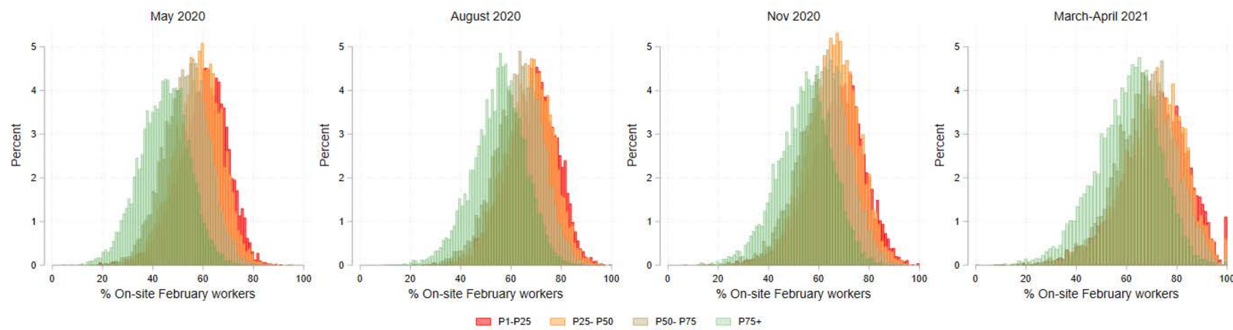
Panel B: Share of Workers in Finance and Insurance



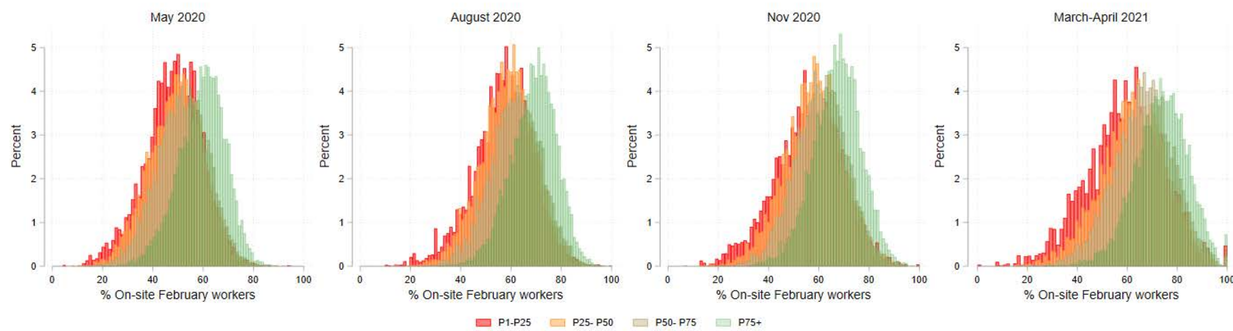
Panel C: Share of Residents Age 25 and Older with a College Degree



Panel D: Logarithm of Mean Household Income



Panel E: Share of Votes for Donald Trump in 2016 Presidential Election



Source: Authors' tabulations from MTI/CATT Lab database.

Table 1: Percent in On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019, Sample of Tracts with 10 or More Devices

	Among those working on-site in February 2020, percent OSW as of:				Among those working on-site in February 2019, percent OSW as of:	
	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019
Mean	53.0	62.6	61.4	66.7	82.6	78.4
p10	38.1	47.9	46.0	49.1	71.6	64.0
p50	53.6	63.2	62.1	67.6	83.4	79.5
p90	67.1	76.5	75.8	83.1	92.8	91.6

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 10 or more devices with a home location in every period. N=28,125 Census tracts.

Table 2: Percent in On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020, Comparison of MTI/CATT and Real-time Population Survey (RPS) Estimates

	Among those working on-site in February 2020, percent OSW as of:			
	May 2020	August 2020	November 2020	March-April 2021
MTI/CATT	53.0	62.6	61.4	66.7
RPS	52.9	60.5	66.1	64.5
Difference	0.1	2.1	-4.7	2.2

Source: Authors' calculations, MTI/CATT Lab mobile device location database and Real-time Population Survey (RPS) public domain database.
Note: MTI/CATT Lab sample restricted to tracts with 10 or more devices with a home location in every period. RPS sample size is 2,939 in May 2020, 4,021 in August 2020, 2,245 in November 2020, and 4,287 in March/April 2021.

Table 3: Factors Affecting Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019, Sample Restricted to Tracts with 10 or More Devices

		OSW February 2020 and:				OSW February 2019 and:			
Explanatory variables	Mean (standard deviation)	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:									
Age 25-64	52.8	0.137	0.088	0.082	0.099	0.100	0.079	0.037	0.008
	(5.8)	(0.009)	(0.010)	(0.012)	(0.014)	(0.010)	(0.013)	(0.014)	(0.016)
Age 65 plus	14.8	0.0101	-0.022	0.000	0.005	0.026	0.000	-0.016	-0.023
	(6.0)	(0.009)	(0.010)	(0.012)	(0.014)	(0.010)	(0.013)	(0.014)	(0.016)
White, non-Hispanic	66.0	-0.003	0.015	-0.022	0.020	-0.008	-0.019	0.005	0.034
	(25.2)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
College graduate	43.4	-0.038	-0.026	-0.036	-0.030	-0.029	-0.022	-0.010	-0.005
	(16.5)	(0.008)	(0.009)	(0.010)	(0.012)	(0.009)	(0.012)	(0.012)	(0.014)
ln(mean household income	4.3	-4.832	-2.990	-2.528	-1.289	4.666	1.511	-9.498	-4.501
	(0.4)	(0.228)	(0.254)	(0.285)	(0.340)	(0.247)	(0.325)	(0.335)	(0.399)
Share commute public trans.	14.2	-0.009	-0.020	-0.026	-0.027	-0.010	-0.012	0.001	-0.008
	(14.5)	(0.004)	(0.004)	(0.005)	(0.006)	(0.004)	(0.006)	(0.006)	(0.007)
Share commute 30+ mins.	38.6	-0.045	-0.045	-0.044	-0.036	0.015	0.033	-0.060	-0.079
	(15.7)	(0.004)	(0.004)	(0.005)	(0.006)	(0.004)	(0.005)	(0.005)	(0.007)
Rural yes/no	12.8	0.008	0.005	0.008	0.006	-0.002	0.001	0.010	0.004
	(33.4)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Share Trump vote in 2016	51.0	0.109	0.141	0.154	0.149	0.068	0.135	0.042	0.006
	(18.4)	(0.005)	(0.005)	(0.006)	(0.007)	(0.005)	(0.007)	(0.007)	(0.009)
May 2020 state lockdown	43.1	-0.038	-0.033	-0.032	-0.042	0.004	-0.021	-0.042	-0.012
	(42.4)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
May 2020 local lockdown	4.7	-0.022	-0.018	-0.013	-0.001	0.007	-0.003	-0.029	-0.015
	(17.4)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)
May 2020 cum COVID deaths	1.9	-0.475	-0.099	-0.052	-0.028	-0.043	-0.336	-0.432	0.237
	(3.0)	(0.018)	(0.020)	(0.022)	(0.026)	(0.019)	(0.025)	(0.026)	(0.031)
Industry dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	--	53.0	62.6	61.4	66.7	82.6	78.4	-29.6	-15.8
Dep. var. standard deviation	--	(11.2)	(11.2)	(11.7)	(13.4)	(8.5)	(10.9)	(14.4)	(14.7)
R-squared	--	0.578	0.479	0.396	0.343	0.129	0.0870	0.446	0.243

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 10 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,125 Census tracts.

Table 4: Percent of Variance in Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019 Explained by Various Factors, Sample Restricted to Tracts with 10 or More Devices

	OSW February 2020 and:				OSW February 2019 and:			
Explanatory variables	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:								
Age 25-64	-1.3	-0.9	-0.8	-0.7	0.7	0.1	-0.3	-0.1
Age 65 plus	0.1	-0.2	0.0	0.0	0.1	0.0	0.0	-0.1
White, non-Hispanic	-0.1	0.6	-0.6	0.8	-0.4	-0.5	0.0	0.3
College graduate	3.3	2.0	2.4	1.4	-0.7	0.0	0.6	0.2
ln(mean household income)	8.9	4.8	3.5	1.2	5.5	0.3	14.4	4.7
Share commute public trans.	0.4	0.9	1.1	0.9	0.1	0.2	0.0	0.1
Share commute 30+ mins.	1.5	1.6	1.3	0.8	0.2	0.1	1.6	1.7
Rural (yes/no)	0.8	0.5	0.6	0.4	0.0	0.0	0.6	0.2
Share Trump vote in 2016	7.4	9.8	9.2	8.4	2.6	5.0	1.2	0.1
May 2020 state lockdown	4.7	3.4	2.9	3.3	0.1	1.1	3.6	0.4
May 2020 local lockdown	0.3	0.2	0.1	0.0	0.0	0.0	0.3	0.1
May 2020 cum COVID deaths	4.1	0.6	0.2	0.1	-0.1	1.2	2.5	-0.3
Industry mix	19.5	16.5	14.9	13.0	3.3	1.1	12.5	9.9
Occupation mix	8.2	8.1	4.7	4.7	1.3	0.1	7.8	6.9
Residual	42.2	52.1	60.4	65.7	87.1	91.3	55.4	75.7
Dep. var. mean	53.0	62.6	61.4	66.7	82.6	78.4	-29.6	-15.8
Dep. var. standard deviation	(11.2)	(11.2)	(11.7)	(13.4)	(8.5)	(10.9)	(14.4)	(14.7)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 10 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,125 Census tracts.

Table 5: Actual versus Predicted On-Site Work Prevalence Among Individuals Working On-Site in February 2020 or February 2019, Selected Geographic Areas

	OSW February 2020 and:				OSW February 2019 and:			
Geographic area	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Texas								
Actual	54.7	62.0	62.5	66.7	81.0	77.9	-26.3	-15.9
Predicted	55.7	63.3	62.9	67.4	80.8	78.8	-25.1	-15.5
Florida								
Actual	54.5	62.3	62.0	66.6	78.5	76.0	-24.0	-13.7
Predicted	54.7	62.5	61.9	66.6	78.8	76.4	-24.2	-13.9
New York								
Actual	46.7	58.3	60.3	62.1	81.3	72.4	-34.6	-14.1
Predicted	47.7	59.3	58.3	63.6	82.3	74.5	-34.6	-15.2
California								
Actual	47.9	55.6	55.8	56.8	81.2	76.7	-33.3	-21.1
Predicted	47.5	55.5	55.3	57.7	80.9	75.7	-33.4	-20.2
Houston-The Woodlands-Sugarland TX								
Actual	54.0	60.1	62.7	66.6	81.3	78.0	-27.3	-18.0
Predicted	54.4	61.7	61.6	66.0	81.2	79.0	-26.7	-17.4
Miami-Fort Lauderdale-Lucie, FL								
Actual	51.4	59.7	60.2	65.4	81.8	78.2	-30.3	-18.5
Predicted	51.8	60.0	60.0	64.4	80.7	77.3	-29.0	-17.3
New York-Newark-Jersey City NY-NJ-PA								
Actual	40.1	53.1	53.8	58.0	82.6	71.8	-42.4	-18.7
Predicted	40.4	53.8	53.6	58.2	82.8	73.3	-42.4	-19.5
San Francisco-Oakland-Berkeley, CA								
Actual	40.0	47.6	48.4	49.4	81.4	77.3	-41.4	-29.7
Predicted	39.9	47.3	46.7	49.1	81.6	75.4	-41.8	-28.1

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Predicted values based on estimated coefficients from model reported in Table 3. Sample restricted to Census tracts with 10 or more devices with a home location in every period.

Table 6: Percent of Variance in Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020, Houston and San Francisco,

	Houston		San Francisco	
	OSW February 2020 and:		OSW February 2020 and:	
Explanatory variables	May 2020	March-April 2021	May 2020	March-April 2021
Share of population:				
Age 25-64	-2.1	-1.2	-1.3	-0.2
Age 65 plus	0.0	0.0	0.0	0.0
White, non-Hispanic	0.2	-0.1	0.1	-0.4
College graduate	6.1	2.1	3.4	1.9
ln(mean household income)	13.3	1.1	7.4	1.4
Share commute public trans.	0.2	0.4	0.2	0.5
Share commute 30+ mins.	0.5	0.1	0.3	0.3
Rural (yes/no)	0.1	0.1	0.1	0.1
Share Trump vote in 2016	-0.6	3.6	-0.4	-0.3
May 2020 cum COVID deaths	0.1	0.0	-0.1	0.0
Industry mix	19.0	8.8	18.5	14.4
Occupation mix	14.2	5.6	8.7	6.5
Residual	49.1	79.5	63.0	75.9
Dep. var. mean	54.0	66.6	40.0	49.4
Dep. var. standard deviation	(8.6)	(10.2)	(10.9)	(14.1)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation the Census tract. Dependent variable the share of February 2020 workers with on-site work activity in indicated month. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 10 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,125 Census tracts.

Table 7: Percent of Variance in Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019 Explained by Various Actors, Sample Restricted to Tracts with 10 or More Devices, State Effects Added

	OSW February 2020 and:				OSW February 2019 and:			
Explanatory variables	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:								
Age 25-64	-1.3	-0.8	-0.6	-0.6	0.7	0.1	-0.3	0.0
Age 65 plus	0.1	-0.2	-0.1	0.0	0.1	0.0	0.0	0.0
White, non-Hispanic	0.4	1.1	0.4	0.5	-0.3	0.3	0.0	0.2
College graduate	5.4	4.1	3.3	2.3	-0.4	0.0	2.6	1.9
ln(mean household income)	7.3	2.9	2.0	0.0	5.3	0.5	12.9	4.6
Share commute public trans.	0.4	0.5	0.4	0.8	0.1	-0.1	0.0	0.3
Share commute 30+ mins.	1.3	1.6	1.3	0.8	0.2	0.0	1.4	1.4
Rural (yes/no)	0.5	0.3	0.6	0.2	0.0	0.0	0.5	0.1
Share Trump vote in 2016	6.0	8.0	6.8	7.8	2.9	3.5	0.4	0.4
May 2020 state lockdown	--	--	--	--	--	--	--	--
May 2020 local lockdown	0.3	0.1	0.0	0.0	0.1	-0.1	0.4	0.1
May 2020 cum COVID deaths	1.9	0.0	0.1	0.2	0.0	0.5	1.4	-0.2
Industry mix	20.1	17.2	15.8	13.0	2.8	0.9	12.9	9.1
Occupation mix	7.5	6.7	4.5	4.2	1.2	-0.2	7.1	5.6
State effects	9.5	8.0	7.0	6.6	1.1	7.0	6.8	2.1
Residual	40.5	50.3	58.5	64.2	86.2	87.5	54.0	74.4
Dep. var. mean	53.0	62.6	61.4	66.7	82.6	78.4	-29.6	-15.8
Dep. var. standard deviation	(11.2)	(11.2)	(11.7)	(13.4)	(8.5)	(10.9)	(14.4)	(14.7)
Memo item:								
State effects only	22.4	19.3	15.8	14.4	2.5	7.4	15.8	5.7

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 10 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,125 Census tracts.

Table 8: Percent of Variance in Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019 Explained by Various Factors, Sample Restricted to Tracts with 10 or More Devices, County Effects Added

	OSW February 2020 and:				OSW February 2019 and:			
Explanatory variables	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:								
Age 25-64	-1.4	-1.0	-0.7	-0.7	0.7	0.1	-0.4	-0.1
Age 65 plus	0.1	-0.2	-0.2	-0.1	0.0	0.0	0.0	0.0
White, non-Hispanic	0.0	0.3	0.4	0.2	-0.2	-0.3	0.0	0.2
College graduate	6.1	4.9	4.4	2.5	-0.4	0.0	3.3	2.0
ln(mean household income)	7.0	3.0	1.8	0.2	4.5	0.5	11.6	4.5
Share commute public trans.	0.7	1.0	0.5	0.9	0.1	0.0	0.1	0.3
Share commute 30+ mins.	0.6	1.0	0.8	0.1	0.2	0.0	0.8	0.9
Rural (yes/no)	0.3	0.2	0.6	-0.1	0.0	-0.1	0.5	0.4
Share Trump vote in 2016	5.8	7.2	6.3	7.3	3.4	4.3	-0.1	-0.3
May 2020 state lockdown	--	--	--	--	--	--	--	--
May 2020 local lockdown	--	--	--	--	--	--	--	--
May 2020 cum COVID deaths	--	--	--	--	--	--	--	--
Industry mix	18.0	15.9	15.4	11.7	3.0	0.8	11.2	9.0
Occupation mix	6.1	5.2	3.6	3.5	1.2	-0.1	6.1	4.8
County effects	20.5	17.3	14.0	15.0	9.0	14.8	18.1	9.9
Residual	36.2	45.1	53.2	59.3	78.5	79.9	48.7	68.4
Dep. var. mean	53.0	62.6	61.4	66.7	82.6	78.4	-29.6	-15.8
Dep. var. standard deviation	(11.2)	(11.2)	(11.7)	(13.4)	(8.5)	(10.9)	(14.4)	(14.7)
Memo item:								
County effects only	50.3	45.2	38.4	35.0	13.0	16.7	36.5	22.4

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 10 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,125 Census tracts.

Table 9: Sensitivity of OSW Percentages and Explanatory Power of OSW Regression Models to Minimum Number of Devices with a Home Location Required for Inclusion of Tract in Sample

	Among those working on-site in February 2020, percent OSW as of:				Among those working on-site in February 2019, percent OSW as of:	
	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019
	One or More Devices Per Tract (67,550 Tracts)					
Mean	52.3	61.4	60.2	64.9	81.0	76.6
p10	35.2	43.6	40.9	41.6	66.6	57.7
p50	53.0	62.4	61.2	66.7	82.3	78.3
p90	68.4	77.8	77.9	86.3	94.0	94.7
R-Squared	0.453	0.332	0.253	0.204	0.106	0.057
	Ten or More Devices Per Tract (28,125 Tracts)					
Mean	53.0	62.6	61.4	66.7	82.6	78.4
p10	38.1	47.9	46.0	49.1	71.6	64.0
p50	53.6	63.2	62.1	67.6	83.4	79.5
p90	67.1	76.5	75.8	83.1	92.8	91.6
R-Squared	0.578	0.479	0.396	0.343	0.129	0.0870
	Twenty or More Devices Per Tract (7,932 Tracts)					
Mean	53.5	63.4	62.2	68.0	83.9	80.3
p10	40.4	50.7	49.2	53.5	75.5	69.1
p50	53.9	63.8	62.7	68.6	84.4	81.1
p90	66.1	75.9	74.7	81.7	91.8	90.8
R-Squared	0.652	0.575	0.503	0.440	0.160	0.142
	Thirty or More Devices Per Tract (2,629 Tracts)					
Mean	53.3	63.4	62.5	68.2	84.5	81.4
p10	42.1	51.5	50.7	55.2	77.1	71.8
p50	53.6	63.6	62.9	68.4	84.9	82.2
p90	64.8	75.0	74.1	80.8	91.4	90.5
R-Squared	0.704	0.648	0.593	0.516	0.195	0.189

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across Census tracts. R-squareds are values for models specified as in Table 3.

A. *Measurement Appendix*

This appendix describes the process we follow to estimate the onsite work share at the Census tract level and the covariates used in our decomposition analysis.

We make use of the repository of smart device location observations assembled by the Maryland Transportation Institute (MTI) and the Center for Advanced Transportation Technology (CATT) Lab. Appendix Table B1 provides summary statistics regarding the raw numbers of devices available for analysis. Our main estimates are based on the devices for which we are able to identify both a home and work location in February 2020 and for which we additionally can identify a home location in subsequent months. For some estimates, we make use of devices for which we can identify both a home and work location in February 2019 and for which we additionally can identify a home location in subsequent months. As shown in Appendix Table B1, our main analysis starts with a sample of approximately 4.2 million devices in February 2020 that shrinks through attrition to approximately 1.0 million devices in March-April 2021. The 2019 sample starts with approximately 2.1 million devices. The pace of attrition in the 2019 sample is similar to that in the 2020 sample.

1. *Identifying Home and Work Locations*

To identify each device's home and (if applicable) work location, we first determine the geohash for each sighting of the device in our data. The geohash system is a hierarchical spatial data structure that divides the surface of the earth into a grid. Based on the precision of the location information needed for each step, we utilize geohash information at geohash level 6, 7 or 8, where the higher-numbered levels represent successively smaller areas. The geohash cell dimensions vary with latitude; the table below shows the dimensions for each level in the worst-case scenario at the equator.

Geohash level	Area dimensions
6	1.2km*609.4m
7	152.9m*152.4m
8	38.2m*19m

The first step in our analysis was to filter the devices based on the total number of times the device was observed during a particular month. We removed devices with less than 100 sightings during the month from our dataset before doing anything else.

Next, we sought to identify a home location for each remaining device as follows:

1) Summarize the number of sightings, number of unique hours, and number of unique days for each device in each observed geohash level 7.
2) Identify the initial level 7 geohash candidate home locations for each device. These are the level 7 geohashes observed for at least 14 days and at least 60 unique hours during the month.
3) Summarize the number of sightings and the number of active hours for all candidate level 7 geohashes.
4) Among all candidate home locations at geohash level 7, select the one with the highest number of active hours. If there is more than one geohash level 7 with this number of observed active hours, select the one with the highest number of sightings.
5) After selecting the geohash level 7 home location, filter all sightings of the device within that geohash and calculate the number of unique sightings and unique hours at the geohash level 8 to get more detailed location information (each level 7 geohash has 32 level 8 geohashes)
6) Select the level 8 geohash with the highest number of unique hours within the selected level 7 geohash as the level 8 geohash home location. If there is more than one level 8 geohash with the same number of unique hours, select the one with the largest number of sightings.

For devices for which we could identify a home location, we then sought to identify a work location. The process of identifying a work location identification was mostly similar to the process for identifying the home location. One difference was that we did not allow the work location to be in the same level 6 geohash as the home location. The candidate level 7 geohashes for the work location were those in a different level 6 geohash observed for at least 60 unique hours and during at least two distinct weeks in the month.

The algorithm for identifying the work location also introduced a temporal similarity constraint designed to avoid mistakenly identifying a location near a device' home location as its work location. If a device dwells around the borders of adjacent geohash zones, its location could alternate across one or more of these neighboring or "twin" zones. These twin zones could be competitive with the true workplace zone with regard to visiting frequency, duration and regularity. Imposing a minimum commute distance threshold would be an alternative method for addressing this issue, but that approach runs the risk of compromising the identification of workplaces close to a device's home location. Based on the assumption that an onsite worker commutes from home to work, then stays at work for several consecutive hours, and then commutes back home, we should not observe the home and workplace location during the same hours too frequently. Based on that reasoning, we use a measure of similarity between the times we observe the home location and the times we observe the candidate work location in our procedures for identifying a device's work location.

Our measure of temporal similarity is defined as follows. For all the unique hours when a candidate workplace location was observed during the month, i.e., W^i for candidate location i, count the number of unique hours that overlap with the unique hours when the imputed home location was observed (H). The ratio between the overlapped hours and the total number of hours in W^i is then calculated. The ratio, referred to as temporal similarity ratio, measures the temporal similarity between home and workplace observations. The formula is given as follows:

$$S = \frac{\left|W^i \cap H\right|}{|W^i|}$$

In an ideal situation in which a device with a fixed work location is observed continuously through the day, the ratio should be less than or equal to $\frac{2}{number\ of\ daily\ work\ hours}$. When the commute time is less than one hour, this most often would be approximately 0.25 for someone working 8 hours per day; the inequality might apply in cases where an individual arrived at work or departed from work exactly on the hour. The ratio should always be zero when the commute time is longer than one hour.

In reality, the location observations are not complete. Through empirical experimentation with different thresholds, we selected a similarity ratio threshold of 0.6 as the best for ensuring a reliable work location identification; candidate workplace locations with similarity ratios above that threshold were rejected.

The exact procedure for workplace location identification is as follows:

1) Starting with the set of devices with an identified home location, summarize the number of sightings, number of unique hours, and number of unique days for each device at all level 7 geohashes.
2) Identify the initial level 7 geohash candidate work locations. These are locations that are not the home location or within the same level 6 geohash as the home location; are observed for at least 60 hours and in two different weeks during the month; and do not violate the temporal similarity constraint.
3) Summarize the number of sightings and the number of active hours for all candidate level 7 geohashes.
4) Among all of the candidate level 7 geohashes, select the one with the highest number of unique hours. If there are multiple candidates with the same number of unique hours, select the work location based on the number of sightings.
5) Within the selected geohash level 7, summarize the number of unique hours and number of sightings at geohash level 8.
6) Select the level 8 geohash with the highest number of unique hours as the work location. If multiple candidates with the same number of hours exist, choose based on the number of sightings.

We considered using time of day as an additional criterion for determining home and work locations, giving additional weight to hours observed overnight in selecting the home location and to hours observed during the day in selecting the work location. This did not seem to lead to any improvement and we treat all hours in the same way, regardless of the time of day.

2. *Protecting Against Misclassification*

The process described above is implemented in all months. For the 2020 and 2021 analysis, we restrict our sample to the devices for which we can identify a home and work location in February 2020 and a home location in subsequent months. For example, the May 2020 sample consists of the 2.8 million devices for which we observe a home and work location in February 2020 and a home location in May 2020. There are at least three reasons we might not be able to identify a work location in a later month:

1) The quality of the underlying data deteriorates, so that even if a person is going to work at a fixed work location, the second most observed location does not meet the 60 hours threshold.
2) The person is working from home.
3) The person is not working.

To protect against not identifying a work location because of the first of these factors, we implement a modified hours threshold in later months. We base the modified threshold on the ratio of the total number of unique hours observed in the follow-up month to the total number of unique hours observed in February. For example, if we observe a device for 300 unique hours in February and for 150 unique hours in the follow up month, that is a 50 percent reduction. The same proportional reduction is applied to the threshold for identifying a work location in the later month, so that the threshold is set at 30 hours rather than 60 hours. The lowered threshold was applied only for identifying the work location, not the home location.

3. *Attrition Analysis*

The share of devices that we observe in later months falls relative to the number observed in February 2020 or February 2019. To correct for possible bias associated with nonrandom attrition, we reweight the devices so that they are more representative. We do this using block group level information for the identified February home location from the 2015-2019 American Community Survey (ACS). We fit a logit model with a zero/one dummy variable for whether the device was observed in follow-up month as the dependent variable. The explanatory variables in the attrition model are the share of residents in the block group aged 25-64, the share aged 65 or older, the share who were White non-Hispanic, the share aged 25 and older with a college degree, the logarithm of mean household income in the block group, and a dummy variable identifying whether the block group is in a tract designated as rural, all defined for the block group where the device was observed in February 2020.

The attrition weights are calculated as the inverse of the predicted probability of observing a device in a given month as a function of the characteristics of the block group of the device's February home location. If information for the block group is missing, we use an attrition weight of one.

4. *Iterative Proportional Fitting*

We also have constructed a second set of device weights designed to make our weighted sample consistent with the information in the Census Bureau's LEHD Origin-Destination Employment Statistics (LODES) file. This file contains estimates of the number of workers living in each block group by the block group of their job location. The goal of this weighting step is to match the marginal home and work location distributions at the county level. To accomplish this, we apply an iterative proportional fitting (IPF) algorithm. All devices that share the same home and work county end up with the same IPF weight. One caution is that this procedure generates considerable dispersion in the IPF weights as, in some cases, the share of sample devices in a marginal cell differs considerably from that in the LODES data. Briefly, the algorithm we apply is as follows:

1) First, create a seed matrix by aggregating the attrition weights for the devices we observe in February 2020 (or February 2019) with a home location in county *i* and a work location in county j. This produces a two-way table with home counties as the rows and work counties as the columns. The entry in each cell is x_{ij}, the sum of the attrition weights for the relevant devices, where *i* represents the county of the device's home location and *j* the county of the device's work location.
2) Next, using the LODES data, calculate the marginal number of residents and workers at the county level, u_i and v_j
3) The goal for the process is to derive a new table consisting of entries $\widehat{m}_{ij} = a_i b_j x_{ij}$ for all i and j such that the marginals become equal to: $\sum_j \widehat{m}_{ij} = u_i$, and $\sum_i \widehat{m}_{ij} = v_j$, where u_i is the number of workers resident in the county and v_i is the number of jobs located in the county.
4) Start with the initial value of $\widehat{m}_{ij}^{(0)} = x_{ij}$
5) Then, in each step, revise the estimates according to:
   a) $\widehat{m}_{ij}^{(2n-1)} = \frac{\widehat{m}_{ij}^{(2n-2)} u_i}{\sum_{k=1}^{j} \widehat{m}_{ik}^{(2n-2)}}$

   b) $\widehat{m}_{ij}^{(2n)} = \frac{\widehat{m}_{ij}^{(2n-1)} v_j}{\sum_{k=1}^{i} \widehat{m}_{kj}^{(2n-1)}}$
6) Repeat step 5 until the row and column totals are sufficiently close to u_i and v_j (within 0.1 percent).

5. *Device-level Composite Weight*

Once the IPF weights have been derived, we calculate the device level final weights as the product of the attrition weight for that device and the IPF weight corresponding to the home and work location of the device at the county level. The attrition weights assign higher weights to devices in block groups that, because of their sociodemographic characteristics, are more likely to be missing in the current month. The IPF weights assign a higher weight to devices that are in county pairs, consisting of an origin (home) county and a destination (work) county, that are less common in our data than expected based on the LODES administrative data.

6. *Tract Level Weights*

Using the composite weights and an indicator variable for whether the device has a work location in a given month, we estimate the weighted share of devices in a Census tract for which we identified an onsite work location in the month. This leaves us with one observation per tract—an estimate of the share of onsite workers from February 2020 or February 2019 who are onsite workers in the current month. For both the descriptive statistics and the regression analysis, the estimates we report are weighted using 2019 tract-level employment from the American Community Survey (ACS).

7. *Explanatory Variables Used for Attrition Weighting and Included in Onsite Work Regressions*

The explanatory variables used in the attrition weighting models and onsite work regressions were constructed using information from the following sources.

American Community Survey 2015-2019 averages

Share of tract residents aged 25-64 and aged 65 plus

Share of tract residents age 25 and older with a college degree

Share of tract residents who are White non-Hispanics

Mean household income in the tract in thousands of 2019 dollars

For tracts with missing information on education or income, those variables are assigned a value of zero and an indicator variable for missing education or income, as appropriate, assigned a value of one.

Data from the ACS 5-year files and downloaded from IPUMS at https://usa.ipums.org/usa/

Department of Agriculture

Zero-one dummy variable for whether tract was rural (rather than urban) as of 2010.

Data downloaded from https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/

Voting and Elections Science Team

Average share of votes for Donald Trump for president in 2016 in precincts where the devices in a tract are located. Given information on device location, each device can be mapped to a precinct. Tract-level estimate is the weighted average of the vote shares based on the precinct information for the devices in a given tract.

Data downloaded from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NH5S2I

Goolsbee et al. (2020) COVID restriction information

Goolsbee et al. (2020) compiled information on the start and end dates for stay-at-home orders at the state and (if applicable) county level. We used that information to calculate the share of the month of May 2020 during which these orders were in effect. These same variables were used in the regressions for all months, not just May 2020.

Data downloaded from https://bfi.uchicago.edu/working-paper/2020-116/

New York Times COVID database

Cumulative county-level COVID deaths as of the end of May 2020.

Data downloaded from https://github.com/nytimes/covid-19-data#cumulative-cases-and-deaths

LODES data

Share of employed residents who are employed in each two-digit NAICS sector. Data are for 2019 for most states but, for a few states, the latest available year is earlier (2016 for Alaska, 2017 for Arkansas and 2018 for Mississippi).

Data downloaded from https://lehd.ces.census.gov/data/lodes/LODES7/

B. Supplemental Tables

Appendix Table B1: Unweighted Mobile Device Counts by Month

Selection criterion	February 2020	May 2020	August 2020	November 2020	March-April 2021
Home location identified	11,659,409	15,565,504	15,038,155	13,860,961	13,710,998
Work location identified | home location identified	4,230,606	3,186,920	3,478,169	3,200,293	5,206,351
Home location identified | work in February 2020	4,230,606	2,804,839	2,058,125	1,571,430	986,972
Work location identified | work in February 2020	4,230,606	1,099,422	946,327	664,907	609,562

Selection criterion	February 2019	May 2019	August 2019
Home location identified	9,392,392	9,625,420	7,454,499
Work location identified | home location identified	2,082,327	2,499,623	1,911,941
Home location identified | work in February 2019	2,082,327	1,276,135	846,028
Work location identified | work in February 2019	2,082,327	972,265	607,577

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Appendix Table B2: Unweighted Distribution of Observed Hours for Devices for Which Home and Work Location Identified in February 2020 or February 2019 and Home Location Identified in a Subsequent Month

	February 2020-May 2020 Sample	February 2020-August 2020 Sample	February 2020-November 2020 Sample	February 2020-March-April 2021 Sample	February 2019-May 2019 Sample	February 2019-August 2019 Sample
Number of devices in sample	2,804,839	2,058,125	1,571,430	986,972	1,276,135	846,028
Hours devices observed in February						
Mean	579	587	602	612	419	431
10th percentile	321	331	377	418	266	294
50th percentile	650	655	660	663	394	398
90th percentile	687	688	688	690	605	612
Hours devices observed in end month						
Mean	381	380	381	558	505	534
10th percentile	196	212	209	286	238	232
50th percentile	372	376	380	632	547	611
90th percentile	584	555	552	722	702	725

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Appendix Table B3: Industry Effects on Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019, Sample Restricted to Tracts with 10 or More Devices

		OSW February 2020 and:				OSW February 2019 and:			
Industry	Mean employment share (percent)	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Agriculture	0.6	0.498	0.334	0.397	0.313	-0.151	0.013	0.649	0.321
		(0.054)	(0.060)	(0.067)	(0.080)	(0.058)	(0.077)	(0.079)	(0.094)
Mining	0.5	0.273	0.243	0.190	0.339	-0.211	-0.139	0.485	0.381
		(0.040)	(0.044)	(0.050)	(0.059)	(0.043)	(0.057)	(0.058)	(0.070)
Utilities	0.6	0.487	0.701	0.584	0.526	0.346	1.014	0.141	-0.313
		(0.120)	(0.133)	(0.150)	(0.179)	(0.130)	(0.171)	(0.176)	(0.210)
Construction	5.4	0.593	0.244	0.352	0.317	-0.297	-0.226	0.890	0.470
		(0.038)	(0.043)	(0.048)	(0.058)	(0.042)	(0.055)	(0.057)	(0.067)
Manufacturing	9.2	0.347	0.448	0.362	0.315	-0.034	0.049	0.381	0.398
		(0.028)	(0.031)	(0.035)	(0.042)	(0.030)	(0.040)	(0.041)	(0.049)
Wholesale trade	4.3	0.219	0.449	0.344	0.258	0.286	0.245	-0.067	0.204
		(0.050)	(0.056)	(0.063)	(0.075)	(0.055)	(0.072)	(0.074)	(0.088)
Retail trade	10.7	0.686	0.685	0.633	0.698	-0.142	-0.220	0.828	0.905
		(0.042)	(0.046)	(0.052)	(0.062)	(0.045)	(0.059)	(0.061)	(0.073)
Transportation and	3.9	0.379	0.320	0.283	0.249	-0.196	0.102	0.575	0.225
warehousing		(0.039)	(0.043)	(0.049)	(0.058)	(0.042)	(0.056)	(0.057)	(0.068)
Information	2.0	-0.011	0.023	-0.065	-0.295	-0.071	0.174	0.059	-0.151
		(0.041)	(0.045)	(0.051)	(0.061)	(0.044)	(0.058)	(0.060)	(0.071)
Finance and insurance	4.3	--	--	--	--	--	--	--	--
		--	--	--	--	--	--	--	--
Real Estate	1.5	-0.166	-0.531	-0.295	-0.618	-0.603	0.132	0.437	-0.663
		(0.117)	(0.130)	(0.146)	(0.175)	(0.127)	(0.167)	(0.172)	(0.205)
Professional services	6.6	-0.042	0.032	-0.066	-0.110	-0.213	-0.079	0.172	0.111
		(0.036)	(0.040)	(0.045)	(0.053)	(0.039)	(0.051)	(0.053)	(0.063)
Management of	1.8	-0.048	-0.106	-0.192	-0.141	-0.131	0.037	0.083	-0.142
companies		(0.055)	(0.062)	(0.069)	(0.083)	(0.060)	(0.079)	(0.081)	(0.097)
Administrative and	6.2	0.231	0.489	0.322	0.211	-0.327	-0.073	0.558	0.562
support services		(0.046)	(0.051)	(0.057)	(0.068)	(0.049)	(0.065)	(0.067)	(0.080)
Education	9.4	0.123	0.303	0.332	0.333	-0.083	-0.020	0.206	0.323
		(0.031)	(0.034)	(0.038)	(0.046)	(0.033)	(0.044)	(0.045)	(0.054)
Health	14.6	0.473	0.489	0.473	0.428	-0.061	0.073	0.534	0.416
		(0.032)	(0.036)	(0.040)	(0.048)	(0.035)	(0.046)	(0.047)	(0.056)
Arts and entertainment	1.8	-0.020	0.126	0.196	0.205	-0.243	-0.227	0.223	0.353
		(0.049)	(0.055)	(0.062)	(0.074)	(0.053)	(0.070)	(0.072)	(0.086)
Accommodations and	9.1	0.075	0.421	0.345	0.329	-0.172	0.021	0.248	0.401
food services		(0.032)	(0.036)	(0.040)	(0.048)	(0.035)	(0.046)	(0.047)	(0.056)
Other services	3.0	0.317	0.529	0.437	0.677	0.043	-0.315	0.273	0.844
		(0.075)	(0.083)	(0.094)	(0.112)	(0.081)	(0.107)	(0.110)	(0.131)
Public administration	4.4	0.286	0.379	0.142	0.147	-0.070	0.044	0.356	0.334
		(0.033)	(0.036)	(0.041)	(0.049)	(0.035)	(0.047)	(0.048)	(0.057)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 10 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,125 Census tracts.

Appendix Table B4: Occupation Effects on Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019, Sample Restricted to Tracts with 10 or More Devices

		OSW February 2020 and:				OSW February 2019 and:			
Occupation	Mean employment share	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Management, business and financial	16.3	0.018	0.057	0.005	0.090	-0.035	-0.014	0.054	0.071
		(0.019)	(0.021)	(0.024)	(0.028)	(0.020)	(0.027)	(0.028)	(0.033)
Computer, engineering and science	9.9	--	--	--	--	--	--	--	--
		--	--	--	--	--	--	--	--
Education, legal, community service, arts and media	10.6	0.084	0.148	0.120	0.164	-0.075	0.001	0.160	0.148
		(0.020)	(0.022)	(0.025)	(0.030)	(0.021)	(0.028)	(0.029)	(0.035)
Healthcare practicioners and technical	6.5	0.157	0.224	0.124	0.247	-0.020	0.027	0.177	0.198
		(0.023)	(0.026)	(0.029)	(0.035)	(0.025)	(0.033)	(0.034)	(0.041)
Healthcare support	2.8	0.105	0.104	0.032	0.103	-0.039	-0.008	0.144	0.112
		(0.029)	(0.033)	(0.037)	(0.044)	(0.032)	(0.042)	(0.043)	(0.052)
Protective service	2.2	0.067	0.135	0.073	0.181	-0.055	-0.086	0.122	0.221
		(0.032)	(0.036)	(0.040)	(0.048)	(0.035)	(0.046)	(0.048)	(0.057)
Food preparation and serving	5.3	0.106	0.168	0.145	0.181	-0.120	-0.027	0.225	0.195
		(0.023)	(0.026)	(0.029)	(0.035)	(0.025)	(0.033)	(0.034)	(0.041)
Building and grounds cleaning and maintenance	3.3	0.144	0.217	0.104	0.202	-0.041	-0.018	0.186	0.235
		(0.026)	(0.029)	(0.033)	(0.039)	(0.028)	(0.037)	(0.038)	(0.046)
Personal care and service	2.7	-0.031	0.129	0.105	0.136	-0.062	-0.014	0.031	0.143
		(0.030)	(0.034)	(0.038)	(0.045)	(0.033)	(0.043)	(0.045)	(0.053)
Sales and related	10.6	0.052	0.098	0.101	0.128	-0.055	-0.019	0.107	0.117
		(0.020)	(0.022)	(0.025)	(0.030)	(0.022)	(0.029)	(0.030)	(0.035)
Office and administrative support	11.7	0.090	0.139	0.095	0.106	-0.038	-0.021	0.128	0.160
		(0.021)	(0.023)	(0.026)	(0.031)	(0.023)	(0.030)	(0.031)	(0.037)
Farming, fishing and forestry	0.5	0.142	0.185	0.036	0.178	-0.151	-0.194	0.293	0.380
		(0.051)	(0.057)	(0.064)	(0.076)	(0.055)	(0.073)	(0.075)	(0.089)
Construction and extraction	4.9	0.147	0.163	0.122	0.172	-0.079	-0.118	0.226	0.281
		(0.023)	(0.025)	(0.028)	(0.034)	(0.025)	(0.032)	(0.033)	(0.040)
Installation, maintenance, and repair	3.3	0.199	0.200	0.130	0.222	-0.067	-0.137	0.266	0.337
		(0.029)	(0.032)	(0.036)	(0.044)	(0.032)	(0.042)	(0.043)	(0.051)
Production	5.7	0.180	0.251	0.163	0.272	-0.078	-0.059	0.258	0.310
		(0.022)	(0.025)	(0.028)	(0.034)	(0.024)	(0.032)	(0.033)	(0.039)
Transportation	3.7	0.138	0.214	0.089	0.168	-0.027	-0.045	0.165	0.259
		(0.027)	(0.030)	(0.034)	(0.040)	(0.029)	(0.038)	(0.039)	(0.047)
Material moving	3.5	0.146	0.160	0.093	0.136	-0.094	-0.045	0.240	0.205
		(0.026)	(0.029)	(0.033)	(0.039)	(0.028)	(0.037)	(0.038)	(0.046)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 10 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,125 Census tracts.

Appendix Table B5.A: Factors Affecting Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019, All Tracts with One or More Devices

Explanatory variables	Mean (standard deviation)	OSW February 2020 and: May 2020	August 2020	November 2020	March-April 2021	OSW February 2019 and: May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:									
Age 25-64	52.9	0.155	0.103	0.098	0.117	0.119	0.101	0.035	0.002
	(6.9)	(0.007)	(0.008)	(0.009)	(0.012)	(0.008)	(0.010)	(0.010)	(0.013)
Age 65 plus	15.3	0.014	-0.029	-0.015	-0.022	0.009	-0.021	0.005	-0.008
	(6.8)	(0.007)	(0.008)	(0.010)	(0.012)	(0.008)	(0.011)	(0.010)	(0.013)
White, non-Hispanic	62.4	-0.006	0.003	-0.024	0.021	-0.024	-0.040	0.018	0.042
	(28.2)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)	(0.005)
College graduate	41.9	-0.049	-0.037	-0.030	-0.035	-0.032	-0.023	-0.017	-0.014
	(18.6)	(0.007)	(0.008)	(0.009)	(0.011)	(0.007)	(0.010)	(0.010)	(0.013)
ln(mean household income	4.2	-3.877	-2.220	-1.903	-0.213	4.734	2.161	-8.612	-4.380
	(0.4)	(0.175)	(0.206)	(0.238)	(0.299)	(0.196)	(0.265)	(0.261)	(0.328)
Share commute public trans.	17.8	-0.008	-0.027	-0.035	-0.038	-0.011	-0.018	0.003	-0.008
	(17.5)	(0.003)	(0.004)	(0.004)	(0.005)	(0.003)	(0.005)	(0.005)	(0.006)
Share commute 30+ mins.	37.9	-0.046	-0.038	-0.034	-0.031	0.021	0.023	-0.067	-0.061
	(15.9)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
Rural yes/no	14.9	0.009	0.005	0.004	0.006	-0.005	-0.004	0.013	0.009
	(35.6)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Share Trump vote in 2016	44.9	0.100	0.137	0.147	0.148	0.087	0.132	0.012	0.004
	(21.1)	(0.004)	(0.005)	(0.005)	(0.007)	(0.004)	(0.006)	(0.006)	(0.007)
May 2020 state lockdown	51.3	-0.039	-0.032	-0.029	-0.041	0.004	-0.018	-0.042	-0.014
	(42.7)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
May 2020 local lockdown	5.2	-0.017	-0.011	-0.013	-0.008	-0.003	-0.007	-0.014	-0.003
	(18.6)	(0.002)	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.004)
May 2020 cum COVID deaths	2.1	-0.481	-0.089	-0.045	0.005	-0.058	-0.294	-0.422	0.205
	(3.2)	(0.014)	(0.016)	(0.019)	(0.024)	(0.016)	(0.021)	(0.021)	(0.026)
Industry dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	--	52.3	61.4	60.2	64.9	81.0	76.6	-28.7	-15.2
Dep. var. standard deviation	--	(13.3)	(14.1)	(15.4)	(18.7)	(11.6)	(15.2)	(17.8)	(20.0)
R-squared	--	0.453	0.332	0.253	0.204	0.106	0.057	0.331	0.156

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample includes all Census tracts for which a home location is observed for at least one device in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=67,550.

Appendix Table B5.B: Factors Affecting Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019, All Tracts with 20 or More Devices

Explanatory variables	Mean (standard deviation)	OSW February 2020 and: May 2020	August 2020	November 2020	March-April 2021	OSW February 2019 and: May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:									
Age 25-64	52.9	0.142	0.118	0.0797	0.122	0.0879	0.0532	0.0538	0.0647
	(5.0)	(0.016)	(0.017)	(0.019)	(0.022)	(0.016)	(0.022)	(0.023)	(0.026)
Age 65 plus	13.5	-0.022	-0.036	-0.014	-0.027	0.014	-0.020	-0.036	-0.016
	(5.4)	(0.015)	(0.017)	(0.019)	(0.022)	(0.016)	(0.021)	(0.022)	(0.026)
White, non-Hispanic	67.5	0.018	0.027	-0.018	0.028	0.001	-0.006	0.017	0.033
	(23.4)	(0.006)	(0.006)	(0.007)	(0.008)	(0.006)	(0.008)	(0.008)	(0.009)
College graduate	45.7	-0.026	0.003	-0.041	-0.014	-0.037	-0.005	0.011	0.008
	(15.3)	(0.013)	(0.014)	(0.016)	(0.018)	(0.013)	(0.018)	(0.019)	(0.021)
ln(mean household income	4.3	-5.512	-3.658	-3.611	-2.923	4.526	1.209	-10.040	-4.867
	(0.3)	(0.387)	(0.427)	(0.470)	(0.552)	(0.401)	(0.529)	(0.560)	(0.644)
Share commute public trans.	10.5	-0.018	-0.035	-0.043	-0.058	-0.004	-0.028	-0.014	-0.007
	(10.7)	(0.007)	(0.008)	(0.009)	(0.011)	(0.008)	(0.010)	(0.011)	(0.012)
Share commute 30+ mins.	40.4	-0.033	-0.040	-0.030	-0.024	0.012	0.029	-0.045	-0.068
	(15.9)	(0.006)	(0.006)	(0.007)	(0.008)	(0.006)	(0.008)	(0.008)	(0.010)
Rural yes/no	9.7	0.009	0.007	0.009	0.011	-0.003	0.004	0.012	0.003
	(29.5)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Share Trump vote in 2016	56.2	0.114	0.172	0.183	0.148	0.054	0.148	0.060	0.025
	(17.0)	(0.008)	(0.009)	(0.010)	(0.011)	(0.008)	(0.011)	(0.012)	(0.013)
May 2020 state lockdown	29.6	-0.038	-0.033	-0.031	-0.035	0.001	-0.024	-0.039	-0.009
	(38.8)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)
May 2020 local lockdown	3.7	-0.023	-0.024	-0.001	0.007	0.011	0.006	-0.034	-0.031
	(14.9)	(0.005)	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)	(0.008)
May 2020 cum COVID deaths	1.3	-0.482	-0.047	-0.103	-0.052	-0.010	-0.358	-0.472	0.311
	(2.2)	(0.035)	(0.038)	(0.042)	(0.050)	(0.036)	(0.048)	(0.050)	(0.058)
Industry dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	--	53.5	63.4	62.2	68.0	83.9	80.3	-30.4	-16.9
Dep. var. standard deviation	--	(9.8)	(9.8)	(10.0)	(11.1)	(6.6)	(8.6)	(12.1)	(11.8)
R-squared	--	0.652	0.575	0.503	0.440	0.160	0.142	0.522	0.331

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019.Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 20 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=7,932.

Appendix Table B5.C: Factors Affecting Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019, All Tracts with 30 or More Devices

Explanatory variables	Mean (standard deviation)	OSW February 2020 and: May 2020	August 2020	November 2020	March-April 2021	OSW February 2019 and: May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:									
ge 25-64	53.04	-0.016	0.007	-0.022	-0.007	-0.009	0.034	-0.007	-0.026
	(4.44)	(0.026)	(0.029)	(0.031)	(0.037)	(0.027)	(0.035)	(0.038)	(0.043)
ge 65 plus	12.16	0.099	0.099	0.021	0.055	0.059	0.044	0.040	0.056
	(4.94)	(0.026)	(0.029)	(0.031)	(0.036)	(0.027)	(0.035)	(0.037)	(0.042)
hite, non-Hispanic	66.26	0.005	0.006	-0.036	0.008	-0.004	-0.021	0.009	0.027
	(22.96)	(0.009)	(0.010)	(0.011)	(0.013)	(0.010)	(0.012)	(0.013)	(0.015)
ollege graduate	48.05	-0.008	0.019	-0.050	-0.001	-0.040	0.027	0.031	-0.008
	(14.60)	(0.020)	(0.022)	(0.024)	(0.028)	(0.021)	(0.027)	(0.029)	(0.033)
ln(mean household income	4.41	-6.444	-4.194	-4.038	-3.512	5.053	1.169	-11.500	-5.363
	(0.32)	(0.627)	(0.698)	(0.746)	(0.889)	(0.657)	(0.847)	(0.909)	(1.035)
Share commute public trans.	8.80	-0.002	-0.034	-0.024	-0.054	0.010	-0.022	-0.011	-0.012
	(8.61)	(0.013)	(0.014)	(0.015)	(0.018)	(0.013)	(0.017)	(0.018)	(0.021)
Share commute 30+ mins.	42.83	-0.039	-0.038	-0.034	-0.029	0.006	0.037	-0.045	-0.075
	(16.13)	(0.008)	(0.009)	(0.010)	(0.012)	(0.009)	(0.011)	(0.012)	(0.014)
l yes/no	6.47	0.001	-0.002	0.006	0.000	-0.002	0.005	0.003	-0.006
	(24.60)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)	(0.008)
Share Trump vote in 2016	57.90	0.142	0.213	0.224	0.177	0.058	0.168	0.085	0.046
	(16.55)	(0.013)	(0.015)	(0.016)	(0.018)	(0.014)	(0.018)	(0.019)	(0.022)
May 2020 state lockdown	20.83	-0.035	-0.029	-0.026	-0.027	0.001	-0.024	-0.036	-0.005
	(34.03)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
May 2020 local lockdown	2.38	-0.026	-0.032	0.002	0.011	0.015	0.005	-0.040	-0.037
	(11.19)	(0.009)	(0.010)	(0.011)	(0.013)	(0.010)	(0.013)	(0.014)	(0.015)
May 2020 cum COVID deaths	1.02	-0.417	0.054	-0.167	-0.102	-0.040	-0.213	-0.377	0.267
	(1.49)	(0.072)	(0.081)	(0.086)	(0.103)	(0.076)	(0.098)	(0.105)	(0.119)
Industry dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	--	53.3	63.4	62.5	68.2	84.5	81.4	-31.1	-18.0
Dep. var. standard deviation	--	(9.0)	(9.1)	(9.1)	(9.9)	(5.7)	(7.3)	(10.8)	(10.4)
R-squared	--	0.704	0.648	0.593	0.516	0.195	0.189	0.574	0.399

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 30 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=2,629.

Appendix Table B6.A: Percent of Variance in Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019 Explained by Various Factors, All Tracts with One or More Devices

	OSW February 2020 and:				OSW February 2019 and:			
Explanatory variables	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:								
Age 25-64	-1.3	-0.8	-0.7	-0.5	0.7	0.2	-0.2	0.0
Age 65 plus	0.0	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0
White, non-Hispanic	-0.1	0.1	-0.3	0.4	-0.9	-0.8	-0.2	0.0
College graduate	3.6	2.1	1.4	1.0	-0.6	-0.1	0.9	0.4
ln(mean household income)	5.7	2.4	1.6	0.1	4.4	0.6	10.2	3.1
Share commute public trans.	0.4	1.1	1.2	1.0	0.1	0.2	-0.1	0.1
Share commute 30+ mins.	1.2	0.9	0.6	0.4	0.2	0.0	1.3	0.8
Rural (yes/no)	0.7	0.3	0.2	0.3	0.0	0.0	0.6	0.3
Share Trump vote in 2016	5.5	7.2	6.2	5.4	2.7	3.2	0.2	0.1
May 2020 state lockdown	3.5	2.2	1.6	1.8	0.0	0.5	2.3	0.3
May 2020 local lockdown	0.2	0.1	0.1	0.0	0.0	0.0	0.1	0.0
May 2020 cum COVID deaths	3.4	0.4	0.2	0.0	0.0	0.5	1.7	-0.2
Industry mix	15.7	11.6	9.9	8.2	3.0	0.9	10.6	6.4
Occupation mix	6.9	5.9	3.5	2.5	1.0	0.4	5.8	4.5
Residual	54.7	66.8	74.7	79.6	89.4	94.3	66.9	84.4
Dep. var. mean	53.2	62.1	60.8	65.4	80.2	75.9	-26.9	-13.8
Dep. var. standard deviation	(14.3)	(15.3)	(17.1)	(21.1)	(13.3)	(17.6)	(19.7)	(22.6)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample includes all Census tracts for which a home location is observed for at least one device in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=67,550.

Appendix Table B6.B: Percent of Variance in Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019 Explained by Various Factors, All Tracts with 20 or More Devices

	OSW February 2020 and:				OSW February 2019 and:			
Explanatory variables	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:								
Age 25-64	-1.3	-1.2	-0.8	-1.0	0.8	0.0	-0.5	-0.5
Age 65 plus	-0.3	-0.5	-0.2	-0.3	0.1	0.0	-0.2	-0.1
White, non-Hispanic	0.8	1.7	-0.7	1.5	0.1	-0.3	0.2	0.7
College graduate	2.5	-0.3	3.2	0.9	-1.0	0.0	-0.8	-0.5
ln(mean household income)	11.2	6.5	6.1	3.7	6.4	0.3	17.5	6.6
Share commute public trans.	0.7	1.3	1.6	1.8	0.0	0.6	0.4	0.1
Share commute 30+ mins.	1.3	1.8	1.1	0.7	0.2	0.0	1.4	2.2
Rural (yes/no)	0.9	0.7	0.7	0.8	0.1	0.1	0.8	0.2
Share Trump vote in 2016	8.8	14.4	13.5	10.1	2.5	8.1	2.2	0.7
May 2020 state lockdown	4.5	3.2	3.2	2.9	0.0	2.0	3.5	0.2
May 2020 local lockdown	0.5	0.5	0.0	-0.1	0.1	-0.1	0.5	0.3
May 2020 cum COVID deaths	2.8	0.2	0.4	0.1	0.0	1.3	2.1	-0.2
Industry mix	21.4	18.0	18.4	16.1	5.0	2.0	14.6	12.7
Occupation mix	11.3	11.1	3.9	6.7	1.8	0.2	10.6	10.7
Residual	34.8	42.5	49.7	56.0	84.0	85.8	47.8	66.9
Dep. var. mean	53.5	63.4	67.2	68.0	83.9	80.3	-30.4	-16.9
Dep. var. standard deviation	(9.8)	(9.8)	(10.0)	(11.1)	(6.6)	(8.6)	(12.1)	(11.8)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019.Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 20 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=7,932.

Appendix Table B6.C: Percent of Variance in Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019 Explained by Various Factors, All Tracts with 30 or More Devices

Explanatory variables	OSW February 2020 and: May 2020	August 2020	November 2020	March-April 2021	OSW February 2019 and: May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:								
Age 25-64	-1.0	-1.0	-0.2	-0.5	0.4	0.0	-0.3	-0.4
Age 65 plus	-0.3	0.1	-0.3	-0.1	0.0	0.2	-0.1	-0.3
White, non-Hispanic	0.3	0.5	-1.9	0.5	-0.3	-1.4	0.2	0.8
College graduate	0.9	-1.7	4.5	0.1	-1.3	0.1	-2.5	0.5
ln(mean household income)	13.7	7.5	7.3	4.9	8.0	0.4	21.7	8.4
Share commute public trans.	0.0	1.0	0.7	1.4	0.0	0.3	0.2	0.2
Share commute 30+ mins.	1.9	2.1	1.5	1.2	0.1	0.0	1.6	3.2
Rural (yes/no)	0.1	-0.1	0.4	0.0	0.0	0.1	0.2	-0.3
Share Trump vote in 2016	12.5	20.4	19.1	13.9	3.3	12.4	3.7	1.7
May 2020 state lockdown	3.6	2.4	2.5	2.1	0.0	1.9	2.9	0.1
May 2020 local lockdown	0.4	0.5	0.0	-0.1	0.1	0.0	0.5	0.3
May 2020 cum COVID deaths	0.9	0.0	0.3	0.1	-0.1	0.3	0.7	0.0
Industry mix	23.3	20.9	22.5	21.8	7.1	3.2	16.7	15.2
Occupation mix	14.2	12.1	3.0	6.1	2.0	1.4	11.9	10.5
Residual	29.6	35.2	40.7	48.4	80.5	81.1	42.6	60.1
Dep. var. mean	53.3	63.4	62.5	68.2	84.5	81.4	-31.1	-18.0
Dep. var. standard deviation	(9.0)	(9.1)	(9.1)	(9.9)	(5.7)	(7.3)	(10.8)	(10.4)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample restricted to Census tracts with 30 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=2,629.

Appendix Table B7: Factors Affecting Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019, Work Sites within 100 Yards of a School Dropped from Analysis, Sample Restricted to Tracts with 10 or More Devices

Explanatory variables	Mean (standard deviation)	OSW February 2020 and: May 2020	August 2020	November 2020	March-April 2021	OSW February 2019 and: May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:									
Age 25-64	52.8	0.137	0.087	0.081	0.098	0.100	0.079	0.037	0.00868
	(5.8)	(0.009)	(0.010)	(0.011)	(0.014)	(0.010)	(0.013)	(0.014)	(0.016)
Age 65 plus	14.8	0.010	-0.023	-0.002	0.004	0.026	0.000	-0.0162	-0.0228
	(6.0)	(0.009)	(0.010)	(0.012)	(0.014)	(0.010)	(0.013)	(0.014)	(0.016)
White, non-Hispanic	66.0	-0.002	0.016	-0.023	0.020	-0.006	-0.017	0.004	0.033
	(25.2)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
College graduate	43.4	-0.038	-0.026	-0.035	-0.029	-0.030	-0.023	-0.009	-0.004
	(16.5)	(0.008)	(0.009)	(0.010)	(0.012)	(0.009)	(0.012)	(0.012)	(0.014)
ln(mean household income	4.3	-4.791	-2.944	-2.500	-1.253	4.710	1.571	-9.501	-4.516
	(0.4)	(0.228)	(0.253)	(0.285)	(0.340)	(0.247)	(0.325)	(0.335)	(0.399)
Share commute public trans.	14.2	-0.009	-0.020	-0.026	-0.026	-0.010	-0.012	0.001	-0.008
	(14.5)	(0.004)	(0.004)	(0.005)	(0.006)	(0.004)	(0.006)	(0.006)	(0.007)
Share commute 30+ mins.	38.6	-0.044	-0.045	-0.044	-0.035	0.015	0.034	-0.060	-0.079
	(15.7)	(0.004)	(0.004)	(0.005)	(0.006)	(0.004)	(0.005)	(0.005)	(0.007)
Rural yes/no	12.8	0.008	0.005	0.008	0.006	-0.002	0.001	0.010	0.004
	(33.4)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Share Trump vote in 2016	51.0	0.108	0.142	0.158	0.151	0.064	0.132	0.045	0.009
	(18.4)	(0.005)	(0.005)	(0.006)	(0.007)	(0.005)	(0.007)	(0.007)	(0.009)
May 2020 state lockdown	43.1	-0.038	-0.034	-0.031	-0.042	0.004	-0.021	-0.042	-0.012
	(42.4)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
May 2020 local lockdown	4.7	-0.022	-0.018	-0.013	-0.001	0.007	-0.003	-0.029	-0.015
	(17.4)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)
May 2020 cum COVID deaths	1.9	-0.478	-0.104	-0.056	-0.032	-0.045	-0.340	-0.433	0.236
	(3.0)	(0.018)	(0.020)	(0.022)	(0.026)	(0.019)	(0.025)	(0.026)	(0.031)
Industry dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	--	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	--	53.0	62.6	61.4	66.7	82.6	78.4	-29.6	-15.8
Dep. var. standard deviation	--	(11.2)	(11.2)	(11.7)	(13.4)	(8.5)	(10.9)	(14.4)	(14.7)
R-squared	--	0.578	0.479	0.397	0.343	0.128	0.087	0.446	0.243

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Sample restricted to tracts for which a home location for at least one device observed in every period. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across Census tracts. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,124.

Appendix Table B8: Percent of Variance in Prevalence of On-Site Work (OSW) in Later Months Among Individuals Working On-Site in February 2020 or February 2019 Explained by Various Factors, Work Sites within 100 Yards of a School Dropped from Analysis, Sample Restricted to Tracts with 10 or More Devices

	OSW February 2020 and:				OSW February 2019 and:			
Explanatory variables	May 2020	August 2020	November 2020	March-April 2021	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:								
Age 25-64	-1.2	-0.9	-0.8	-0.7	0.7	0.1	-0.3	-0.1
Age 65 plus	0.1	-0.2	0.0	0.0	0.1	0.0	0.0	-0.1
White, non-Hispanic	-0.1	0.6	-0.6	0.8	-0.3	-0.5	0.0	0.3
College graduate	3.3	2.0	2.3	1.4	-0.7	0.0	0.5	0.2
ln(mean household income)	8.8	4.7	3.5	1.2	5.6	0.3	14.5	4.7
Share commute public trans.	0.4	0.9	1.0	0.9	0.1	0.2	0.0	0.1
Share commute 30+ mins.	1.5	1.6	1.3	0.8	0.2	0.1	1.6	1.7
Rural (yes/no)	0.8	0.5	0.6	0.4	0.0	0.0	0.6	0.2
Share Trump vote in 2016	7.4	9.9	9.5	8.5	2.4	4.9	1.3	0.2
May 2020 state lockdown	4.7	3.4	2.9	3.3	0.1	1.1	3.6	0.4
May 2020 local lockdown	0.3	0.2	0.1	0.0	0.0	0.0	0.3	0.1
May 2020 cum COVID deaths	4.1	0.6	0.3	0.1	-0.1	1.2	2.5	-0.3
Industry mix	19.5	16.5	14.8	12.9	3.3	1.1	12.5	9.9
Occupation mix	8.3	8.1	4.8	4.7	1.3	0.1	7.8	7.0
Residual	42.2	52.1	60.3	65.7	87.2	91.3	55.4	75.7
Dep. var. mean	53.0	62.6	61.4	66.7	82.6	78.4	-29.6	-15.8
Dep. var. standard deviation	(11.2)	(11.2)	(11.7)	(13.4)	(8.5)	(10.9)	(14.4)	(14.7)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.

Note: Unit of observation the Census tract. Dependent variable the share of February 2020 or February 2019 workers with on-site work activity in indicated month or, in final two columns, the difference between those shares in 2020 versus 2019. Algorithm for determining on-site work activity as described in text. Estimates weighted to account for device attrition and to match initial residence-to-employment flows across counties. Sample then restricted to Census tracts with 10 or more devices with a home location in every period. Explanatory variables from American Community Survey, except for rural, share Trump 2016 vote and COVID variables. See text for details. N=28,124 tracts.

C. The Variance Decomposition Methodology

In this appendix, we describe the methodology used in the main text to decompose the contributions of various explanatory factors to the overall variation in onsite work (OWS) (see Tables 4, 6, 7 and 8). The method we use was developed by Hottman, Redding and Weinstein (2016). We first describe how the method is implemented and then discuss the theoretical underpinnings of the empirical calculations.

1. Empirical Method

Consider a dependent variable Y (e.g., the OSW share in the tract) and independent variables $X_1, X_2, \dots, X_n$ (e.g., all of the covariates in Table 3).

- **<u>Step 1:</u>** Estimate an OLS regression of Y on a constant and $X_1, X_2, \dots, X_n$, denoting the OLS estimator of the coefficient on X_k as $\widehat{\beta_k}$. That is:

$$Y = \alpha + \sum_{k=1}^{n} \widehat{\beta_k} X_k \ + \ \hat{u}$$

  where α is a constant, the $\widehat{\beta_k}$ are the estimated coefficients on the X_k and $\hat{u}$ is a vector of residuals. This is what we do in Table 3 for our baseline results.

- **<u>Step 2:</u>** To measure how much each X_k contributes to the variance of Y, run an OLS regression of $\widehat{\beta_k} X_k$ on Y plus a constant:

$$\widehat{\beta_k} X_k = \alpha_k + \widehat{\delta_k} Y \ + \ \widehat{\epsilon_k}$$

  where α_k is the constant and $\widehat{\epsilon_k}$ is the residual. Note that $\widehat{\epsilon_k}$ and Y are orthogonal by the properties of OLS. The OLS coefficient $\widehat{\delta_k}$ is the contribution of X_k to the variance of Y.

- **<u>Step 3:</u>** To measure how much the variance of the residual $\hat{u}$ contributes to the variance of Y, run another OLS regression, again including a constant:

$$\hat{u} = \alpha_u + \widehat{\delta_u} Y \ + \ \widehat{\epsilon_u}$$

where α_u is the constant and $\widehat{\epsilon_u}$ is the residual. $\widehat{\epsilon_u}$ is orthogonal to Y by the properties of OLS. $\widehat{\delta_u}$ is the contribution of the residuals to the variance of Y.

2. Theoretical Underpinnings

This method yields a decomposition of variance such that the contribution of a covariate to the variance of a dependent variable is equal to its direct contribution plus half of all of the contributions attributable to the covariances of that covariate with other explanatory variables. To see that this property holds, note that the contribution of X_k to the variance of Y is defined as

$$V_k = var\left(\widehat{\beta_k} X_k\right) + \sum_{l:\, l=1, l\neq k}^{n} cov\left(\widehat{\beta_k} X_k, \widehat{\beta}_l X_l\right)$$

Observe that the second term on the righthand side of this expression allocates half of the contribution due to the explanatory variable covariances, since the full decomposition would include the second term times two. To establish the claimed result, we must show that $\frac{V_k}{var(Y)} = \widehat{\delta_k}$, with $\widehat{\delta_k}$ as defined in Section 1. First, note that $var\left(\widehat{\beta_k} X_k\right) = cov(\widehat{\beta_k} X_k, \widehat{\beta_k} X_k)$ so V_k can be simplified to

$$V_k = \sum_{l:\, l=1}^{n} cov\left(\widehat{\beta_k} X_k, \widehat{\beta}_l X_l\right)$$

The proof consists of three steps.

<u>Step 1:</u> We first simplify V_k . Replacing $\widehat{\beta_k} X_k = \alpha_k + \widehat{\delta_k} Y + \widehat{\epsilon_k}$ and $\widehat{\beta}_l X_l = \alpha_l + \widehat{\delta}_l Y + \widehat{\epsilon}_l$ for all values of *l* in V_k, we have

$$V_k = \sum_{l:\, l=1}^{n} cov\left(\alpha_k + \widehat{\delta_k} Y + \widehat{\epsilon_k}, \alpha_l + \widehat{\delta}_l Y + \widehat{\epsilon}_l\right)$$

Note that we have $cov\left(\alpha_k + \widehat{\delta_k} Y + \widehat{\epsilon_k}, \alpha_l + \widehat{\delta}_l Y + \widehat{\epsilon}_l\right) = cov\left(\widehat{\delta_k} Y, \widehat{\delta}_l Y\right) + cov(\widehat{\epsilon_k}, \widehat{\epsilon}_l)$ because

- α_k, α_l are constants and;
- Y and $\widehat{\epsilon_k}$ are independent and;
- Y and $\widehat{\epsilon}_l$ are independent.

Also, we can write $cov\left(\widehat{\delta_k}Y, \widehat{\delta_l}Y\right) = \widehat{\delta_k}\widehat{\delta_l}cov(Y,Y) = \widehat{\delta_k}\widehat{\delta_l}var(Y)$ because

- $\widehat{\delta_k}$, $\widehat{\delta_l}$ are constants and;
- $cov(Y,Y) = var(Y)$.

Therefore, we can simplify V_k as

$$V_k = \sum_{l:\, l=1}^{n}\left[\widehat{\delta_k}\widehat{\delta_l}var(Y) + cov(\widehat{\epsilon_k}, \widehat{\epsilon_l})\right] = \widehat{\delta_k}var(Y)\sum_{l:\, l=1}^{n}\widehat{\delta_l} + cov\left(\widehat{\epsilon_k}, \sum_{l=1}^{n}\widehat{\epsilon_l}\right)$$

where the second equation uses

$$\sum_{l:\, l=1}^{n} cov(\widehat{\epsilon_k}, \widehat{\epsilon_l}) = cov\left(\widehat{\epsilon_k}, \sum_{l=1}^{n}\widehat{\epsilon_l}\right)$$

<u>Step 2:</u> We use regression equations to further simplify V_k. Note that

$$Y = \alpha + \sum_{k=1}^{n}\widehat{\beta_k}X_k \; + \; \hat{u}$$

Replacing $\widehat{\beta_k}X_k$ with $\alpha_k + \widehat{\delta_k}Y \; + \; \widehat{\epsilon_k}$ for all k in the equation above, we have

$$Y = \alpha + \sum_{k=1}^{n}\left[\alpha_k + \widehat{\delta_k}Y \; + \; \widehat{\epsilon_k}\right] \; + \; \hat{u}$$

The above equation is equivalent to

$$\left(1 - \sum_{l=1}^{n}\widehat{\delta_l}\right)Y - \hat{u} = \sum_{l=1}^{n}\widehat{\epsilon_l} \; + \alpha + \sum_{l=1}^{n}\alpha_l$$

Hence,

$$cov\left(\widehat{\epsilon_k}, \sum_{l=1}^{n}\widehat{\epsilon_l}\right) = cov\left(\widehat{\epsilon_k}, \sum_{l=1}^{n}\widehat{\epsilon_l} + \alpha + \sum_{l=1}^{n}\alpha_l\right) = cov\left(\widehat{\epsilon_k}, \left(1 - \sum_{l=1}^{n}\widehat{\delta_l}\right)Y - \hat{u}\right)$$
$$= -cov(\widehat{\epsilon_k}, \hat{u})$$

The first equation holds because $\alpha + \sum_{l=1}^{n} \alpha_l$ is a constant. In the second equation, we replace $\sum_{k=1}^{n} \widehat{\epsilon_k} + \alpha + \sum_{k=1}^{n} \alpha_k$ by $\left(1 - \sum_{l=1}^{n} \widehat{\delta_l}\right)Y - \hat{u}$. The last equation comes from the fact that $cov(\widehat{\epsilon_k}, Y) = 0$ since Y and $\widehat{\epsilon_k}$ are independent.

We have $\widehat{\beta_k} X_k = \alpha_k + \widehat{\delta_k} Y + \widehat{\epsilon_k}$ and $cov\left(\widehat{\beta_k} X_k, \hat{u}\right) = 0$ because $\hat{u}$ is the residuals in the OLS regression of Y on $X_1, X_2, \ldots, X_n$. It follows that $cov\left(\alpha_k + \widehat{\delta_k} Y + \widehat{\epsilon_k}, \hat{u}\right) = 0$. Equivalently,

$$cov(\alpha_k, \hat{u}) + cov\left(\widehat{\delta_k} Y, \hat{u}\right) + cov(\widehat{\epsilon_k}, \hat{u}) = 0$$

or $-cov(\widehat{\epsilon_k}, \hat{u}) = \widehat{\delta_k} cov(Y, \hat{u})$ because $cov(\alpha_k, \hat{u}) = 0$ since α_k is a constant. Note that

$$-cov(\widehat{\epsilon_k}, \hat{u}) = \widehat{\delta_k} cov(Y, \hat{u}) = \widehat{\delta_k} cov\left(Y, \alpha_u + \widehat{\delta_u} Y + \widehat{\epsilon_u}\right) = \widehat{\delta_k}\widehat{\delta_u} var(Y)$$

where the second equation uses $\hat{u} = \alpha_u + \widehat{\delta_u} Y + \widehat{\epsilon_u}$ and the last equation uses $cov(Y, Y) = var(Y)$. Therefore, using the formula of V_k in step 1:

$$V_k = \widehat{\delta_k} var(Y) \sum_{l=1}^{n} \widehat{\delta_l} + cov\left(\widehat{\epsilon_k}, \sum_{l=1}^{n} \widehat{\epsilon_l}\right) = \widehat{\delta_k} var(Y) \sum_{l=1}^{n} \widehat{\delta_l} + \widehat{\delta_k}\widehat{\delta_u} var(Y)$$
$$= \widehat{\delta_k}\left(\widehat{\delta_u} + \sum_{l=1}^{n} \widehat{\delta_l}\right) var(Y)$$

Step 3: We will show that

$$\widehat{\delta_u} + \sum_{l=1}^{n} \widehat{\delta_l} = 1$$

so it follows that $V_k = \widehat{\delta_k} var(Y)$, which is what we want to show. From step 2, note that we have

$$\left(1-\sum_{l=1}^{n}\widehat{\delta_k}\right)Y-\hat{u}=\sum_{l=1}^{n}\widehat{\epsilon_k}+\sum_{l=1}^{n}\alpha_k$$

Using $\hat{u}=\alpha_u+\widehat{\delta_u}Y+\widehat{\epsilon_u}$, we get

$$\left(1-\widehat{\delta_u}-\sum_{l=1}^{n}\widehat{\delta_l}\right)Y=\widehat{\epsilon_u}+\sum_{l=1}^{n}\widehat{\epsilon_l}+\alpha+\alpha_u+\sum_{l=1}^{n}\alpha_l$$

Therefore,

$$cov\left(Y,\left(1-\widehat{\delta_u}-\sum_{l=1}^{n}\widehat{\delta_l}\right)Y\right)=cov\left(Y,\widehat{\epsilon_u}+\sum_{l=1}^{n}\widehat{\epsilon_l}+\alpha+\alpha_u+\sum_{l=1}^{n}\alpha_l\right)$$

Note that that RHS of the equation above is 0 because

- $\alpha, \alpha_u, \alpha_k$ are constants and;
- Y and $\widehat{\epsilon_u}$ are independent and;
- Y and $\widehat{\epsilon_k}$ are independent.

The LHS of the equation above is equal to

$$var(Y)\left(1-\widehat{\delta_u}-\sum_{l=1}^{n}\widehat{\delta_l}\right)$$

Since $var(Y)>0$, it is equal to 0 if and only if $1-\widehat{\delta_u}-\sum_{l=1}^{n}\widehat{\delta_l}=0$, which is what we need.