Using Mobile Device Activity Data to Study Local Variation in Onsite Work

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

February 25, 2025

Abstract: Mobile device location data suitable for a variety of research and commercial purposes have become increasingly available. We use these data to provide new evidence on the evolution of onsite work (OSW) following the pandemic. In one analysis, 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, then track those individuals' onsite work activity in May and August 2020. We then carry out a parallel analysis for 2019 and compare the persistence in OSW across the two time periods. In a second analysis, we analyze the ratio of measured OSW activity in September 2020, September 2021 and September 2022 to measured OSW activity in February 2020. In both analyses, we work with Census-tract-level estimates, documenting considerable cross-tract variation in OSW outcomes nationally, within states and cities, and even within counties. Observable characteristics such as industry, occupation, and household income in the tract account for much of the observed variation, but there is also substantial unexplained residual variation. Our results imply considerable heterogeneity in how the pandemic affected where the resident populations of U.S. neighborhoods spend their days, a finding that has significant implications for businesses, workers, and policymakers. We use this study of the evolution of OSW following the pandemic to evaluate the strengths and weaknesses of mobile device location data for tracking economic activity.

*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.

I. Introduction

The use of smart mobile devices with a variety of applications that provide information to users based on their location has become ubiquitous. Data aggregators have pooled the device location data generated by these applications, making it feasible to use the information for a variety of commercial and research purposes. An advantage of the aggregated mobile device location data is that the large sample sizes permit analysis at a very granular spatial level of aggregation. Moreover, the data are available at a higher frequency than the data from household surveys and potentially with a minimal lag, making it possible in principle to use them for real-time analysis. In this paper, we explore the use of mobile device location data to track changes in the pattern of onsite work following the pandemic. Our goal is both to characterize the geographic pattern of changes in onsite work over the period from 2019 through 2022 and to highlight the strengths and limitations of mobile device location data for this sort of analysis. Although there are significant challenges in using the mobile device location data to study the evolution of onsite work, they nonetheless can yield unique insights.

Our exploration makes use of mobile device location data from a repository created by researchers at the Maryland Transportation Institute (MTI) and Center for Advanced Transportation Technology Laboratory (CATT Lab). This infrastructure, based on location information collected through numerous common smart device applications, allows us to track the locations of millions of smart devices. The primary application of the MTI/CATT Lab data has been to monitor and model transportation activity (see, e.g., Zhang, Darzi, Pan et al., 2023). At the start of the pandemic, MTI researchers used the mobile device location data to create a real-time dashboard, described in Zhang, Darzi, Ghader et al. (2023), that included geographically disaggregated measures of social distancing.

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). For many workers, continued work from home at least on a hybrid basis appears to be the new normal (Barrero et al. 2021; Aksoy et al. 2022; Bick, Blandin and Mertens 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. We explore two related but distinct approaches to using the MTI/CATT device location data to study the evolution of onsite work and the geographical variation in that evolution, one that conditions on having onsite employment in February 2020 (the conditional analysis) and one that does not (the unconditional analysis).

For the conditional analysis, we begin with a sample of about 4.2 million people for whom we are able to algorithmically identify probable February 2020 home and onsite work locations. Then, we track these individuals forward in time to produce snapshots for May 2020 and August 2020, using the same algorithmic approach to identify the current home location and, if there was one, the current 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 the home and work locations of these individuals forward to May 2019 and August 2019. Our thresholds for identifying 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. For our baseline conditional analysis, we require that each tract have at least usable 10 devices in each studied month.

This conditional analysis reveals a large decline in onsite work among those who had been working onsite in February 2020, with many of these individuals no longer having an identifiable onsite work location in May 2020 or August 2020. Reassuringly for our interpretation of the conditional analysis results, our topline estimates of the decline in onsite work are very similar to conceptually similar estimates based on the Real-Time Population Survey (RPS) (Bick, Blandin and Mertens 2022).

Some of the decline we observe could be because the owners of the devices in our sample were no longer employed rather than because they had switched from onsite to remote work. Insofar as any declines in the share of February 2020 workers who would have ceased onsite work because of labor market turnover are well proxied by the corresponding 2019 changes in onsite work, however, the gap between the declines in onsite work that occurred in 2020 and the declines that occurred in 2019 should give us reasonable estimates of the effects of the pandemic on continuing workers' work locations. Between February and May, the average decline in onsite work among those working onsite in February was 29.6 percentage points larger in 2020 than in 2019; for the change from February to August, the average decline was 15.8 percentage points larger.

While other research has documented large post-pandemic increases in the prevalence of remote work, implying corresponding declines in the prevalence of onsite work, that research generally has had little to say about differences across geographic areas in those changes. The results from our conditional analysis show not only that the overall declines in onsite work among previously onsite workers in the early months of the pandemic were much larger than the

declines over the same months of 2019, but also that there was also enormous spatial variation in these outcomes. For the 10th percentile (employment weighted) Census tract (measured by place of residence), onsite work among those who had worked onsite in February declined between February and May by 48.4 percentage points more in 2020 than in 2019; in the 90th percentile tract, this difference was just 11.5 percentage points. The corresponding gap for August was also large. The dispersion in the 2020 minus 2019 difference across tracts is mostly attributable to differences in the 2020 tract-level changes; the correlation between the raw 2020 change and 2020-net-of-2019 change in Census tract estimates is 0.93 for May and 0.81 for August.

For the unconditional analysis, we use the device data to compute the percentages of individuals for whom a home location can be identified in repeated cross sections created from device observations contained in the MTI/CATT data repository who also have an identifiable work location. We construct these measures for February 2020, September 2020, September 2021 and September 2022. The number of devices underlying the monthly estimates in these repeated cross sections is much larger than the number available for our conditional analysis, in the tens of millions. This means that, even requiring a minimum of 100 devices for a tract to be included, our tract-level analysis sample is much larger.

A challenge with the unconditional analysis is that, even in February 2020 when remote work was quite uncommon, the overall share of devices with a home location for which we also are able to identify a work location is significantly smaller than the employment rate as measured in Bureau of Labor Statistics data. Among other possible factors, employed device users may not regularly use the applications that generate the location information that populates the MTI/CATT repository while they are at work. The measure we use to track the evolution of onsite work among residents of a tract is the ratio of the share of devices with identifiable onsite

work (OSW) in a later month (September 2020, September 2021 or September 2022) to the share in February 2020. To the extent that the tract-level shares of employed people with an onsite work location who use location-generating applications regularly throughout the day have been relatively stable over time and changes in the prevalence of onsite work have been large relative to changes in the tract-level employment rate, however, changes in the share of devices with a home location for which a work location can be identified will be informative about how the prevalence of onsite work is changing. It is reassuring that, as we show below, similar calculations using data from the RPS and the American Community Survey (ACS) that track the share of onsite work more directly yield top-side results that are very consistent with the results we obtain using the mobile device data.

In our data, the ratio of measured OSW in September 2020 to measured OSW in February 2020 has an (employment-weighted) mean of 0.75, which we interpret as implying a roughly 25 percentage-point decrease in the prevalence of onsite work (or, equivalently, a 25 percentage-point increase in the prevalence of remote work). By September 2022, the corresponding ratio had recovered to a mean of 0.91. As with the conditional analysis, we find considerable cross-tract variation, in this case persisting through 2022. A key take-away from both the conditional and unconditional analyses is that the pandemic led to large neighborhood-level differences in the evolution of onsite work. This is variation that cannot readily be observed or analyzed in data from alternative sources and especially not on a timely basis.

Much of the variation in both the conditional and unconditional measures of changes in OSW can be explained by tract level characteristics measured in external data sources. For example, we are able to account for about 58% of the May 2020 cross-tract variation in our conditional onsite work measure, with the industry and occupation mix of employment among

residents of the tract playing especially important roles. The results for a first-difference outcome measure, constructed as the difference between the conditional OSW measure for May 2020 and that for May 2019, are in most respects quite similar, suggesting that differences in the response to the pandemic rather than longer-standing differences in employment patterns are most important in accounting for the cross-tract variation in the May 2020 numbers. Observable covariates also account for a significant fraction of the variation in the unconditional ratios, with industry and occupation again the covariates that generally have the greatest explanatory power.

While we view these results as interesting and informative, our analysis also highlights the challenges of working with mobile device data to learn about changes in work location patterns. One challenge is that device use is not ubiquitous, meaning that some onsite work is not observed. This requires the analyst to develop metrics that take this into account. In the conditional analysis, we do this by restricting our sample to people we are able to identify as onsite workers in the base period (i.e., to people who have demonstrated themselves to be regular users of location-generating applications) and tracking their onsite work behavior over time. In the unconditional analysis, we focus on the ratios of measures for different time periods, reasoning that if the share of onsite work captured in tracts is relatively stable and employment rates have not changed too greatly, those ratio measures should be informative. A further potential complication is that the landscape for collecting and sharing device location data seems to be changing, meaning that these data may be less available going forward. Future analysts will need to consider whether and how changes to the mobile device data landscape may affect the conclusions that can be drawn from these data.

The paper proceeds as follows. Section II provides a more extensive review of the alternative sources of data for learning about the prevalence of onsite work. Section III describes

the data and measurement approaches that underlie our analysis. Section IV presents some basic facts about the spatial variation in the changes in onsite work during the post-pandemic period. We analyze the factors that account for this spatial variation in Section V. Section VI compares our findings to results from more conventional data sources and briefly describes the robustness and sensitivity analyses we have carried out. In Section VII, we conclude with a discussion of our key findings, their implications, and the prospects for future work using mobile device location data to study shocks to the prevalence of onsite work.

II. Alternative Sources of Data for Measuring Onsite Work

While remote work had begun to rise even before 2020, it jumped dramatically during the pandemic. Several approaches have been used to predict and track remote versus onsite work activity across individuals and locations.

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% 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 household surveys have provided valuable information on trends in work-from-home (WFH) during the post-pandemic period. 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). Over time, 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. Beginning in October 2022, the telework questions on the CPS were revised to collect total hours of telework during the survey reference week, with separate questions about telework prior to the pandemic. 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, making it possible to estimate both the level and the change in WFH over the period the survey was fielded (Bick, Blandin and Mertens 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).

Lastly, even before the pandemic, the American Community Survey (ACS) included questions about commuting patterns that can be used to measure remote work and onsite work activity (Burrows, Burd and McKenzie 2023). The key ACS question is “How did this person usually get to work LAST WEEK?” The response options “worked from home” should capture people who were mainly or exclusively remote. A strength of the ACS is that its relatively large sample size supports more geographically disaggregated estimates than the other surveys just

mentioned. Estimates for large counties can be computed at an annual frequency. Estimates for smaller geographies including tracts, however, are available only for five-year periods, meaning that the ACS estimates for those levels of disaggregation will be slow to capture changes in behavior.

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 following the start of the pandemic (in the CPS, ACS and RPS) and has been substantially more prevalent (in the same surveys plus the SWAA) among workers with higher levels of education. Estimates from the ACS, 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 the broad patterns of WFH, because of sample size limitations, surveys are generally not well suited for the production of geographically disaggregated statistics, especially at a monthly or annual frequency.

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 covers 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 such source of information (see, e.g., Sehra et al. 2020, Jacobsen and Jacobsen 2020, Mendolia, Stavrunova and Yerokhin 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). They link devices to the neighborhoods where their owners live and conclude that, in the first months of the pandemic, travel to work fell more in higher-income neighborhoods. Limitations of the Jay et al. (2020) measure of travel to work are that, because the measure is based on activity for a single

day, visiting a friend, running errands or other non-work trips might show up as travel to work. It also will miss work that occurs outside of daytime hours.

Our approach is closest in spirit to the Jay et al. (2020) study using device-level 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. Our conditional analysis takes advantage of the fact that we are able to follow devices longitudinally, which allows us to see how the probability of OSW changed in the immediate aftermath of the pandemic for individuals with clearly identified pre-existing onsite work attachments. In our unconditional analysis, for which the samples are much larger, we examine how the share of devices with a home location that also have an onsite work location changed over a longer period. A focus of both our conditional and unconditional analyses is to better understand the considerable cross-tract variation in the evolution of onsite work following the pandemic. To that end, we link devices to the Census tract where their owners live and use characteristics of the owners' neighborhoods to explain the variation in our OSW outcome variables.

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 underlying data utilized in this analysis were gathered by one of the leading U.S. location-based services data aggregators and processed for analysis by MTI and CATT Lab staff. Processing included essential data cleaning such as removing entries with invalid values and deleting duplicate observations (Zhang, Darzi, Pan et al. 2023). 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, though not all of these devices were observed with sufficient frequency to be usable for our purposes.

We adopt the general approach to identifying home and work locations described by Pan et al. (2023). In essence, it involves clustering device sightings based on their latitude and longitude, then sorting the data to identify places that are plausible home and work locations. For our conditional analysis, we experimented with different criteria for selecting home and work locations before choosing the specific rules discussed below. For the unconditional analysis, we made use of data files created for a different project that identified home and work locations using the same general framework but a somewhat different set of criteria.[1] Appendix A describes both approaches in greater detail.

Conditional Analysis Data Infrastructure

For our conditional analysis, we constructed a sample of mobile devices for which we were able to identify both a February 2020 home location and a February 2020 onsite work location. Because mobile devices typically are not shared, we refer to these interchangeably as device or individual locations. In a preliminary processing step, we dropped devices that were not observed at least 100 times during the month. The February 2020 home location is the location where the device was observed most frequently during the month, provided it was 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 was observed for a minimum of 60

[1] Processing the mobile device location data is computationally intensive. The work to develop the data file used in the conditional analysis we conducted initially was supported with funding from the National Science Foundation. The fact that we could build on work already done for another purpose made it possible to add the unconditional analysis.

distinct hours spread across at least two different weeks. Note that this definition should capture hybrid work—working a few days a week onsite and the rest of the time at home or in another location—along with full-time onsite work provided the person works a sufficient number of onsite days. For the conditional analysis, 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.[2] Additional steps were taken to ensure that the home and work locations did 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 then searched for records for those same devices in May 2020 and August 2020. In each of those months, our sample consists of the devices for which we were able to identify a home location using the criteria described above. Then, for each of these devices, we attempted 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 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]

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

[3] In Abraham et al. (2024) (an earlier working paper version of the current paper) we also evaluated conditional OSW activity for November 2020 and March/April 2021. The main reason for not including those results here is that we are interested primarily in results based on the differences between OSW persistence in 2020 and OSW persistence over the corresponding period in 2019. Our 2019 sample is not large enough to produce reliable results for November 2019 or later months and, even if it were, the interval from February 2019 to March/April 2020 would be problematic for establishing a benchmark.

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 conditional analysis. The number of usable devices fell to about 2.8 million in May 2020 and about 2.1 million in August 2020. 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. This fell to about 1.3 million in May 2019 and about 0.8 million in August 2020.[4]

One impediment to identifying work locations in the months that followed February 2020 is that, for the devices we are able to track longitudinally, the average number of observations per device drops off substantially. In the May 2020 sample, for example, which consists of devices with a home and work location in February 2020 and a home location in May 2020, the average number of observations per month was 579 in February and 381 in May. The declines for the August 2020 sample were similar. For this reason, we lower the thresholds used to define the work location in the later months. For example, for a device with 600 observed hours in February 2020 and 300 observed hours in the later month, we cut the hours threshold for identifying the device's work location in the later month in half. [5] Although we do not observe any drop-off in the average number of observed hours in the end month for the May 2019 and August 2019 samples, for consistency we applied the same procedures to those samples.[6]

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

[5] 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.

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

Another feature of our data is that the rate of attrition from the samples varies across geography. In the 2020 data, we address this by constructing separate block-group-level attrition weights for the May 2020 and August 2020 samples. The first step in constructing these weights was to 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 and zero otherwise on block-group characteristics of the device's February 2020 residence location. The block-group characteristics are measures based on data from the 2015-2019 ACS.[7] The device's attrition weight then is computed 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.[8]

Most of the analyses we report use data aggregated to the tract level. For that purpose, we restrict our baseline sample to tracts with at least 10 devices in every period for which we have constructed estimates.[9] Because labor market turnover implies that, even in normal times, a substantial fraction of workers observed in February of a given year will not be working in subsequent months, our main focus in the conditional analysis is on the differences between the

[7] These characteristics are 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.

[8] We also would like our sample for the conditional analysis 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. The data appendix provides details of this procedure. All of the within-tract measures based on the device data used in our analyses of conditional OSW activity 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.

[9] This includes having at least 10 usable devices in November 2020 and March/April 2021, for which we produced estimates included in the earlier working paper version of the current paper.

patterns of OSW persistence observed in 2020 and those observed in 2019. For example, for understanding the impact of the pandemic on OSW in May 2020, our preferred estimates are based on the difference between the share of workers in a tract persisting in OSW in May 2020 among those who had been working onsite in February 2020 and the share of workers persisting in OSW in May 2019 among those who had been working onsite in February 2019.

Unconditional Analysis Data Infrastructure

Analyzing the evolution of OSW by starting with a sample of people we can feel confident were working onsite in a baseline period has some clear advantages. Because we need to be able to follow the same devices over time and there is significant sample attrition, however, the conditional analysis is most suitable for looking at relatively short-term changes. For that reason, we also have carried out an unconditional analysis based on a series of cross-sectional samples that we use to estimate the share of devices with a home location for which we also can identify an onsite work location. Using this approach, we look at how OSW changed from February 2020 through September 2022.

The methodology for this alternative approach aligns with the framework used in the conditional analysis though the specific algorithm used to identify home and work locations applies somewhat different rules. The home location in a month is still the location where the device is observed for the largest number of hours. Now, however, the requirement is that it be observed on at least 3 days and on at least half of the observed days during the month, for a minimum of 2 hours on each of those days. A work location is identified if there is a location other than the home location that satisfies the same restrictions based on weekday (non-holiday Mondays through Fridays) observations. Both the home and work locations are identified at

geohash level 7, corresponding to an area of no more than 152.9 meters by 152.4 meters, rather than geohash level 8. Again, we took additional steps to ensure that the home and work locations did not overlap. As can be seen in Appendix Table B3, the sample of devices underlying the unconditional analysis is much larger than that for the conditional analysis. The February 2020 sample consists of about 52.2 million devices that have an identified home location; of these, 21 million have an identified work location. The number of devices with an identified home location fluctuates somewhat across months, but never falls below about 35.6 million and reaches 77.8 million in September 2022. An advantage of the large sample sizes available for the unconditional analysis is that, even though we restrict the sample of tracts to those with more than 100 devices in each of the relevant months, it includes many more tracts (62,829) than the baseline conditional analysis sample for which we required only 10 devices per tract (28,125).

Although the criteria used to identify home and work locations for the unconditional analysis require fewer observed days and hours, as shown in Appendix Table B4, we still observe a substantial number of days and hours for the included devices. In the February 2020 sample, for example, devices are observed for an average of 21 days at home and 10 days at work. In the same sample, included devices average 187 home hours and 51 weekday work hours.

For the unconditional analysis, the measure of onsite work is the share of devices with a home location that also have an onsite work location. The shares of devices with a home location for which we observe a work location are considerably below the employment rates for the same months. In February 2020, for example, we observe a work location for 40.6% of devices with a home location, whereas the (non-seasonally-adjusted) employment-to-population ratio for the population age 16 plus was 60.9%. An important reason the two do not agree is that onsite

workers don't necessarily use the applications that record their locations while on the job.[10] Further, the share of workers whose locations are recorded may differ across tracts. We are more interested, however, in the *change* in OSW and how that has varied across geography than in its level. For this reason, we focus on the *ratio* of tract-level onsite work in the months of September 2020, September 2021 and September 2022 to tract-level onsite work in February 2020 rather than on the raw tract-level estimates. The ratio measure has the property that it nets out any time-invariant proportional tract-level factors that might affect the device-based measurements, as the same factor would appear in both the numerator and the denominator of the tract-level ratio.

It is true, of course, that even if the device data captured 100 percent of onsite work, the ratio of the onsite work share in the later months to the onsite work share in February 2020 could be affected both by changes in the overall employment-to-population ratio and by changes in the propensity for those who are employed to work onsite. Our interpretation of changes in the spatial variation in the ratio as indicative of spatial variation in the change in the probability that people who are employed work onsite rests on the assumption that the spatial variation in the change in the employment-to-population ratio is small relative to the spatial variation in the change in the probability that employed individuals work onsite. Other evidence suggests that relative employment to population ratios across tracts do indeed change slowly so that, given the

[10] Ownership of smartphones among children and teenagers under age 16 may be another reason they don't agree. Even if everyone aged 10 to 15 owned a smartphone and was recorded in our data as having a home location but not a work location, however, that would account for only about 5 percentage points of the 20-percentage-point February 2020 discrepancy.

sharp increase in remote work during the pandemic, this would appear to be a reasonable assumption. [11]

Location Characteristics

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 for use in both our conditional and unconditional regression analyses that capture community characteristics found in other research to be associated with social distancing and/or the prevalence of remote work. Several of our tract-level explanatory variables come from the 2015-2019 ACS. In addition to the variables used for our attrition analysis (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, all 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 LEHD Origin-Destination Employment Statistics (LODES) data. The Department of Agriculture is our source for a dummy variable indicating whether a tract is rural.[12] To measure political preferences among tract residents, we use Donald Trump's

[11] To illustrate, using ACS data, we computed tract-level employment-to-population ratios for the non-overlapping 2010-14 and 2015-19 periods. Then, we estimated the correlation across tracts in the employment-to-population ratios for these two periods. The employment-weighted correlation across these five-year intervals is 0.85. Assuming an AR(1) process, this implies an average annual correlation of the employment-to-population ratios across tracts of 0.97.

[12] 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.

share of the 2016 presidential vote. The source for this measure is the 2016 Precinct-Level 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 initial 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 2015-2019 ACS.

IV. Basic Facts

We begin our analysis by reporting some basic facts about the geographic distribution of changes in OSW during the pandemic.

Conditional Analysis

Starting with our conditional analysis, Figure 1 maps 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.[13] Darker shading in the

[13] 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

figure is associated with lower OSW rates. Some 40.5% of counties, accounting for 82.4% of employment, have OSW shares in May 2020 among those who had been working onsite in February 2020 that are below 60%. Further analysis of the underlying data shows that these rates are highest in rural counties and, among counties in urban areas, fall monotonically with CBSA size. The shares of workers continuing in OSW 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 working onsite in May 2019. More than 90% of counties, accounting for more than 99% of employment, had May 2019 OSW rates for those who had been working onsite in February 2019 of at least 70%. Indeed, more than 60% of counties accounting for about two-thirds of employment had OSW rates among those who had been working onsite in February 2019 of 80% 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. For this reason, as already discussed, we are especially interested in the *changes* in the conditional OSW rates associated with the pandemic. Using the same underlying data as were used to create the previous two figures, Figure 3 displays the 2020 minus 2019 differences in the share of workers working onsite in February who were also working onsite in May. Most counties experienced a decline in OSW persistence from 2019 to

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.

2020 and the regional patterns are broadly similar to those in Figure 1. The parts of the country with the smallest share of continuing OSW workers in May 2020 also tend to be places where the share continuing in OSW fell the most between May 2019 and May 2020. The employment-weighted county-level correlation of the OSW persistence rates mapped in Figure 1 and the difference measures mapped in Figure 3 is 0.93.

Appendix Figures B1, B2 and B3 present maps corresponding to Figures 1, 2 and 3 but for August rather than May. Comparing Appendix Figure B1 with Figure 1, OSW among those who had been working onsite in February 2020 were somewhat higher in August 2020 than in May 2020 (apparent in the lighter overall shading in the August map). Still, it was generally lower than OSW in August 2019 for February 2019 onsite workers shown in Appendix Figure B2. Appendix Figure B3 shows the 2020 minus 2019 difference in the share of February OSW workers who also were working onsite in August. There is a strong weighted county-level correlation between the estimates underlying Appendix Figure B1 and those underlying Appendix Figure B3 (0.81), suggesting that, as for May 2020, the August 2020 differences across counties were mainly attributable to differing responses to the pandemic rather than to differences in normal labor market dynamics.

To further illustrate the very different early impact of the pandemic on onsite work in different parts of the country, Figure 4 shows the May 2020 minus May 2019 and August 2020 and August 2019 difference in the OSW continuation percentages for two cities, Houston and San Francisco. Houston is a city where these differences were relatively small (represented by lighter shading) and San Francisco is a city where they were relatively large (represented by darker shading). 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%. In contrast, all five counties in the San Francisco-Oakland-Berkeley, CA CBSA experienced OSW declines in excess of 30% and two of the five counties experienced declines in excess of 40%. The maps displaying the August differences also show a considerably greater decline in OSW persistence in San Francisco than in Houston.

While striking, even the county-level displays mask significant tract-level variation. Figure 5 summarizes the cross-tract variation in OSW percentages among previously on-site workers in the form of a histogram showing data for May 2019 (for those working onsite in February 2019), May 2020 (for those working onsite in February 2020), and the difference between the two. 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 for the Figure 5 histograms along with corresponding summary statistics for August.[14] Of those who were working onsite in February 2019 in the included tracts, 82.6% were still doing so in May 2019. Although normal labor market dynamics mean this percentage is not 100%, it is still considerably higher than the 53.0% of people working onsite in February 2020 who were still doing so in May 2020. The histogram showing the change from May 2019 to May 2020 is centered around minus 30 percentage points. There is also, however, considerable variation across tracts, with the 10th percentile tract having a change of minus 48.4 percentage points and the 90th percentile tract a change of minus 11.5 percentage points. OSW work conditional on February 2020 employment remained lower in August 2020 than one would have expected based on the 2019 patterns and, again, there is considerable cross-tract variation.

[14] Distributions across tracts for August are shown in Appendix Figure B4.

Unconditional Analysis

Our unconditional analysis provides useful insights regarding the changes in the pattern of onsite work activity through September 2022. Figure 6 depicts the ratio of the county-level OSW percentages based on the device data for September 2022 to the same measure in February 2020. As for the conditional analysis, we display counties rather than tracts to make the figure more readable.[15] The average (employment-weighted) county has a ratio of 0.90, which we interpret to mean that the share of workers who were (primarily) onsite was about 90% as large in September 2022 as it had been in February 2020. This is considerably higher than the analogous ratio of 0.72 for September 2020. Perhaps equally important is that spatial variation in these ratios remains high through September 2022; even that far removed from the onset of the pandemic, the 10th percentile county has a ratio of 0.83 while the 90th percentile county has a ratio of 0.97.

As a further illustration, the evolution of these ratios for the counties in the Houston and San Francisco CBSAs is depicted in Figure 7. In both of these cities, the ratio increased from September 2020 to September 2022, captured by progressively lighter shading in the figure. It also is evident that, across the three months displayed, the San Francisco counties have consistently lower ratios than the Houston counties. For Houston, the (employment-weighted) ratio to the February 2020 baseline increases from 0.71 in September 2020 to 0.88 in September 2022. For San Francisco, the ratio increased from 0.51 to 0.80.

As was true for the conditional analysis, the county-level maps are instructive but mask considerable variation across tracts. Figure 8 displays the employment-weighted histograms of

[15] Similar county-level figures for September 2020 and September 2021 are shown in Appendix Figures B5 and B6.

tract-level OSW ratios. Table 2 reports summary statistics for the data underlying the Figure 8 histograms. The upward shift in the distribution of the ratios from September 2020 to September 2021 to September 2022 is evident. Still, considerable tract level dispersion persists through September 2022. As of September 2020, the 90th percentile tract had an OSW ratio of 0.96 and the 10th percentile tract a ratio of 0.54, for a 90-10 differential of 0.42. The 90-10 differential had declined somewhat by September 2022 but only to 0.30. The 90-10 differential in the OSW ratios across tracts is also large and persistent within major cities; as of September 2022, it was 0.28 within San Francisco and 0.24 within Houston, of the same order of magnitude in both cases as the national 90-10 differential.

The enormous local variation in OSW activity across tracts within cities is a core finding from both the conditional and the unconditional analysis. Others have noted the spatial variation in OSW. For example, consistent with our results, Hansen et al. (2023) find that, as of 2022, San Francisco was a city with a relatively large share of vacancy postings offering remote work while the share of remote vacancy posting was much lower in Houston. Hansen et al. (2023) also note that their cross-city patterns are similar to the patterns of change in WFH that emerge in American Community Survey (ACS) data between 2019 and 2021.[16] Our results imply, however, that a city-level analysis misses substantial cross-tract variation. As we will show below, the same is true even within counties.

[16] Decker and Haltiwanger (2023) find a much weaker correlation between the ACS data and the Hansen et al. (2023) 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.

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 geographical variation in OSW. Our analysis consists of a series of regressions that relate tract-level OSW measures based on either the conditional or the unconditional measurement approach to observable characteristics of the tracts.

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

In these regressions, Y_{it} is the conditional or unconditional OSW measure for tract i in month t and X_i is a set of tract characteristics. All of the regressions are weighted by tract employment. The covariates included in the regressions are basic demographics plus the other tract characteristics described in Section III that previous research has suggested may be associated with onsite work.

Conditional Analysis

The outcomes of interest for the conditional regression analysis are the May 2020 and August 2020 measures of OSW for those we identify as working onsite in February 2020; the corresponding measures for 2019; and, the outcomes that will be our primary focus, the differences between the 2020 and 2019 measures.

Regression Coefficients. Estimated coefficients for the covariates other than industry and occupation from the conditional OSW regressions are reported in Table 3; the industry and occupation coefficients from these regressions are shown in Appendix Tables B5 and B6. There are systematic relationships between many of the covariates and the likelihood that a person working onsite in February 2020 was still working onsite in May or August of 2020. Overall, as shown in the first two columns of Table 3, observable characteristics account for 57.8% of the

cross-tract variation in conditional May 2020 OSW and 47.9% of the cross-tract variation in conditional August 2020 OSW. In contrast, these same covariates account for very little of the variation in conditional OSW across tracts in 2019—just 12.9% in May 2019 and 8.7% in August 2019.

A useful way to frame the differences between the 2020 and the 2019 regressions is to think of the 2020 coefficients as reflecting patterns during the COVID period and the 2019 coefficients as reflecting patterns during a normal period. The difference between the two estimated coefficients then is arguably a better estimate of the COVID impact than the coefficients from the May 2020 regressions. This difference is reported directly in the final two columns of Table 3 (and, for industry and occupation, in the final two columns of Appendix Tables B5 and B6). These columns summarize the results of regressions in which the dependent variable is the change in the OSW rate for previously-OSW workers from February 2020 to either May 2020 or August 2020 minus the corresponding change for May 2019 or August 2019. Many of the same factors that helped to explain differences in the conditional level of OSW in May and August 2020 also are important in the change regressions, but there are some differences.

In both the level and change regressions, for example, higher mean household income has a negative effect on the outcome variable, but the effect is larger in the change regressions than in the level regressions. This reflects the fact 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.

In other cases, the COVID effect measured in the change regressions is smaller than in the level regressions. The share of votes in the tract for Trump in the 2016 election, for example, has a sizeable and positive coefficient both in the 2020 level regressions and the 2019 level regressions. Taken on their own, the 2020 regression coefficients would seem to imply that political attitudes were an important source of variation in the COVID effect on onsite work, as argued in some previous research (see, e.g., Allcott et al. 2020). The fact that the coefficient on the 2016 Trump vote share is smaller in the change regressions, however, suggests that, in the 2020 level regressions, the Trump 2016 vote is at least partially standing in for something else.

Differences between the level and change regression coefficients also can be observed for some of the industry and occupation coefficients. As one example, relative to the omitted industry group (finance and insurance), the share of workers in a tract employed in construction has a positive coefficient in the 2020 level regression but a negative coefficient in the 2019 level regression, making its coefficient in the change regression much larger than its effect in the 2020 level regression. In normal times, the jobs of construction workers are less stable than those of workers in finance and insurance, but during COVID construction workers were much less likely to be able to work remotely and therefore more likely to be measured as having an onsite job.

Variance Decomposition. Although the coefficient estimates reported in Table 3 and the two appendix tables 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, they do not translate simply into the share of the variation in the onsite work percentages that each explains. To better understand 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 (2024). 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.[17] By construction, this method yields a decomposition in which the terms, including the residual, sum to one.

Table 4 reports the results of this variance decomposition for the Table 3 models. In May 2020, the largest contributors to the 57.8% of the variance accounted for by the covariates are industry mix, which accounts for 19.5%, and occupation mix, which accounts for 8.2%. Other large contributors are log(mean household income) at 8.9% and the 2016 share of votes for Trump at 7.4%.[18] Consistent with the R-squareds reported in Table 3, the contribution of the covariates to explaining the variance in our conditional OSW measure is somewhat lower in August 2020, but the same four factors (industry mix, occupation mix, mean household income and Trump voters) continue to top the list in terms of their contribution to explained variance. In contrast, the covariates explain relatively little of the variance in our conditional OSW measure in either May 2019 or August 2019.

[17] 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 and August 2020. In 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.

[18] 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.

Most of the same covariates that account for significant portions of the variance in our conditional OSW measure 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, 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 conditional 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 conditional OSW change, though as for the level models, the explained portion of the first-difference variance is lower in August than in May.

One way to assess how well the covariates account for observed spatial variation in conditional OSW is to compare the actual and predicted rates at the state and city level. The estimated models do a good job of accounting for the broad regional variation in OSW levels and OSW changes shown in the maps provided earlier. Some selected illustrative results are reported in appendix Table B7. For Texas and California, there is a close correspondence between the actual and predicted OSW patterns for levels and changes. The same is true at the city level for Houston and San Francisco. The result that our models do a good job of predicting OSW rates at the state and city level holds more generally.

It is also of interest to examine the variation in OSW outcomes within these broader geographies. By way of illustration, we have carried out an exercise like that in Table 4 that reports on a decomposition of the cross-tract variation in Houston and San Francisco (see Appendix Table B8). Both metropolitan areas have a large cross-tract dispersion in the changes from May 2019 to May 2020 and in the changes from August 2019 to August 2020. Using the

coefficients estimated for the nation as a whole, the share of the dispersion in the May 2019 to May 2020 change accounted for by covariates is 48.8% in Houston and 31.7% in San Francisco. In both cities, consistent with the national results, income, industry and occupation mix are the dominant explanatory factors. Covariates account for less of the August 2019 to August 2020 changes but it is the same covariates that are especially important. Even within cities, then, there is considerable cross-tract variation, much but not all of which can be accounted for by the covariates in our regression models.

What Does a Tract-Level Analysis Add? One of the key contributions of our analysis is that we are able to examine differences in the evolution of onsite work at the level of the neighborhood (i.e., tract) as opposed to differences across broader geographic areas (state and county). One might reasonably ask whether drilling down to the neighborhood level actually makes a significant difference or whether the systematic variation we see is mostly attributable to differences across broader geographic areas. To answer this question, we have estimated the specifications summarized in Tables 3 and 4 with just state and county fixed effects and also with these effects added to the existing models. Table 5 summarizes our findings; as it shows, the covariates included in our regression models do more to explain the variation in the conditional outcomes across tracts than do state effects or even county effects.

On their own, state fixed effects explain only 15.8% of the change between May 2020 and May 2019. This compares to the 44.6% explained by our list of covariates. Adding state fixed effects to the covariate model raises the explained variance by only 1.4 percentage points. In the full model, just 6.8% of the cross-tract variance—or about 15% of the explained variance—is absorbed by the state fixed effects, and at least a portion 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.

County fixed effects capture considerably more of the conditional 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 36.5% of the variation in the May 2020 minus May 2019 difference in tract-level OSW, compared to 44.6% 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 6.7 percentage points. In the combined model, county fixed effects absorb 18.1% 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.

Summarizing these results, neither state nor county fixed effects capture the cross-tract variation in the conditional OSW outcomes we would like to understand. On their own, they explain less of the variance in tract-level OSW shares than our covariates; they add little explanatory power when added to models that include the covariates; and, in models that include both covariates and the state fixed effects, they absorb only a modest portion of the cross-tract OSW variance. There also is a good deal of cross-tract variance that is not explained either by our covariates or by the state or county fixed effects.

Unconditional Analysis

As explained earlier, the outcome examined in our unconditional analyses is a measure of onsite work activity derived from repeated cross sections. Our discussion of the factors that

account for the considerable spatial variation in this measure parallels the discussion of the conditional OSW results.

Regression Coefficients and Variance Decomposition. The regression results summarized in Tables 6 and 7 are the unconditional analysis analogues to the regression results reported in Tables 3 and 4 for the conditional analysis.[19] Observable covariates account for 59.6% of the cross-tract variation in the ratio of measured OSW activity in September 2020 to that in February 2020. Similar to the results for the conditional analysis, the covariates that make the largest contribution to accounting for the variation in the September 2020 ratio are industry and occupation mix, household income, and the Trump vote share in 2016. Given that the conditional and unconditional analyses take quite different approaches, with the former capturing changes in the persistence in onsite work among previously onsite workers between 2019 and 2020 and the latter employing data for repeated cross sections to construct a ratio measure, the broad similarity in the findings regarding the factors that affected changes in the prevalence of onsite work early in the pandemic is notable.

The cross-tract variation in the unconditional ratio measure is almost as large in September 2021 as in September 2020, but the regression covariates account for much less of that variation (21.4% versus 59.6%). Industry mix and household income in the tract continue to have notable explanatory power, but the contributions of the other factors in the model are generally much smaller than in the September 2020 model.[20] By September 2022, although the cross-tract

[19] For the ratio coefficient estimates in Table 6, the dependent variable has been multiplied by 100. This does not influence the variance decomposition in Table 7. Industry and occupation estimates for the unconditional analysis are reported in appendix Tables B9 and B10.

[20] The one notable exception to this generalization is the share White, non-Hispanic in the tract, which plays a larger role in the September 2021 model and also the September 2022 model than in the September 2020 model.

variation in the OSW ratio measure remains large, covariates account for only 15.1% of the cross-tract dispersion.

As with the conditional analysis, the regression models do a good job of accounting for the differences across states and cities in our unconditional outcome measures. This is illustrated for some selected examples in Table B11, which shows the actual values of the unconditional OSW measure for two states (Texas and California) and two cities (Houston and San Francisco) together with predicted values based on the regression models reported in Table 6 for the same jurisdictions. Consistent with the conditional analysis results, the correspondence between the actual and the predicted measures for the September 2020 ratios is very close. Some of the gaps between the actual and predicted September 2021 ratios are wider, but none are very large and the September 2022 ratios are again very close.

Even within cities, as with the conditional measures, however, there is considerable cross-tract variation in the unconditional ratio measures. This can be seen in Appendix Table B12. Using the estimated coefficients from the Table 6 regression, covariates account for less of the within-city early-pandemic cross-tract variation than they do of the national variation, but they nonetheless have substantial explanatory power, accounting for 42.1% of the cross-tract variation in the ratio of September 2020 OSW to February 2020 OSW in Houston and 38.3% of the variation in that ratio in San Francisco. There are some differences in the magnitudes of the contributions of the different explanatory variables across the two cities, with industry mix more important in San Francisco and the Trump vote share more important in Houston. In both cities, occupation mix and household income are quantitatively important in explaining the September 2020 ratio. By 2022, much as for the nation as a whole, although cross-tract dispersion within both Houston and San Francisco remained large, our explanatory variables are less successful in

explaining that variation. They account for just 13.0% of the cross-tract dispersion in the ratio of September 2022 to February 2020 OSW in Houston and just 10.3% in San Francisco.

What Does a Tract-Level Analysis Add? As with the conditional analysis, we would like to know how much is added by working with data for tracts as opposed to examining geographic variation in our OSW ratio measure at the level of the state or county. Table 8 shows that state and county effects do a somewhat better job of accounting for spatial variation in the unconditional ratios than was the case for the OSW measures examined in our conditional analysis. On their own, state effects explain 27.7% of the cross-tract variation in the September 2020 OSW ratio, less than half as much as the 59.6% of the cross-tract variation explained by our covariates on their own. When both are included in the September 2020 model, state effects account for just 14.3% of the OSW ratio variation as compared to the 48.6% accounted for by our covariates, exclusive of the state-level lockdown variable. County effects on their own account for nearly as much of the cross-tract variation in the September 2020 OSW ratio as our covariates, but when both are included in the model, the contribution of county effects to explaining the September 2020 OSW ratio variance drops significantly to 32.1% versus 35.6% for our covariates, exclusive of the effects of covariates that can be measured only at the county level.

State and county effects do much less well in explaining the September 2021 and September OSW ratios. For example, state effects account for just 10.4% of the variation in the September 2022 ratio on their own (versus 15.1% for our covariates on their own) and just 8.5% of the variation when included in the model with other controls (versus 10.8% for the other controls). County effects by themselves account for only 21.6% of the variation in the September 2022 ratio (versus the 15.1 % already cited for our covariates on their own) and 16.5% of the

variation when included in the model along with other controls (versus 9.0% for the other controls). While county effects capture more of the variation in the OSW ratio measures than state effects, in both the 2021 and the 2022 county effect models, the residual cross-tract variation dwarfs the variation explained by the county effects.

For all three of the ratio measures examined in our unconditional analysis—the ratios of our September 2020, September 2021 and September 2022 OSW measure to the February 2020 OSW measure—using state or even county as the unit of observation would miss considerable tract-level variation, a sizeable portion of which can be accounted for by observable tract-level characteristics.

VI. *Comparisons to Other Estimates and Sensitivity Checks*

Mobility device location data are largely unfamiliar to economists. For that reason, it is perhaps especially important to the credibility of our findings to ask how they align with results from other sources and whether they are sensitive to the specific assumptions we have made in carrying out our analysis.

Comparisons with Estimates from Other Sources

As a check on the device-based patterns, it is useful to compare our estimates of onsite work to estimates from other sources. For the conditional measures, we use data 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 conditional 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.

The top panel of Table 9 compares our estimates of the share of February 2020 onsite workers who were working onsite in May 2020 and August 2020 with estimates based on the RPS data. The similarity of the two sets of estimates bolsters our confidence in our conditional OSW measures.[21]

Although the RPS data are not available for September 2021 or September 2022, they do allow us to construct unconditional estimates of the ratio of September 2020 OSW to February 2020 OSW that we can compare to our estimates. As shown in the bottom panel of Table 9, the RPS estimate of this ratio is very similar to that we obtain from the device data.[22]

As a benchmark for the September 2021 and September 2022 ratio estimates, we turn to the ACS one-year estimates at the county level, available for counties with populations of 65,000 or more. The timing of the ACS estimates is somewhat different than for our estimates. The ACS numbers are ratios of the OSW estimates for calendar years 2022 and 2021 to the ACS OSW estimate for calendar year 2019, whereas our estimates based on the device data are ratios of

[21] Although the RPS sample size is relatively small, just 2,939 in May 2020 and 4,021 in August 2020, so that the subnational RPS estimates should be viewed with considerable caution, the data do permit calculations at the Census Division level. We take additional reassurance from the fact that our estimates and the RPS estimates are reasonably highly correlated across Census Divisions. The Pearson correlation between the comparable OSW estimates in the RPS and MTI/CATT Lab data at the Census Division level is 0.66 in May 2020 and 0.47 in August 2020.

[22] The Pearson correlation at the Census Division level between the comparable ratio OSW estimates in the RPS and the MTI/CATT Lab is 0.81.

September 2022 and September 2021 OSW estimates to the February 2020 OSW estimate. The two sets of estimates are otherwise conceptually very comparable.[23] Again, we find the external estimates to be very similar to our estimates based on the device data.

Sample Size Sensitivity

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 the conditional analysis, a possible concern is that, even though we have restricted our sample to tracts with at least 10 devices in every period, some of this variation could be simply noise. Noise due to small sample sizes is not an especially relevant issue for our unconditional analysis, as the minimum tract-level device count in the sample for that analysis is 100 devices.

To assess the sensitivity of our baseline conditional estimates to sampling error, we have looked at how the mean rate of persistence in OSW and the cross-tract distribution of persistence in OSW 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 appendix Table B13. Not surprisingly, the larger the minimum number of usable devices required for inclusion in the sample, the smaller the number of tracts. The mean conditional OSW share is relatively insensitive to the choice of sample restriction. The gaps between the 90th percentile and 10th percentile conditional OSW

[23] To construct the ACS based ratios at the county level, we need to know, for both the starting and the ending year, the working age population, the number of workers and, of those workers, the number who work from home. We have the necessary information to construct the 2021 to 2019 ratio for 771 counties and to construct the 2022 to 2019 ratio for 777 counties. All of these large counties are represented in the MTI/CATT Lab device database.

rates are somewhat smaller when the sample of tracts is restricted to those with a larger number of usable devices, but the declines are modest in size.[24] Moreover, the changes from May to August in these measures are broadly robust across the different samples.

In Appendix Table B14, we also report variance decompositions for the difference specifications (May 2020 minus May 2019 and August 2020 minus May 2019) for the different samples. Consistent with sampling error being smaller when the underlying sample is larger, the residual variation in these decompositions is smaller when the tract inclusion criterion is more restrictive. In the May 2020 minus May 2019 regressions, for example, the residual variation falls from 66.9% when requiring tracts to have just one device to 55.4% with a minimum of 10 devices, 47.8% with a minimum of 20 devices and 42.6% with a minimum of 30 devices. Although the explained variation is greater in the samples with a higher minimum device count, the relative importance of the different covariates is quite robust across samples.

Schools?

Another possible concern is that the identification of work locations may be confounded with commuting to other locations. The most obvious source of concern is that our OSW measures could be picking up students who are spending their days at school. In Appendix Table B15, we report conditional 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 B16.

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

VII. Conclusion

Since the pandemic, Americans in different communities have had very different experiences with regard to work from home (WFH) or, conversely, onsite work (OSW). Other research has found that cities such as San Francisco exhibited larger and more persistent declines in OSW than cities such as Houston, but the existing literature has not systematically explored how the changes in OSW have varied at a more disaggregated level. We establish that, within cities and even within counties, there has been substantial cross-neighborhood variation in the evolution of OSW.

We draw our inferences on the spatial variation in how OSW has changed over time using two related but distinct approaches to analyzing the mobile device location data. First, in our conditional analysis, we track individuals who were working onsite in February 2020 over the next three to six months, then contrast the persistence of OSW for that group with the persistence among individuals who had been working onsite in February 2019. There were dramatic declines in OSW persistence at the national level but also substantial cross-tract variation in the magnitude of this decline in the early post-pandemic period. Second, in our unconditional analysis, we compute the ratio of measured OSW in September 2020, September 2021 and September 2022 to that in February 2020. At the national level, this OSW ratio dropped sharply in September 2020 and had only partially recovered by September 2022. In each of the three endpoint months we examine, however, there was substantial variation across tracts nationally, within cities and states, and even within counties.

We identify a number of factors that help to explain the significant cross-tract variation in our OSW measures. In both our conditional and unconditional analyses, the most important

contributors are the industry and occupation mix of the workers who live in the tract. Median household income also plays a strong supporting role. Although we are able to account for much of the cross-tract variation in OSW, a significant fraction is not explained even by the rich set of tract characteristics that we consider and this is more true in later months.

The finding that the change in OSW has varied so much across neighborhoods has important implications for analysts and policymakers. In neighborhoods with fewer residents who are working onsite, more people will be at home during the day and correspondingly fewer people will be 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 (2024) find that the surge in business formation since the start of the pandemic exhibits spatial variation at the local level consistent with changing WFH/OSW patterns. Fully adapting to a situation in which there have been lasting neighborhood-by-neighborhood impacts on the share of workers who are at home during the day, however, is likely to take time.

We are able to look at how OSW changed in such spatially granular detail thanks to the millions of mobile devices for which we have high frequency location observations. We can imagine similar information being valuable for tracking the impact of future events on WFH/OSW behavior, especially if the information could be generated on a close-to-real-time basis. Working with these data, however, poses some inherent challenges. In addition to the technical challenges associated with transforming many millions of latitude and longitude coordinates into usable data, there is the challenge that some onsite workers may not be identifiable as such. We have dealt with this in our conditional analysis by restricting our

attention to individuals for which we have strong evidence of onsite work in a base period and then following them longitudinally; in our unconditional analyses, we did so by focusing on the ratio of measured OSW between different points in time rather than on the level of the OSW measure. There are assumptions underlying both approaches, but it is reassuring that our topside estimates are consistent with estimates from other sources.

Going forward, whether databases of the size we were able to access will continue to be available for research purposes may be just as important an issue as the technical challenges of working with the data. Consumer privacy concerns, company responses and government interventions all have affected the ability of data aggregators to obtain and disseminate mobile device location data. In 2021, Apple made changes to its iOS operating system to require that users give explicit permission to apps to track their behavior and sell their data. Google made similar changes affecting Android phone owners in 2022. The Federal Trade Commission has followed up with greater monitoring and regulation of mobile device location data aggregators.

The most obvious effect of these changes has been to reduce the size of available high-quality samples. In some cases, data aggregators have turned to providing model-based data that are problematic for the types of exercises we have undertaken.[25] A challenge going forward will be whether the amount of data available is sufficient to support analyses of the sort we have described. More broadly, as with other forms of naturally occurring administrative data, we see these data as having great potential but also that there are challenges to using them, including challenges related to the necessity of protecting individual privacy.

[25] In working with data for other projects, the MTI/CATT Lab members of our team noticed problems with the data from certain providers. For example, in some cases, the data contained observations showing devices moving across bridges that had collapsed or large numbers of devices visiting Chik-fil-A locations on Sundays when those restaurants are closed. We have not used data supplied by any of these providers in our analyses.

References

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

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. 2022. "Real-Time Population Survey." Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi. org/10.3886/E158081V4

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.

Burrows, Michael, Charlynn Burd and Brian McKenzie. 2023. "Home-Based Workers and the COVID-19 Pandemic," U.S. Census Bureau, American Community Survey Report ACS-52.

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. 2024. "Surging Business Formation in the Pandemic: Causes and Consequences?" *Brookings Papers on Economic Activity*, *Fall 2023*: 249-302.

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.

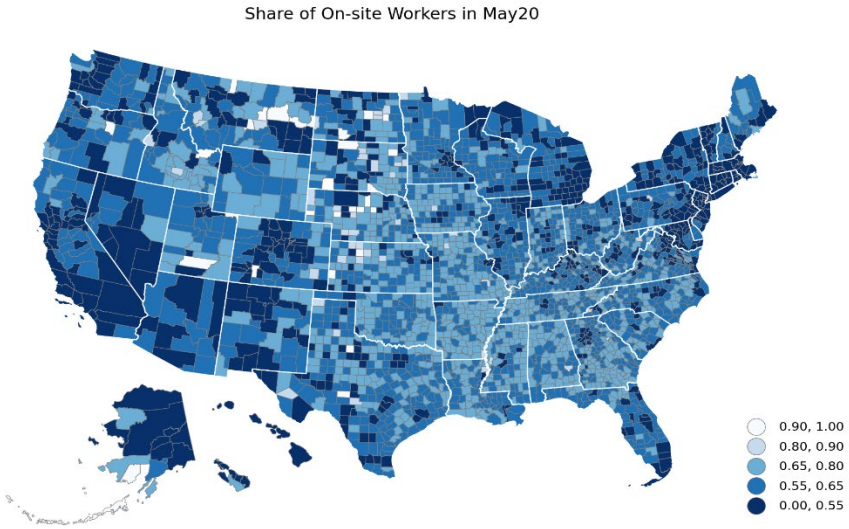
Rafiq, Rezwana, Michael G. McNally, Yusuf Sarwar Uddin, and Tanjeeb Ahmed. 2022. “Impact of Working from Home on Activity-Travel Behavior During the COVID-19 Pandemic: An Aggregate Structural Analysis.” *Transportation Research Part A: Policy and Practice,* 159: 35-54.

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. 2023. "Interactive COVID-19 Mobility Impact and Social Distancing Analysis Platform." *Transportation Research Record,* 2677(4): 168-180.

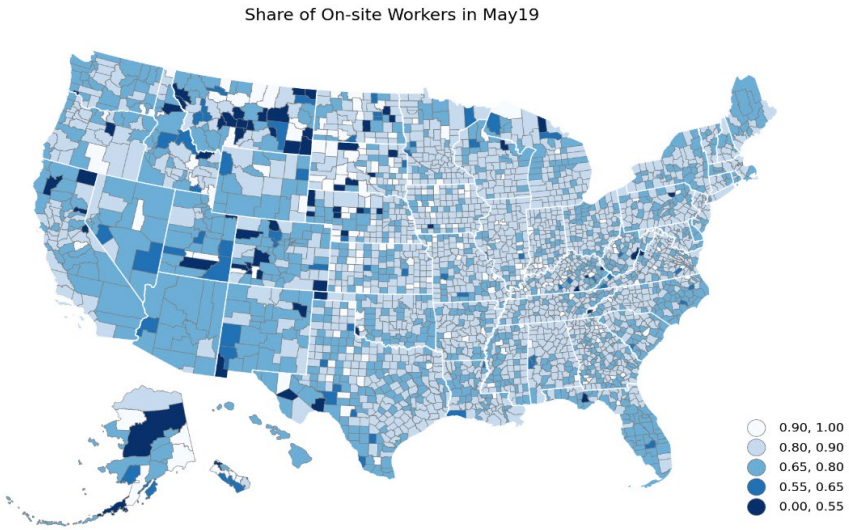
Zhang, L., A. Darzi, Y. Pan, M. Yang, Q. Sun, A. Kabiri, G. Zhao, and C. Xiong. 2023. "Next Generation National Household Travel Survey National Origin Destination Data Passenger Origin-Destination Data Methodology Documentation." Available online at https://nhts.ornl.gov/od/assets/doc/2020_NextGen_NHTS_Passenger_OD_Data_Methodology_v2.pdf. Accessed February 20, 2025.

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



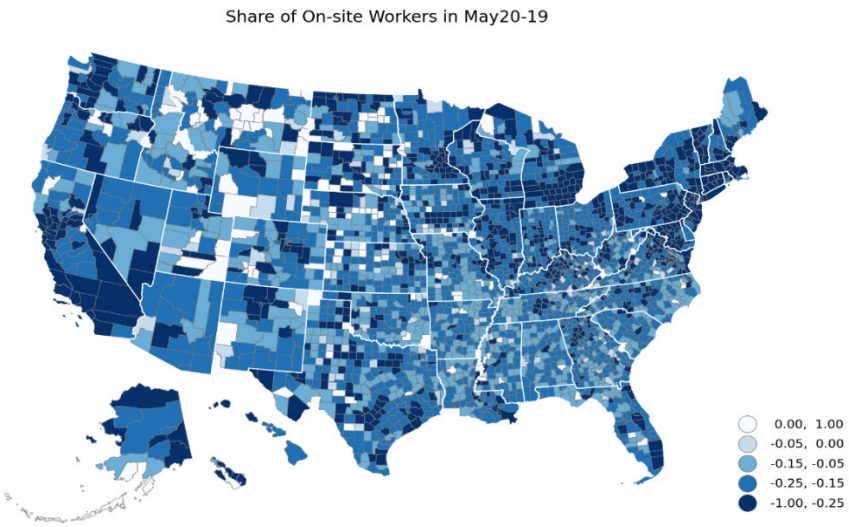
Source: Authors' tabulations from MTI/CATT Lab database.

Figure 2: Shares of Workers 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

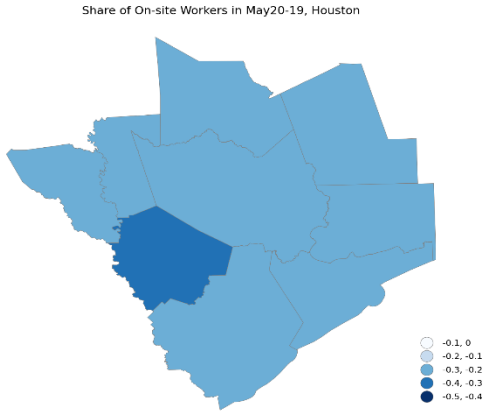


Source: Authors' tabulations from MTI/CATT Lab database.

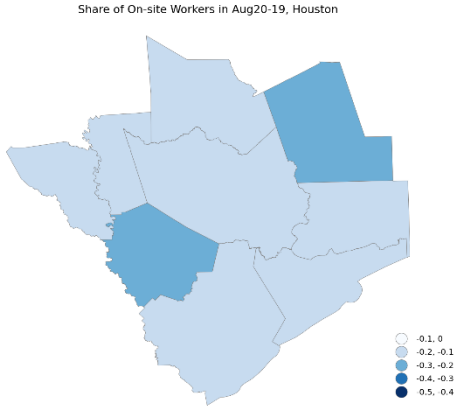
Figure 4: Difference in Share of Workers Persisting in Onsite Work from February, May 2020 minus May 2019 and August 2020 minus August 2019, Houston and San Francisco

Houston

May 20 minus May 2019)

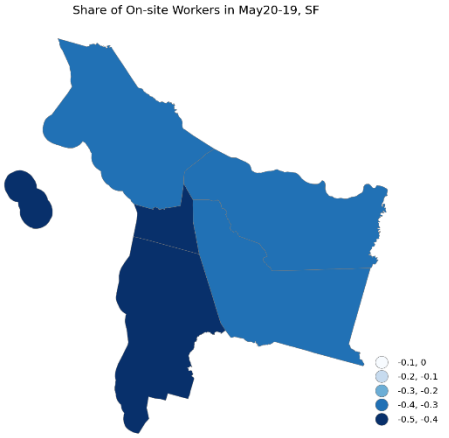


August 2020 minus August 2019

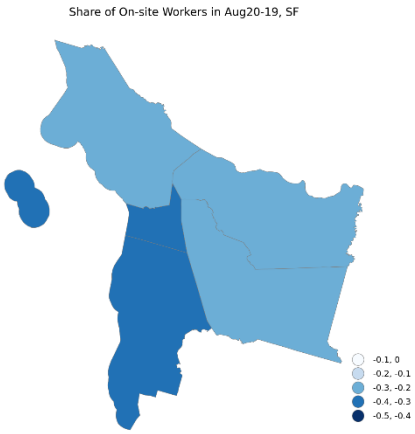


San Francisco

May 20 minus May 2019

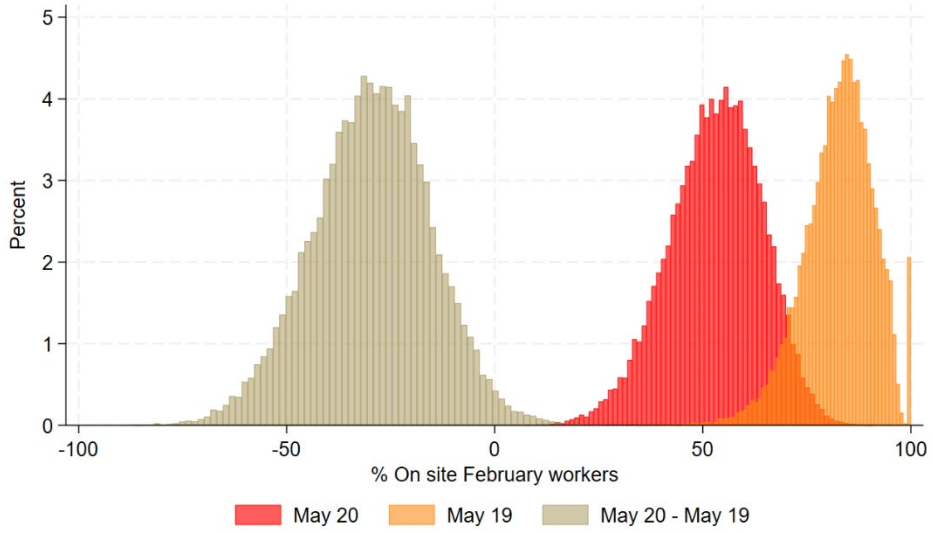


August 2020 minus August 2019)



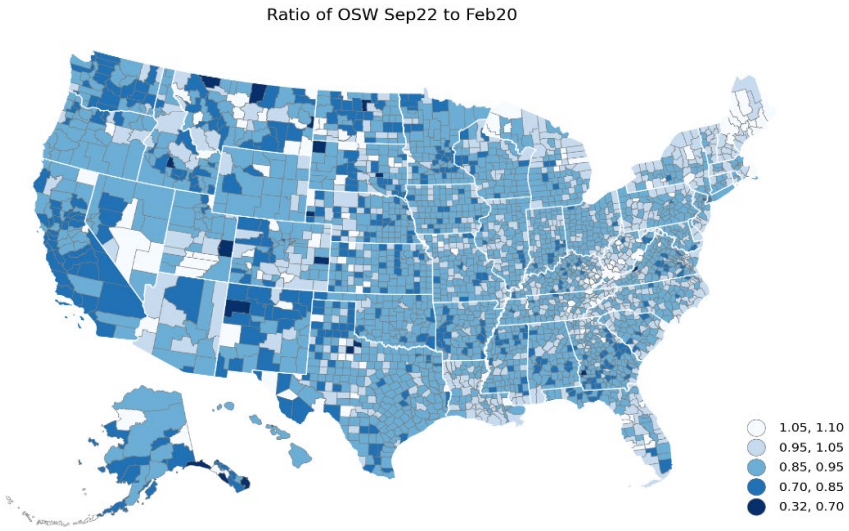
Source: Authors' tabulations from MTI/CATT Lab database.

Figure 5: Tract-Level Distribution of Onsite Work among February 2020 and 2019 Onsite Workers in May 2020 and May 2019 and Difference across Two Years



Source: Authors' tabulations from MTI/CATT Lab database.
Note: Estimates are employment weighted. Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Figure 6: Ratio of OSW in September 2022 to OSW in February 2020

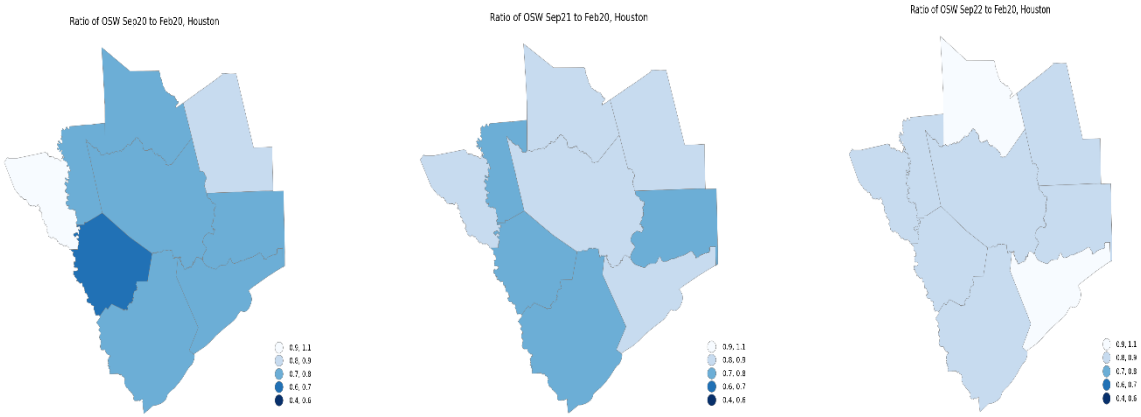


Source: Authors' tabulations from MTI/CATT Lab database.

Figure 7: Ratio of OSW in September 2020, September 2021 and September 2022 to OSW in February 2020, Houston and San Francisco

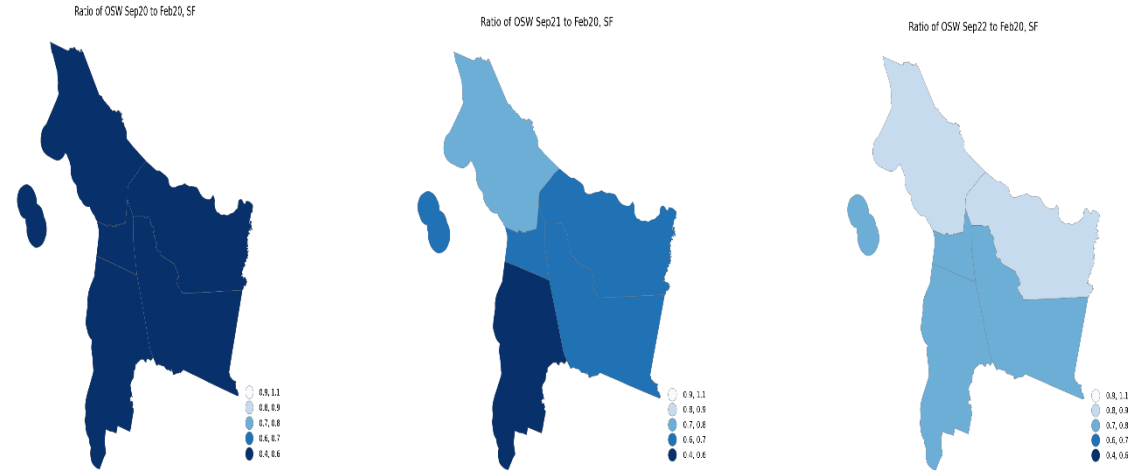
Houston

Ratio Sep 20 Ratio Sep 21 Ratio Sep 22



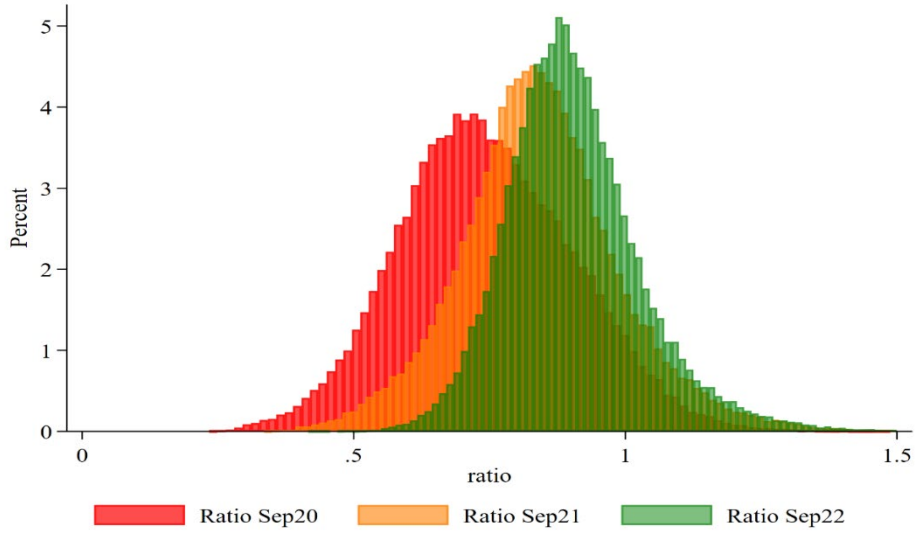
San Francisco

Ratio Sep 20 Ratio Sep 21 Ratio Sep 22



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

Figure 8: Tract-Level Distribution of Ratio of OSW in September 2020, September 2021 and September 2022 to OSW in February 2020



Source: Authors' tabulations from MTI/CATT Lab database.
Note: Estimates are employment weighted. Sample restricted to Census tracts with 100 or more devices with a home location in every period. N=62,829 Census tracts.

Table 1: Percent in Onsite Work (OSW) in Later Months Among Individuals in OSW in February 2020 or February 2019

	Among those in OSW in February 2020, percent in OSW as of:		Among those in OSW in February 2019, percent in OSW as of:		Difference in share of workers persisting in OSW from February	
	May 2020	August 2020	May 2019	August 2019	May 2020-May2019	August 2020-August 2019
Mean	53.0	62.6	82.6	78.4	-29.6	-15.8
p10	38.1	47.9	71.6	64.0	-48.4	-34.0
p50	53.6	63.2	83.4	79.5	-29.3	-16.0
p90	67.1	76.5	92.8	91.6	-11.5	2.7

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Algorithm for determining OSW described in text. Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Table 2: Ratio of Percent in Onsite Work (OSW) in Later Months to Percent in OSW in February 2020

	September 2020 to February 2020 OSW ratio	September 2021 to February 2020 OSW ratio	September 2022 to February 2020 OSW ratio
Mean	0.75	0.84	0.91
p10	0.54	0.66	0.76
p50	0.74	0.84	0.89
p90	0.96	1.03	1.06

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Algorithm for determining OSW described in text. Sample restricted to Census tracts with 100 or more devices with a home location in all sample periods. N=62,829 Census tracts.

Table 3: Factors Affecting Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February 2020 or February 2019

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

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the share of those working onsite in February 2020 or February 2019 with OSW in indicated month or, in final two columns, the 2020 minus 2019 difference in those shares. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Table 4: Percent of Variance in Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February 2020 or February 2019 Explained by Various Factors

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

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the share of those working onsite in February 2020 or February 2019 with OSW in indicated month or, in final two columns, the 2020 minus 2019 difference in those shares. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Table 5: Percent Contribution of State and County Effects to Explaining the Variance in the Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February 2020 or February 2019

	OSW February 2020 and:		OSW February 2019 and:			
Explanatory variables	May 2020	August 2020	May 2019	August 2019	May 2020 minus May 2019	August 2020 minus August 2019
State effects only	22.4	19.3	2.5	7.4	15.8	5.7
State effects with controls	9.5	8.0	1.1	7.0	6.8	2.1
Residual (full model including state effects)	40.5	50.3	86.2	87.5	54.0	74.4
County effects only	50.3	45.2	13.0	16.7	36.5	22.4
County effects with controls	20.5	17.3	9.0	14.8	18.1	9.9
Residual (full model including county effects)	36.2	45.1	78.5	79.9	48.7	68.4
Dep. var. mean	53.0	62.6	82.6	78.4	-29.6	-15.8
Dep. var. standard deviation	(11.2)	(11.2)	(8.5)	(10.9)	(14.4)	(14.7)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the share of those working onsite in February 2020 or February 2019 with OSW in indicated month or, in final two columns, the 2020 minus 2019 difference in those shares. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Table 6: Factors Affecting Ratio of Prevalence of Onsite Work (OSW) in Later Months to Prevalence of OSW in February 2020

Explanatory variables	Mean (standard deviation)	Ratio of OSW in month to OSW in February 2020		
		September 2020	September 2021	September 2022
Share of population:				
Age 25-64	52.6	0.183	-0.086	-0.105
	(6.5)	(0.008)	(0.010)	(0.009)
Age 65 plus	15.5	0.163	0.15	0.236
	(6.8)	(0.008)	(0.010)	(0.009)
White, non-Hispanic	63.7	0.071	0.157	0.094
	(27.7)	(0.003)	(0.004)	(0.004)
College graduate	41.5	-0.091	0.007	0.017
	(18.2)	(0.007)	(0.010)	(0.008)
ln(mean household income	4.2	-7.53	-6.373	-1.593
	(0.4)	(0.197)	(0.255)	(0.223)
Share commute public trans.	16.6	-0.007	-0.011	-0.056
	(16.3)	(0.003)	(0.004)	(0.004)
Share commute 30+ mins.	37.5	-0.048	-0.029	0.021
	(15.8)	(0.004)	(0.004)	(0.005)
Rural yes/no	16.1	0.2	0.013	-0.014
	(36.7)	(0.004)	(0.002)	(0.002)
Share Trump vote in 2016	46.6	0.109	-0.045	0.051
	(20.5)	(0.005)	(0.006)	(0.005)
May 2020 state lockdown	49.6	-0.071	-0.051	-0.014
	(42.6)	(0.001)	(0.002)	(0.001)
May 2020 local lockdown	5.0	-0.023	-0.001	-0.012
	(18.2)	(0.002)	(0.003)	(0.003)
May 2020 cum COVID deaths	2.1	0.273	0.468	0.279
	(3.2)	(0.016)	(0.020)	(0.018)
Industry dummies	--	Yes	Yes	Yes
Occupation dummies	--	Yes	Yes	Yes
Dep. var. mean	--	74.2	84.4	90.5
Dep. var. standard deviation	--	(16.5)	(15.4)	(12.9)
R-squared	--	0.596	0.214	0.151

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the ratio of OSW share in indicated month to OSW share in February 2020. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 100 or more devices with a home location in every period. N=62,829 Census tracts.

Table 7: Percent of Variance in Ratio of Prevalence of Onsite Work (OSW) in Later Months to Prevalence of OSW in February 2020 Explained by Various Factors

	Ratio of OSW in month to OSW in February 2020		
Explanatory variables	September 2020	September 2021	September 2022
Share of population:			
Age 25-64	-1.4	0.7	0.8
Age 65 plus	1.4	1.1	2.9
White, non-Hispanic	3.0	4.8	4.6
College graduate	5.2	-0.1	0.0
ln(mean household income)	9.5	5.2	0.3
Share commute public trans.	0.3	0.1	1.2
Share commute 30+ mins.	1.3	0.7	-0.3
Rural (yes/no)	4.4	0.5	-0.2
Share Trump vote in 2016	12.3	-0.9	1.6
May 2020 state lockdown	5.9	2.6	0.5
May 2020 local lockdown	0.1	0.0	0.0
May 2020 cum COVID deaths	-0.9	0.0	0.1
Industry mix	11.4	5.3	2.8
Occupation mix	7.3	1.3	1.0
Residual	40.4	78.6	84.9
Dep. var. mean	0.75	0.84	0.91
Dep. var. standard deviation	(0.16)	(0.15)	(0.13)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the ratio of OSW share in indicated month to OSW share in February 2020. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 100 or more devices with a home location in every period. N=62,829 Census tracts.

Table 8: Percent Contribution of State and County Effects to Explaining Variance in Ratio of Prevalence of Onsite Work (OSW) in Later Months to OSW in February 2020

	Ratio of OSW in month to OSW in February 2020		
Explanatory variables	September 2020	September 2021	September 2022
State effects only	27.7	17.5	10.4
State effects with controls	14.3	14.5	8.5
Residual (full model including state effects)	37.1	72.9	80.7
County effects only	59.3	27.2	21.6
County effects with controls	32.1	22.8	16.5
Residual (full model including county effects)	32.3	68.2	74.5
Dep. var. mean	0.75	0.84	0.91
Dep. var. standard deviation	(0.16)	(0.15)	(0.13)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the ratio of OSW share in indicated month to OSW share in February 2020. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 100 or more devices with a home location in every period. N=62,829 Census tracts.

Table 9: Comparisons of Onsite Work (OSW) Patterns in MTI/CATT Lab, Realtime Population Survey (RPS) and American Communithy Survey (ACS) Data

A: Conditional Analysis

	Among those working onsite in February 2020, percent OSW as of:	
	May 2020	August 2020
MTI/CATT Lab	53.0	62.6
RPS	52.9	60.5
Difference	0.1	2.1

B: Unconditional Analysis

	Ratio of OSW in later month to OSW in February 2020		
	September 2020	September 2021	September 2022
MTI/CATT Lab	0.75	0.84	0.91
RPS	0.77	na	na
Difference	-0.02	na	na
	Ratio of OSW in later year to OSW in 2019		
	na	2021	2022
ACS	na	0.83	0.89
Difference	na	-0.01	-0.02

Source: Authors' calculations, MTI/CATT Lab mobile device location database, Real-time Population Survey (RPS) public domain database, and American Community Survey (ACS).

Note: For conditional analysis, MTI/CATT Lab sample restricted to tracts with 10 or more devices with a home location in all sample periods. RPS N= 2,939 for May 2020, 4,021 for August 2020, 2,245 for November 2020, and 4,287 for March/April 2021. For unconditional analysis, MTI/CATT Lab sample restricted to tracts with 100 or more devices with a home location in every period. RPS N=73,895 for February 2020 and 8,423 for September 2020. ACS OSW one year estimates for large counties. For this table, MTI/CATT Lab estimates for September 2021 and September 2022 restricted to these counties.

A. Measurement Appendix

This appendix describes the processes we follow to estimate the onsite work share at the Census tract level and the covariates used in our decomposition analysis. For both the conditional and the unconditional analysis, we make use of the repository of smart device location observations assembled by the Maryland Transportation Institute (MTI) and Center for Advanced Transportation Technology (CATT) Lab. The details differ somewhat between the conditional and the unconditional analysis, but the general framework for identifying devices' home and work locations is described in Pan et al. (2023).

Measurement Framework for the Conditional Analysis

Appendix Table B1 provides summary statistics regarding the raw numbers of devices available for the conditional analysis. These estimates are based on 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, together with 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 2020 dataset begins with a sample of approximately 4.2 million devices observed in February 2020 that shrinks through attrition to approximately 2.1 million devices in August 2020. Our 2019 dataset 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 for the conditional analysis, 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 conditional 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's 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 identifying workplace locations for the conditional analysis 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.

2. Protecting Against Misclassification

The process described above is implemented in all months. For the 2020 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 Longitudinal Employer-Household Dynamics (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. These shares and differences between the shares for a given month are used as our dependent variables. For both the descriptive statistics and the regression analysis, the estimates we report are weighted using 2015-2019 tract-level employment from the American Community Survey (ACS).

Measurement Framework for the Unconditional Analysis

The framework that we implemented for the unconditional analysis applies the same basic logic as the framework for the conditional analysis just described. Like that methodology, it utilizes geohash encoding to efficiently cluster location sightings based on latitude and longitude, enabling the identification of meaningful home and work candidates. The algorithm for the unconditional analysis begins by identifying home and work locations at the level-6 geohash scale (grids of approximately 1220 m × 610 m at the equator) and subsequently refines these locations to the more precise level-7 geohash scale (grids of approximately 152.9m*152.4m at the equator) within the boundaries of the identified level-6 geohash.

1. *Home Location Identification*

The identification of home locations for the unconditional analysis is performed in two stages. First, a set of candidate level 6 geohashes is created and the one that best satisfies our specified criteria is selected:

1) Select the level 6 geohashes that are observed an average of 2 or more hours per day on at least $\max\left\{3, \text{integer}\left(\frac{\text{Number of observed days}}{2}\right)+1\right\}$ days
2) Sort the home level 6 geohash location candidates by the observed number of days, average daily number of observed hours, and average number of hourly sightings;
3) Keep the 3 top-ranked level 6 geohash home location candidates from step 2) and sort them by observed number of nights, average daily number of observed nighttime hours, and average number of hourly sightings during nighttime;
4) Select the top-ranked level-6 geohash from step 3) as the home location; if a tie exists, select the top-ranked one based on step 2).

Then, a level 7 geohash within the chosen level 6 geohash is identified as the home location:

1) Starting from all the level 7 geohashes within the level 6 geohash home location, sort the level 7 geohash candidates by observed number of days, average daily number of observed hours, and average number of hourly sightings;
2) Keep 3 top-ranked level 7 geohash candidates and sort them by observed number of nights, average daily number of observed nighttime hours, and average number of hourly sightings during nighttime;
3) Select the top-ranked level 7 geohash as the home location; if a tie exists, select the top-ranked one based on step 1).

One difference between the algorithm used to identify the home location for the conditional analysis and that used for the unconditional analysis is the consideration of observed nighttime hours in the latter. The definition of nighttime was derived from the American Time Use Survey (ATUS). According to the 2017, 2018, and 2019 ATUS data, over 80% of full-time and part-time workers observed at home during the survey day remained at home between 9:00 PM and 5:59 AM. This timeframe was designated as the nighttime window in our algorithm for the unconditional analysis.

Another difference is that home locations for the conditional analysis were defined as level 8 geohashes, whereas for the unconditional analysis they were defined as (larger) level 7 geohashes.

2. *Work Location Identification*

The objective of fixed workplace identification is to recognize one's major work location that is not the home location. Therefore, workplace location candidates are level 6 geohashes that are not one's home geohash. The algorithm applies a temporal similarity ratio similar to the one described for the conditional analysis to exclude level 6 geohashes in which there would be too much overlap between the hours that the device is observed in the home location and the hours it is observed in the work location.

Similar to the identification of the home location, the work location is identified in two steps. First, we identify a level 6 geohash for the work location:

1) Select the level 6 geohashes that are observed an average of 2 or more hours per day on at least $\max\left\{3, \text{integer}\left(\frac{\text{Number of observed workdays}}{2}\right) + 1\right\}$ workdays
2) Sort the work location candidates by observed number of workdays, average workday number of observed hours, and average workday number of hourly sightings;
3) Keep 3 top-ranked candidate work locations;
4) Check the hours each candidate work location was observed, i.e., $\{W^i\}$ for candidate i, and count the ratio of hours in $\{W^i\}$ when the device also was observed in the imputed home location. This is the temporal similarity ratio;

5) Sort the top three level 6 geohash work location candidates in ascending order of similarity ratio;

6) Select the top-ranked level 6 geohash as the work location provided it has a temporal similarity ratio smaller than the maximum allowable temporal similarity threshold.

Then, a level 7 geohash within the chosen level 6 geohash is identified as the work location:

1) Sort the level 7 geohash work location candidates by observed number of workdays, average workday number of observed hours, and average workday number of hourly sightings;
2) Select the top-ranked level-7 geohash as the work location.

One key difference between the algorithms used for the conditional and unconditional analyses lies in the treatment of time. For the identification of work locations, theunconditional analysis is restricted to considering only observed hours on workdays, defined as non-holiday weekdays. Additionally, the work locations in the conditional analysis were defined using more precise level 8 geohashes, while the unconditional analysis employed less granular level 7 geohashes.

3. Tract Level Weights

Because we are using repeated cross sections rather than samples of devices that we follow longitudinally, the unconditional analysis does not require accounting for attrition. Because the cross-sectional samples are not restricted to people who were observed in onsite work as of a given date, we also cannot apply the iterative proportional fitting approach that was used for the conditional analysis. Instead, our estimate of the prevalence of onsite work in a tract is simply the share of devices with a home location in that tract for which we observe an onsite work location. For both the descriptive statistics and the regression analysis, the estimates we report are weighted using 2015-2019 tract-level employment from the American Community Survey (ACS).

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

The explanatory variables used in the attrition weighting models for the conditional analysis and the 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/

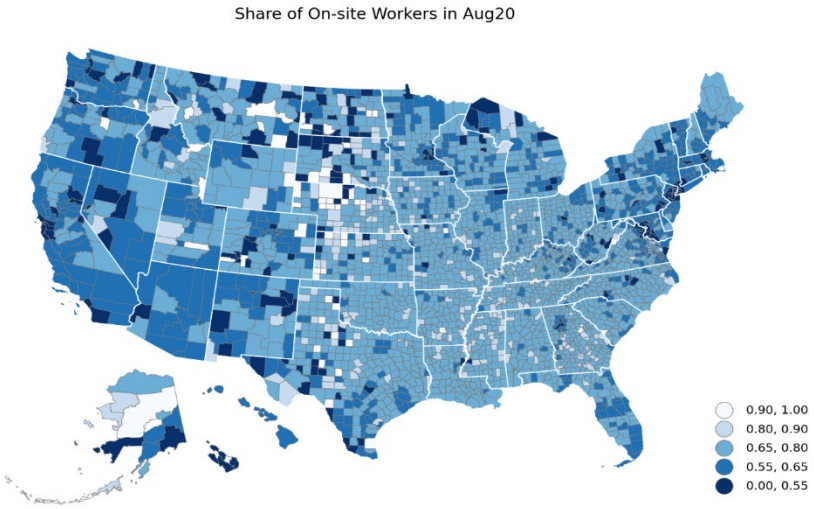
References

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.

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.

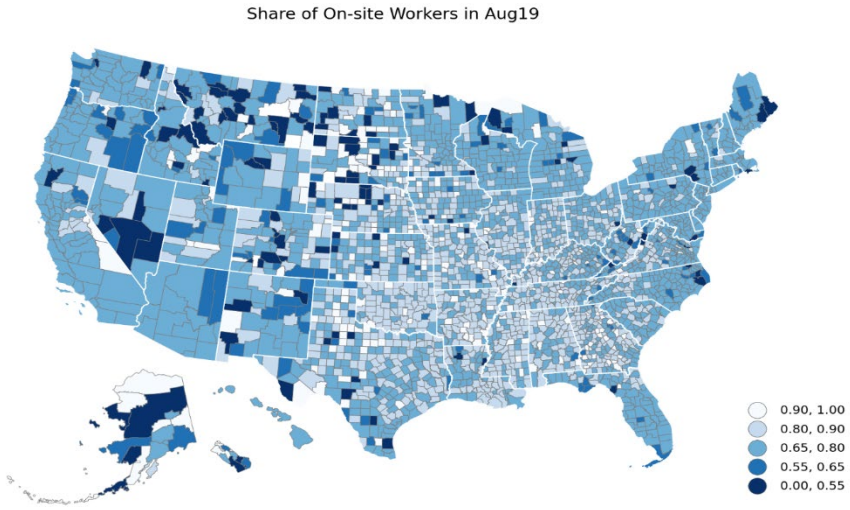
B. Supplemental Figures and Tables

Appendix Figure B1: Shares of Workers Onsite in February 2020 Who Also Were Onsite in August 2020



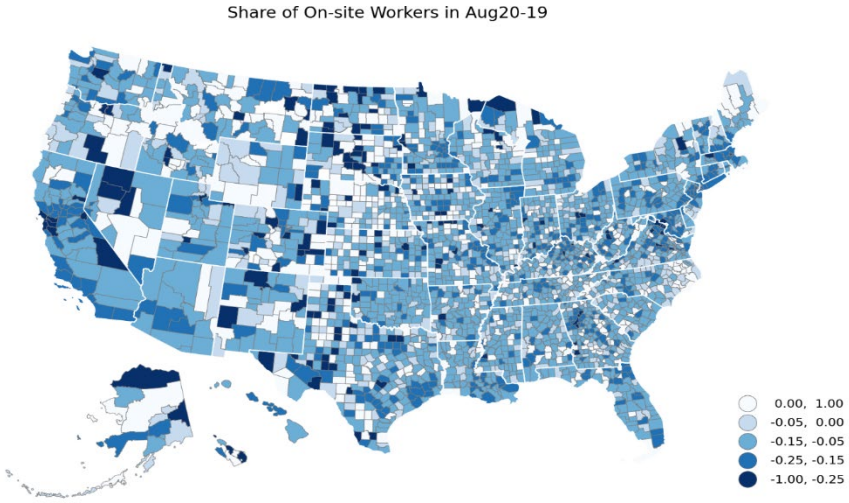
Source: Authors' tabulations from MTI/CATT Lab database.

Appendix Figure B2: Shares of Workers Who Were Onsite in February 2019 Who Also Were Onsite in August 2019



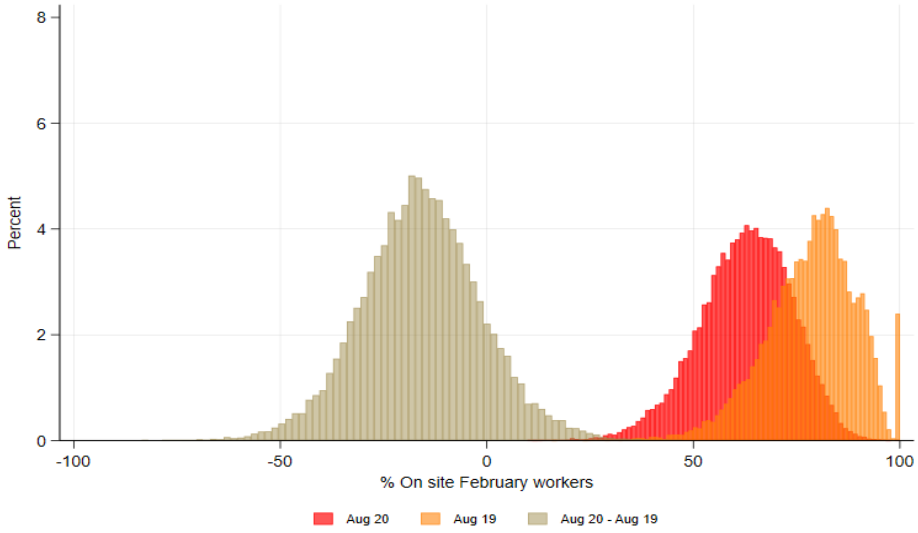
Source: Authors' tabulations from MTI/CATT Lab database.

Appendix Figure B3: Difference in Share of Workers Persisting in Onsite Work from February, August 2020 minus August 2019



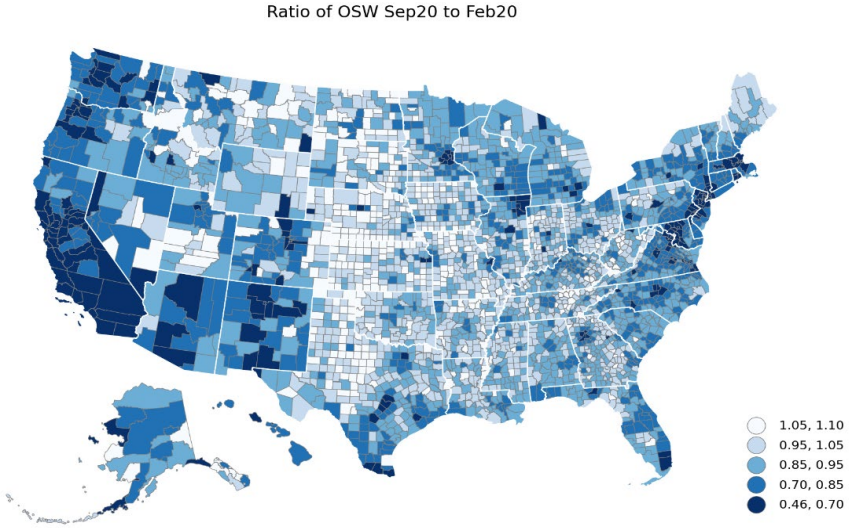
Source: Authors' tabulations from MTI/CATT Lab database.

Figure 5: Tract-Level Distribution of Onsite Work among February 2020 and 2019 Onsite Workers in August 2020 and August 2019 and Difference across Two Years



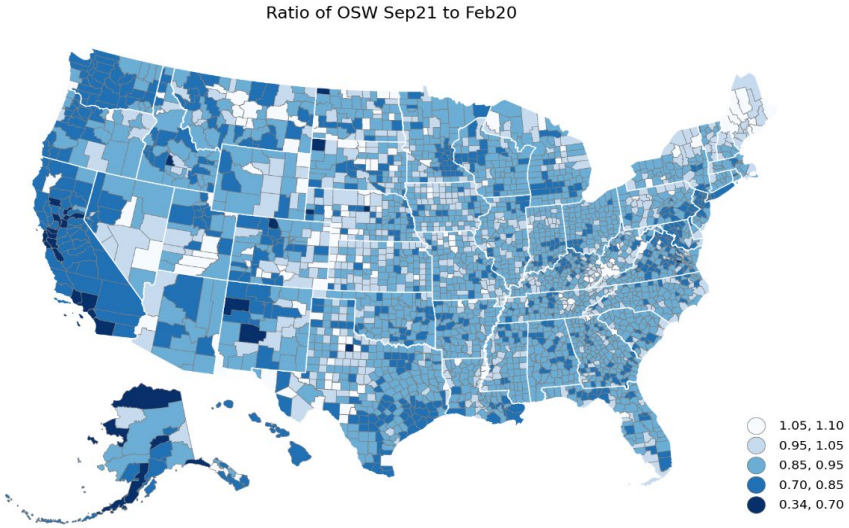
Source: Authors' tabulations from MTI/CATT Lab database.
Note: Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Appendix Figure B5: Ratio of OSW in September 2020 to February 2020



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

Appendix Figure B6: Ratio of OSW in September 2021 to February 2020



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

Appendix Table B1: Unweighted Mobile Device Counts by Month, Conditional Analysis

Selection criterion	February 2020	May 2020	August 2020
Home location	11,659,409	15,565,504	15,038,155
Work location | home location	4,230,606	3,186,920	3,478,169
Home location | work location in February 2020	4,230,606	2,804,839	2,058,125
Work location | work location in February 2020	4,230,606	1,099,422	946,327

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

	February 2020-May 2020 Sample	February 2020-August 2020 Sample	February 2019-May 2019 Sample	February 2019-August 2019 Sample
Number of devices in sample	2,804,839	2,058,125	1,276,135	846,028
Hours devices observed in February				
Mean	579	587	419	431
p10	321	331	266	294
p50	650	655	394	398
p90	687	688	605	612
Hours devices observed in end month				
Mean	381	380	505	534
p10	196	212	238	232
p50	372	376	547	611
p90	584	555	702	725

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

Appendix Table B3: Unweighted Mobile Device Counts by Month, Unconditional Analysis

Selection criterion	February 2020	September 2020	September 2021	September 2022
Home location	52,214,257	53,569,490	35,680,887	77,752,093
Work location | home location	21,201,228	16,861,241	12,208,217	28,564,867

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

Appendix Table B4: Unweighted Distribution of Observed Days and Hours for Devices for Which Home and Work Location Identified in February 2020, September 2020, September 2021 and September 2022, Unconditional Analysis

	February 2020	September 2020	September 2021	September 2022
Number of devices with observed home days	52,214,257	53,569,490	35,680,887	77,752,093
Home days observed				
Mean	21	20	21	23
p10	10	9	10	12
p50	22	20	21	26
p90	29	30	30	30
Home hours observed				
Mean	187	181	174	131
p10	27	30	25	30
p50	110	117	90	91
p90	453	446	472	296
Number of devices with observed work days	21,201,228	16,861,241	12,208,217	28,564,867
Work days observed				
Mean	10	10	12	14
p10	3	3	4	5
p50	11	10	12	14
p90	17	18	20	21
Work hours observed				
Mean	51	48	57	47
p10	6	6	9	10
p50	34	33	39	38
p90	122	113	139	93

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

Appendix Table B5: Industry Effects on Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February 2020 or February 2019

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

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

Note: Unit of observation is the Census tract. Dependent variable is the share of those working onsite in February 2020 or February 2019 with OSW in indicated month or, in final two columns, the 2020 minus 2019 difference in those shares. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Appendix Table B6: Occupation Effects on Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February 2020 or February 2019

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

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the share of those working onsite in February 2020 or February 2019 with OSW in indicated month or, in final two columns, the 2020 minus 2019 difference in those shares. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Appendix Table B7: Actual versus Predicted On-Site Work Prevalence Among Individuals Working On-Site in February 2020 or February 2019, Selected Geographic Areas

Geographic area	OSW February 2020 and: May 2020	OSW February 2020 and: August 2020	OSW February 2019 and: May 2019	OSW February 2019 and: August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Texas						
Actual	54.7	62.0	81.0	77.9	-26.3	-15.9
Predicted	55.7	63.3	80.8	78.8	-25.1	-15.5
California						
Actual	47.9	55.6	81.2	76.7	-33.3	-21.1
Predicted	47.5	55.5	80.9	75.7	-33.4	-20.2
Houston-The Woodlands-Sugarland TX						
Actual	54.0	60.1	81.3	78.0	-27.3	-18.0
Predicted	54.4	61.7	81.2	79.0	-26.7	-17.4
San Francisco-Oakland-Berkeley, CA						
Actual	40.0	47.6	81.4	77.3	-41.4	-29.7
Predicted	39.9	47.3	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.

Appendix Table B8: Percent of Variance in 2020 minus 2019 Difference in Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February Explained by Various Factors, Houston and San Francisco

	Houston		San Francisco	
Explanatory variables	May 2020 minus May 2019	August 2020 minus August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:				
Age 25-64	-0.3	-0.1	-0.2	0.0
Age 65 plus	0.0	0.0	0.0	0.0
White, non-Hispanic	-0.4	-1.8	-0.2	-0.5
College graduate	1.1	0.4	0.6	0.2
ln(mean household income)	21.0	7.4	10.3	3.3
Share commute public trans.	0.0	0.0	0.0	0.1
Share commute 30+ mins.	0.9	0.6	-0.2	0.0
Rural (yes/no)	0.0	0.0	0.0	0.0
Share Trump vote in 2016	-1.3	-0.1	-0.2	0.0
Industry mix	14.0	11.7	13.2	9.3
Occupation mix	13.7	13.0	8.3	7.0
Residual	51.2	68.9	68.3	80.6
Dep. var. mean	-27.9	-18.5	-42.5	-30.8
Dep. var. standard deviation	12.8	13.6	14.9	18.1

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the difference between the share of those working onsite in February 2020 with OSW in May or August 2020 and the corresponding May or August share for 2019. Algorithm for determining OSW described in text. Estimates are based on coefficients from Table 3 model and are employment weighted. Sample restricted to Census tracts with 10 or more devices with a home location in all sample periods. N=28,125 Census tracts.

Appendix Table B9: Industry Effects on Ratio of Prevalence of Onsite Work (OSW) in Later Months to Prevalence of OSW in February 2020

Industry	Mean employment share (percent)	Ratio of OSW in month to OSW in February 2020: September 2020	September 2021	September 2022
Agriculture	0.8	-0.051	-0.399	-0.292
		(0.042)	(0.055)	(0.048)
Mining	0.5	0.546	0.566	0.392
		(0.038)	(0.049)	(0.043)
Utilities	0.6	0.213	-1.349	-0.457
		(0.106)	(0.138)	(0.121)
Construction	5.3	-0.253	-0.539	-0.253
		(0.037)	(0.047)	(0.042)
Manufacturing	9.1	0.363	-0.076	-0.057
		(0.028)	(0.036)	(0.032)
Wholesale trade	4.1	-0.035	-0.333	-0.291
		(0.047)	(0.061)	(0.054)
Retail trade	10.7	0.495	0.477	0.616
		(0.038)	(0.050)	(0.044)
Transportation and warehousing	3.9	0.375	-0.030	0.031
		(0.038)	(0.049)	(0.043)
Information	2.1	-0.228	-0.58	-0.256
		(0.036)	(0.047)	(0.041)
Finance and insurance	4.1	--	--	--
		--	--	--
Real Estate	1.5	-0.95	-1.307	-0.594
		(0.099)	(0.128)	(0.112)
Professional services	6.5	-0.042	-0.040	0.161***
		(0.035)	(0.045)	(0.040)
Management of companies	1.7	-0.198	-0.081	0.231
		(0.057)	(0.073)	(0.064)
Administrative and support services	6.3	0.359	0.665	0.756
		(0.041)	(0.054)	(0.047)
Education	9.2	0.035	0.065	-0.013
		(0.030)	(0.039)	(0.034)
Health	14.9	0.347	-0.012	0.172
		(0.031)	(0.040)	(0.035)
Arts and entertainment	1.8	0.259	-0.276	-0.054
		(0.047)	(0.061)	(0.054)
Accommodations and food services	9.4	0.069	0.219	0.309
		(0.030)	(0.039)	(0.034)
Other services	3.1	0.361	0.029	0.052
		(0.062)	(0.080)	(0.070)
Public administration	4.5	0.069	-0.135	-0.013
		(0.031)	(0.041)	(0.036)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the ratio of OSW share in indicated month to OSW share in February 2020. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 100 or more devices with a home location in every period. N=62,829 Census tracts.

Appendix Table B10: Occupation Effects on Ratio of Prevalence of Onsite Work (OSW) in Later Months to Prevalence of OSW in February 2020

Occupation	Mean employment share	Ratio of OSW in month to OSW in February 2020: September 2020	September 2021	September 2022
Management, business and financial	15.5	0.278	0.266	0.209
		(0.017)	(0.022)	(0.019)
Computer, engineering and science	5.9	--	--	--
		--	--	--
Education, legal, community service, arts and media	10.7	0.279	0.195	0.152
		(0.018)	(0.023)	(0.020)
Healthcare practicioners and technical	6.1	0.336	0.175	0.154
		(0.021)	(0.028)	(0.024)
Healthcare support	3.2	0.245	0.22	0.129
		(0.025)	(0.032)	(0.028)
Protective service	2.1	0.23	0.287	0.205
		(0.029)	(0.038)	(0.033)
Food preparation and serving	5.7	0.24	0.194	0.157
		(0.021)	(0.027)	(0.023)
Building and grounds cleaning and maintenance	3.8	0.458	0.269	0.131
		(0.022)	(0.029)	(0.025)
Personal care and service	2.8	0.143	0.168	0.127
		(0.027)	(0.035)	(0.031)
Sales and related	10.3	0.193	0.167	0.229
		(0.018)	(0.024)	(0.021)
Office and administrative support	11.5	0.103	-0.005	-0.010
		(0.019)	(0.025)	(0.021)
Farming, fishing and forestry	0.7	0.605	0.479	0.357
		(0.032)	(0.042)	(0.036)
Construction and extraction	5.2	0.409	0.279	0.098
		(0.020)	(0.026)	(0.023)
Installation, maintenance, and repair	3.2	0.272	0.078	-0.053
		(0.026)	(0.034)	(0.030)
Production	5.9	0.362	0.225	0.114
		(0.020)	(0.026)	(0.023)
Transportation	3.8	0.303	0.084	-0.035
		(0.024)	(0.031)	(0.027)
Material moving	3.8	0.35	0.27	0.209
		(0.023)	(0.030)	(0.026)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the ratio of OSW share in indicated month to OSW share in February 2020. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample restricted to Census tracts with 100 or more devices with a home location in every period. N=62,829 Census tracts.

Table B11: Actual versus Predicted Ratio of Prevalence of Onsite Work (OSW) in Later Months to Prevalence of OSW in February 2020, Selected Geographic Areas

Geographic area		Ratio of OSW in month to OSW in February 2020: September 2020	September 2021	September 2022
Texas				
	Actual	0.75	0.84	0.88
	Predicted	0.76	0.83	0.89
California				
	Actual	0.60	0.68	0.82
	Predicted	0.62	0.74	0.84
Houston-The Woodlands-Sugarland TX				
	Actual	0.71	0.80	0.88
	Predicted	0.72	0.81	0.87
San Francisco-Oakland-Berkeley, CA				
	Actual	0.51	0.64	0.79
	Predicted	0.52	0.70	0.82

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Predicted values based on estimated coefficients from model reported in Table 6. Sample restricted to Census tracts with 100 or more devices with a home location in every period.

Appendix Table B12: Percent of Variance in Ratio in Prevalence of Onsite Work (OSW) in Later Months to OSW in February 2020 Explained by Various Factors, Houston and San Francisco

	Houston			San Francisco		
	Ratio of OSW in month to Feb 2020			Ratio of OSW in month to Feb 2020		
Explanatory variables	September 2020	September 2021	September 2022	September 2020	September 2021	September 2022
Share of population:						
Age 25-64	-2.6	0.2	0.6	-0.1	0.4	0.8
Age 65 plus	0.7	0.8	3.1	-1.7	0.1	1.5
White, non-Hispanic	0.6	-2.7	8.1	-4.5	1.7	3.1
College graduate	8.7	-0.2	0.7	8.9	-0.1	0.0
ln(mean household income)	12.0	5.9	-1.8	13.8	3.3	0.2
Share commute public trans.	0.1	-0.1	0.6	0.2	0.1	0.4
Share commute 30+ mins.	0.6	0.5	-0.1	0.2	0.0	0.1
Rural (yes/no)	1.0	0.1	-0.1	0.1	0.0	0.0
Share Trump vote in 2016	8.9	1.3	3.5	-1.2	0.1	0.4
May 2020 cum COVID deaths	0.0	0.0	0.0	0.0	-0.1	-0.1
Industry mix	1.2	3.8	-4.5	12.0	2.9	1.6
Occupation mix	10.9	2.3	2.8	10.6	1.9	2.2
Residual	57.9	88.0	87.0	61.7	89.8	89.7
Dep. var. mean	0.72	0.82	0.89	0.51	0.64	0.79
Dep. var. standard deviation	(0.11)	(0.12)	(0.11)	(0.12)	(0.15)	(0.13)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the ratio of OSW share in indicated month to OSW share in February 2020. Algorithm for determining OSW described in text. Estimates are based on coefficients from Table 6 model and are employment weighted. Sample restricted to Census tracts with 100 or more devices with a home location in every period.

Appendix Table B13: Sensitivity of Onsite Work (OSW) Percentages and Explanatory Power of OSW Regression Models for Conditional Analysis 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	May 2019	August 2019
	One or More Devices Per Tract (67,550 Tracts)			
Mean	52.3	61.4	81.0	76.6
p10	35.2	43.6	66.6	57.7
p50	53.0	62.4	82.3	78.3
p90	68.4	77.8	94.0	94.7
R-Squared	0.453	0.332	0.106	0.057
	10 or More Devices Per Tract (28,125 Tracts)			
Mean	53.0	62.6	82.6	78.4
p10	38.1	47.9	71.6	64.0
p50	53.6	63.2	83.4	79.5
p90	67.1	76.5	92.8	91.6
R-Squared	0.578	0.479	0.129	0.0870
	20 or More Devices Per Tract (7,932 Tracts)			
Mean	53.5	63.4	83.9	80.3
p10	40.4	50.7	75.5	69.1
p50	53.9	63.8	84.4	81.1
p90	66.1	75.9	91.8	90.8
R-Squared	0.652	0.575	0.160	0.142
	30 or More Devices Per Tract (2,629 Tracts)			
Mean	53.3	63.4	84.5	81.4
p10	42.1	51.5	77.1	71.8
p50	53.6	63.6	84.9	82.2
p90	64.8	75.0	91.4	90.5
R-Squared	0.704	0.648	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 are employment weighted. R-squareds are values for models specified as in Table 3.

Appendix Table B14: Sensitivity of Percent of Variance in 2020 minus 2019 Difference in Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February Explained by Various Factors to Minimum Number of Devices with a Home Location Required for Inclusion of Tract in Sample

	One or more devices		10 or more devices		20 or more devices		30 or more devices	
Explanatory variables	May 2020 minus May 2019	August 2020 minus August 2019	May 2020 minus May 2019	August 2020 minus August 2019	May 2020 minus May 2019	August 2020 minus August 2019	May 2020 minus May 2019	August 2020 minus August 2019
Share of population:								
Age 25-64	-0.2	0.0	-0.3	-0.1	-0.5	-0.5	-0.3	-0.4
Age 65 plus	0.0	0.0	0.0	-0.1	-0.2	-0.1	-0.1	-0.3
White, non-Hispanic	-0.2	0.0	0.0	0.3	0.2	0.7	0.2	0.8
College graduate	0.9	0.4	0.6	0.2	-0.8	-0.5	-2.5	0.5
ln(mean household income)	10.2	3.1	14.4	4.7	17.5	6.6	21.7	8.4
Share commute public trans.	-0.1	0.1	0.0	0.1	0.4	0.1	0.2	0.2
Share commute 30+ mins.	1.3	0.8	1.6	1.7	1.4	2.2	1.6	3.2
Rural (yes/no)	0.6	0.3	0.6	0.2	0.8	0.2	0.2	-0.3
Share Trump vote in 2016	0.2	0.1	1.2	0.1	2.2	0.7	3.7	1.7
May 2020 state lockdown	2.3	0.3	3.6	0.4	3.5	0.2	2.9	0.1
May 2020 local lockdown	0.1	0.0	0.3	0.1	0.5	0.3	0.5	0.3
May 2020 cum COVID deaths	1.7	-0.2	2.5	-0.3	2.1	-0.2	0.7	0.0
Industry mix	10.6	6.4	12.5	9.9	14.6	12.7	16.7	15.2
Occupation mix	5.8	4.5	7.8	6.9	10.6	10.7	11.9	10.5
Residual	66.9	84.4	55.4	75.7	47.8	66.9	42.6	60.1
Dep. var. mean	-26.9	-13.8	-29.6	-15.8	-30.4	-16.9	-31.1	-18.0
Dep. var. standard deviation	(19.7)	(22.6)	(14.4)	(14.7)	(12.1)	(11.8)	(10.8)	(10.4)

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation the Census tract. Dependent variable is the difference between the share of those working onsite in February 2020 with OSW in May or August 2020 and the corresponding May or August share for 2019. Algorithm for determining on-site work activity as described in text.

Appendix Table B15: Factors Affecting Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February 2020 or February 2019, Work Sites within 100 Yards of a School Dropped from Analysis

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

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the share of those working onsite in February 2020 or February 2019 with OSW in indicated month or, in final two columns, the 2020 minus 2019 difference in those shares. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample of tracts the same as in Table 3 models, except for one tract for which there were no devices after dropping those wihin 100 yards of a school. N=28,124 Census tracts.

Appendix Table B16: Percent of Variance in Prevalence of Onsite Work (OSW) in Later Months Among Individuals Working Onsite in February 2020 or February 2019 Explained by Various Factors, Work Sites within 100 Yards of a School Dropped from Analysis

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

Source: Authors' calculations, MTI/CATT Lab mobile device location database.
Note: Unit of observation is the Census tract. Dependent variable is the share of those working onsite in February 2020 or February 2019 with OSW in indicated month or, in final two columns, the 2020 minus 2019 difference in those shares. Algorithm for determining OSW described in text. Estimates are employment weighted. Sample of tracts the same as in Table 3 models, except for one tract for which there were no devices after dropping those wihin 100 yards of a school. N=28,124 Census 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 $X_1, X_2, \ldots, X_n$ (e.g., all of the covariates in Table 3).

- **Step 1:** Estimate an OLS regression of Y on a constant and $X_1, X_2, \ldots, X_n$, 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.

- **Step 2:** 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.

- **Step 3:** 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

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

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 $var(\widehat{\beta_k}X_k) = cov(\widehat{\beta_k}X_k, \widehat{\beta_k}X_k)$ so V_k can be simplified to

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

The proof consists of three steps.

Step 1: 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

$$V_k = \sum_{l:\, l=1}^{n} cov(\alpha_k + \widehat{\delta_k}Y + \widehat{\epsilon_k}, \alpha_l + \widehat{\delta_l}Y + \widehat{\epsilon_l})$$

Note that we have $cov(\alpha_k + \widehat{\delta_k}Y + \widehat{\epsilon_k}, \alpha_l + \widehat{\delta_l}Y + \widehat{\epsilon_l}) = cov(\widehat{\delta_k}Y, \widehat{\delta_l}Y) + cov(\widehat{\epsilon_k}, \widehat{\epsilon_l})$ 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 $cov(\widehat{\delta_k}Y, \widehat{\delta_l}Y) = \widehat{\delta_k}\widehat{\delta_l}cov(Y,Y) = \widehat{\delta_k}\widehat{\delta_l}var(Y)$ 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:

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

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.