Measuring Work From Home in the Cross-Section

By Augustus Kmetz, John Mondragon, and Johannes F. Wieland*

The shift to work from home (WFH) has been a large and persistent consequence of the pandemic. To quantify the effect of WFH on the macroeconomy, researchers have exploited the fact that local labor markets are differentially exposed to this shock, either in empirical or quantitative spatial settings. See for example, Althoff et al. (2022), Brueckner, Kahn and Lin (2021), Delventhal, Kwon and Parkhomenko (2021), Gupta et al. (2021), Gupta, Mittal and Van Nieuwerburgh (2022), Haslag and Weagley (2021), Kmetz, Mondragon and Wieland (2022), Liu and Su (2021), Mondragon and Wieland (2022), Ramani and Bloom (2021), among others.

These analyses require a measure of WFH at disaggregated levels. In this paper we compare several important measures used in the literature: Barrero, Bloom and Davis (2021), Bick, Blandin and Mertens (2022), Dingel and Neiman (2020), and the American Community Survey (ACS).[1] While these measures differ in how comprehensively they measure WFH (e.g., they may or may not include hybrid work), we show that they are highly correlated in the cross-section. Therefore, these measures will yield similar causal effects once appropriately scaled by the average level of WFH.

We argue that, when choosing a particular measure, researchers should carefully consider the trade-off between how comprehensively WFH is measured and measurement error in the survey at the particular level of geographic aggregation. The large sample size of the ACS makes it uniquely suitable for analysis at micro- and metropolitan levels or even finer levels of aggregation, but it only captures hybrid WFH indirectly. When measuring different types of WFH is important or the cross-sectional analysis can be done at the state or large-city level, then the surveys by Barrero, Bloom and Davis (2021) and Bick, Blandin and Mertens (2022) may be preferable.

I. Measuring Work from Home

We investigate the measurement of remote work in three distinct surveys: the ACS, the Survey of Working Arrangements and Attitudes (SWAA), and the Real-Time Population Survey (RPS).

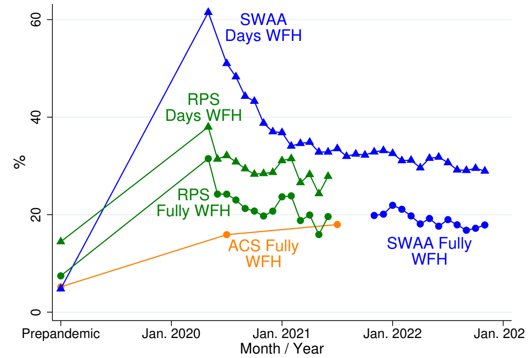
Each year, the ACS reaches around 4.5 million people split among 3.5 million households. Respondents are first sent mailings that allow them to fill out the online survey. Those unable to fill out the online form can return the questionnaire by mail or may be called or visited by a Census employee. WFH is measured using the subject's employment status and transportation to work response: an employed person that reports not having to commute to work is classified as WFH. This is a coarse measure of WFH that likely only captures full-time WFH (since hybrid workers do commute).

The richness of the ACS data also allows an easy mapping from the Dingel and Neiman (2020) teleworkability classifications, which gives another measure of WFH suitability, as well as many demographic and household observables. The large sample size also means the ACS provides decent coverage at low levels of aggre-

* Kmetz: Federal Reserve Bank of San Francisco, 101 Market St., San Francisco, CA 94105, U.S.A., gus.kmetz@sf.frb.org. Mondragon: Federal Reserve Bank of San Francisco, 101 Market St., San Francisco, CA 94105, U.S.A., john.mondragon@sf.frb.org. Wieland: UCSD & NBER, 9500 Gilman Dr. #0508, La Jolla, CA 92093-0508, U.S.A., jfwieland@ucsd.edu.
We thank Jose Barrero, Alex Bick, Adam Blandin, Nick Bloom, Steven Davis, Karel Mertens, and Michael Starsinic for helpful comments. The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

[1] We focus solely on WFH in the U.S. Aksoy et al. (2022), Alon et al. (2022), and Hansen et al. (2022) provide aggregate measures of WFH in other countries.

FIGURE 1. AGGREGATE MEASURES OF WORKING FROM HOME



Note: The SWAA pre-pandemic data point is from the American Time Use Survey (Barrero, Bloom and Davis, 2021). The ACS is reported at an annual frequency so the 2020 WFH share also contains pre-pandemic data.

gation, such as micropolitan areas.[2]

The SWAA, developed by Barrero, Bloom and Davis (2021), is an ongoing monthly survey administered online by Inc-Query and QuestionPro. It receives around 7000 responses a month. WFH is measured using the share of full working days the employee worked from home in the reference week. This allows one to measure both hybrid and fully-remote work and, along with weights based on the Current Population Survey (CPS), allows for nationally-representative measures as well as rich coverage of attitudes, experiences, and expectations regarding WFH.

The Real-Time Population Survey, developed by Bick and Blandin (2022) and Bick, Blandin and Mertens (2022), was a bi-weekly online survey administered using the Qualtrics panel. WFH days are measured as working days in the reference week that do not involve a commute. The survey was designed to match the February 2020 CPS along several demographic characteristics and surveyed around 2000 individuals a month until it was discontinued in June 2021. The RPS provides a detailed look at the intensive margin of WFH, respondent demographics, and nationally-representative weights based on the CPS.

Figure 1 plots the percent of WFH days and the percent of employees fully WFH from Bick, Blandin and Mertens (2022) and Barrero, Bloom and Davis (2021) as well as the percent of employees WFH from the ACS (Mondragon and Wieland, 2022). All three surveys show quantitatively large increases in WFH and high subsequent persistence across all measures. The differences in magnitudes largely reflect differences in measurement. The ACS consistently displays the lowest degree of WFH because it likely captures only full-time WFH. When comparing the percent of full-time WFH, all three surveys agree that the fraction of full-time WFH employees is around 20% in 2021.[3] Similarly, by 2021 both the SWAA and the RPS show that the fraction of WFH days is stable at around 30%. In short, the surveys paint a quantitatively consistent picture in the time series. Our objective is to assess whether this is also true in the cross-section.

II. All Measures are Highly Correlated in the Cross-Section

Using the SWAA and the RPS we construct state-level measures of the fraction of

[2]The public-use microdata use the "PUMA" geography, which is easily mapped to other geographies. Researchers with access to confidential Census data would, in principle, observe census blocks.

[3]Barrero et al. (2023) compare the SWAA WFH time-series with the ACS and other surveys. They also show that much of the remaining difference between the SWAA and the ACS reflects sample design. We do not make adjustments to homogenize the sample, as we are interested in measuring the common variation of the headline WFH measures in the cross-section.

days worked from home in 2021, and using the ACS we construct state-level measures of the fraction of fully WFH employees.[4]. We also construct a measure of the fraction of jobs that may be done from home using Dingel and Neiman (2020) and occupational code mappings into the ACS. Figure 2 constructs ventiles of the ACS 2021 WFH share and then plots this against the the average of each survey's measure within the same bin, as well as regression lines from the underlying data.

As expected, the ACS tends to understate WFH relative to the other measures, but the most important takeaway is that all of the measures comove very strongly with the ACS. While there are some noticeable differences within particular bins, the slopes of the regression lines are effectively indistinguishable. This suggests that, at the state level at least, any of the measures can be used to reasonably capture cross-sectional variation in remote work.

We summarize these relationships quantitatively in Table 1, which reports the correlations of the same 2021 state-level measures. All measures are strongly correlated, with no correlation below 0.61. Some of the disagreement reflects measurement error: when we pool the SWAA and the RPS over the entire sample rather than 2021 alone, then we find that the lowest correlation is 0.74.[5] The Dingel and Neiman (2020) measure applied to the ACS and ACS WFH measure are very strongly related with a correlation of 0.96. This is interesting since the ACS WFH measure only captures fully remote work, while the potential WFH measure can theoretically capture the entire spectrum of WFH arrangements. Ultimately, the strong correlations across the board show that any of the measures will reflect similar underlying variation in actual WFH across states.

III. Trading Off Comprehensiveness and Measurement Error

We argue that which survey a researcher should use depends on the trade-off between how comprehensively a survey measures WFH and the precision of the measurement at a particular level of aggregation. We show how measurement error of WFH depends on sample size and the level of aggregation. For each $i = 1, ..., N$ regional units of observation we estimate the measurement error σ_i for the local remote share, average it across all units, and normalize it using the national remote share S,

$$\text{Normalized error} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_i}{S}.$$

By normalizing the error to the national level we adjust for the fact that dispersion in WFH is increasing in the level, thus making estimates across different years more comparable.

For the SWAA, we implement a standard bootstrap with $B = 250$ draws for each month so that $\sigma_{im} = \frac{1}{B-1}\sum_{j=1}^{B}(S_{ijm} - \hat{S}_{im})$ where $\hat{S}_{im}$ is the survey's monthly estimate of WFH, using survey weights. We then average across months to get an annual estimate.

This procedure must be adjusted to recover correct errors in the RPS as the survey is raked to generate weights that match a set of moments from the CPS.[6] To correctly calculate errors, we incorporate the non-linear raking procedure by bootstrapping the sample and then raking the bootstrapped samples to match the same set of moments in the monthly CPS. We then calculate annual errors by averaging the monthly estimates.

For the 2015-19 and 2021 ACS we use the replicate weights provided by the Census to recover errors directly. Finally, we provide

[4]Monthly surveys are averaged.

[5]The correlation with the 2021 ACS is slightly smaller for 2021 full-time WFH in the SWAA (0.68 when backfilled) and in the RPS (0.6). This suggests that the distinction between full-time and hybrid WFH has greater measurement error than the headline WFH rate, at least at the state-level at an annual frequency. This reinforces our claim that researchers need to carefully consider the trade-off between detailed measurement and measurement error in their analysis.

[6]Raking is a standard iterative procedure of generating weights for a sample that will match a set of target moments (Kolenikov, 2014).

FIGURE 2. BINNED SCATTER PLOT OF WORKING FROM HOME SHARE ACROSS U.S. STATES, 2021

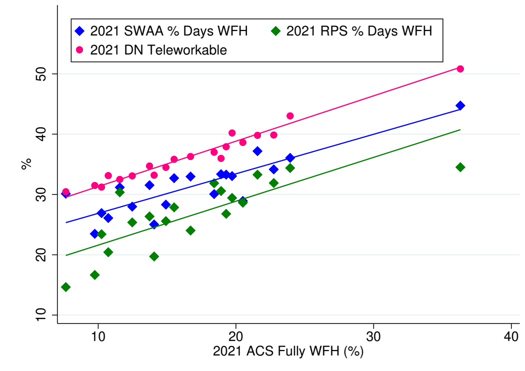


TABLE 1—CORRELATION OF REMOTE WORK ACROSS STATES (N=51), 2021

Variables	ACS	SWAA	RPS	DN
ACS	1.00			
SWAA	0.71	1.00		
RPS	0.67	0.61	1.00	
DN	0.95	0.68	0.64	1.00

two sets of error estimates for the 2020 ACS due to the experimental nature of the 2020 replicate weights (Rothbaum et al., 2021). Our first approach follows standard ACS procedure and uses the experimental replicate weights as given. Our second estimate implements a bootstrap ($B = 250$) that ignores the replicate weights and samples from the baseline survey directly.

Figure 3 plots the normalized errors against the log sample size of each survey. Blue scatter points show measurement errors at the state level and orange scatter points show measurement errors at the CBSA level (ACS only). Measurement errors overall are modest with even the least precise survey having errors equivalent to just 14% of the aggregate WFH rate. Second, all but the 2020 ACS replicate-based estimates clearly lie on a downward-sloping curve where increases in sample size reduce measurement error. Due to its much larger size, the ACS provides the most precise measures of WFH.

We also see, from the orange scatter points, that the ACS's CBSA-level measures are less precise than the ACS state-level measures due to the lower level of aggregation, but they are approximately as precise as the RPS and SWAA measures at the state level. This suggests that the main tradeoff in choosing between the ACS and the RPS or the SWAA is whether the analysis requires measuring WFH at low levels of aggregation, which favors the ACS, or whether the analysis requires a precise measurement of different types of WFH and/or working arrangements, which favors the SWAA and RPS.

Finally, the 2020 ACS estimates using the experimental replicates are significantly below our bootstrapped estimates and generally seem to be very low, with 2020 CBSA-level estimates having comparable precision to 2021 ACS state-level estimates despite the different levels of aggregation. Since the replicate-based estimates seem to lie below the curve traced out by the other surveys, we believe these estimates are too low. The fact that the bootstrapped errors shift back up to the other ACS-based estimates suggest that the experimental replicates did not properly account for sampling error and it may be more appropriate for researchers to rely on a bootstrap to estimate errors.[7]

IV. Conclusion

We compare several important measures of WFH in the cross-section of the United States: the SWAA (Barrero, Bloom and Davis, 2021), the RPS (Bick, Blandin and Mertens, 2022), Dingel and Neiman (2020) implemented in the ACS, and the ACS itself. We show that at the state level these measures are highly correlated. Therefore these measures will yield similar causal effects once appropriately scaled by the average level of WFH. We argue that the main criterion in choosing a particular source for cross-sectional WFH is the trade-off between how comprehensively WFH is measured and the measurement error in the survey at a particular level of aggregation. Finally, we show that the experimental replicates for the 2020 ACS likely understate the error in the survey, but that a bootstrap approach likely recovers correct standard errors which are comparable to other years of the survey.

REFERENCES

Aksoy, Cevat Giray, Jose Maria Barrero, Nicholas Bloom, Steven J Davis, Mathias Dolls, and Pablo Zarate. 2022. "Working from home around the world." *mimeo.*

Alon, Titan, Sena Coskun, Matthias Doepke, David Koll, and Michèle Tertilt. 2022. "From mancession to shecession: Women's employment in regular and pandemic recessions." *NBER Macroeconomics Annual*, 36(1): 83–151.

Althoff, Lukas, Fabian Eckert, Sharat Ganapati, and Conor Walsh. 2022. "The geography of remote work." *Regional Science and Urban Economics*, 93: 103770.

[7]The 2020 ACS replicates were generated using a new raking procedure to attempt to correct for the extraordinary changes in nonresponse bias throughout the year (Rothbaum et al., 2021). The fact that a simple bootstrap gives plausible estimates of the standard error suggests that the nonresponse bias in the survey did not have a serious impact on the survey's power.

Barrero, Jose Maria, Nicholas Bloom, and Steven J Davis. 2021. "Why working from home will stick." *mimeo.*

Barrero, Jose Maria, Nicholas Bloom, Shelby Buckman, and Steven J Davis. 2023. "Benchmarking SWAA Estimates of the Prevalence of Working From Home." *mimeo.*

Bick, Alexander, Adam Blandin, and Karel Mertens. 2022. "Work from home after the COVID-19 Outbreak." *mimeo.*

Bick, Alexander, and Adam Blandin. 2022. "Employer Reallocation During the COVID-19 Pandemic:Validation and Application of a Do-It-Yourself CPS." *mimeo.*

Brueckner, Jan, Matthew E Kahn, and Gary C Lin. 2021. "A new spatial hedonic equilibrium in the emerging work-from-home economy?" *mimeo.*

Delventhal, Matthew J, Eunjee Kwon, and Andrii Parkhomenko. 2021. "JUE Insight: How do cities change when we work from home?" *Journal of Urban Economics*, 103331.

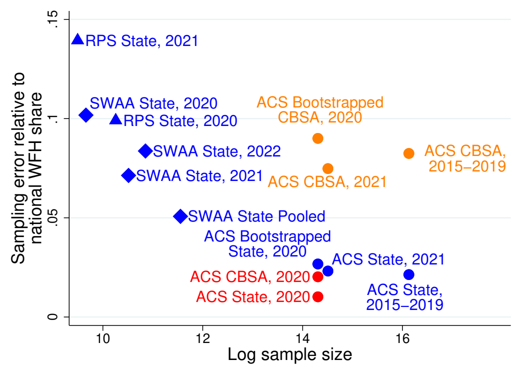
Dingel, Jonathan I, and Brent Neiman. 2020. "How many jobs can be done at home?" *Journal of Public Economics*, 189: 104235.

Gupta, Arpit, Vrinda Mittal, and Stijn Van Nieuwerburgh. 2022. "Work From Home and the Office Real Estate Apocalypse." *mimeo.*

Gupta, Arpit, Vrinda Mittal, Jonas Peeters, and Stijn Van Nieuwerburgh. 2021. "Flattening the curve: pandemic-induced revaluation of urban real estate." *Journal of Financial Economics.*

Hansen, Stephen, Peter John Lambert, Nick Bloom, Steven J Davis, Raffaella Sadun, and Bledi Taska. 2022. "Remote Work across Jobs, Companies, and Countries." *mimeo.*

Figure 3. Sampling Error of Working From Home Share



Haslag, Peter H, and Daniel Weagley. 2021. "From LA to Boise: How migration has changed during the COVID-19 pandemic." *mimeo.*

Kmetz, Augustus, John Mondragon, and Johannes F Wieland. 2022. "Remote Work and Housing Demand." *FRBSF Economic Letter*, 2022(26): 1–5.

Kolenikov, Stanislav. 2014. "Calibrating survey data using iterative proportional fitting (raking)." *The Stata Journal*, 14(1): 22–59.

Liu, Sitian, and Yichen Su. 2021. "The impact of the Covid-19 pandemic on the demand for density: Evidence from the US housing market." *Economics letters*, 207: 110010.

Mondragon, John A, and Johannes Wieland. 2022. "Housing Demand and Remote Work." *mimeo.*

Ramani, Arjun, and Nicholas Bloom. 2021. "The Donut effect of COVID-19 on cities." *mimeo.*

Rothbaum, Jonathan, Jonathan Eggleston, Adam Bee, Mark Klee, and Brian Mendez-Smith. 2021. "Addressing nonresponse bias in the american community survey during the pandemic using administrative data." *American Community Survey Research and Evaluation Report Memorandum Series ACS21-RER-05, US Census Bureau, Washington.*