WHITE PAPER

Characteristics of Workers in Low Work-From-Home and High Personal-Proximity Occupations

Simon Mongey and Alex Weinberg

MARCH 2020

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Characteristics of workers in low work-from-home and high personal-proximity occupations

Simon Mongey* Alex Weinberg†

March 2020

Abstract

We categorize occupations by a measure that captures the likelihood that jobs can be conducted from home (Dingel and Neiman, 2020), as well as a measure of low personal proximity in the workplace. The former relates to how well work can be done under social distancing policies, the latter relates to how quickly occupations might come back online. We then compare characteristics of workers in low work-from-home and high personal-proximity occupations. Relative to workers in high work-from-home occupations, workers in low work-from-home occupations are less likely to be white, have a college degree, or have employer provided healthcare, more likely to be in the bottom half of the income distribution, and more likely to rent their homes. These workers are less likely to have access to *informal insurance channels*: more likely to be single, and less likely to be born in the United States. They are also less likely to have had *stable jobs*: more likely to have been unemployed in the last year, less likely to be employed full-time, and less likely to be employed in large firms. Females are both more likely to be in high work-from-home occupations and more likely to work in high physical-proximity occupations, suggesting that the employment effects of broad social distancing policies on women may be less severe, but later integration into the economy may be more difficult.

Keywords: Coronavirus, occupations, demographics

*Kenneth C. Griffin Department of Economics, University of Chicago and NBER: mongey@uchicago.edu.
†Kenneth C. Griffin Department of Economics, University of Chicago: weinberga@uchicago.edu.

1 Introduction

Absent a vaccine or widespread testing, a key policy implemented during the Coronavirus pandemic is 'social distancing' which requires workers in may jobs to work from home if feasible. Moreover, returning to work will likely occur more slowly for jobs that require a large degree of personal proximity to others. To the extent that jobs that cannot be conducted from home, or have high personal proximity, feature workers that are systematically different to workers in jobs that can be conducted from home, or have low personal proximity these policies will have systematically different effects across groups of individuals. Understanding how individuals vary across these occupations is therefore important for targeting economic policies designed to assist workers.

In this short paper, we combine data sources to study this systematic variation in individuals across occupations. We merge the Bureau of Labor Statistics' *Current Population Survey* (CPS) with a version of the Dingel and Neiman (2020) classification of occupations' capacity to work from home and add a measure of personal proximity in the workplace.[1] We construct these two measures using data from the Department of Labor's *Occupational Information Network* (O*NET) data. In these occupation-level data, occupational classifications are finer than those available in the individual-level data. To merge make the data conformable we use the Bureau of Labor Statistics' *Occupation Employment Statistics* (OES) to employment weight O*NET measures within the coarser occupations defined in the CPS. We then project individual characteristics onto indicators of whether occupations rank above or below the employment weighted medians of these measures.

Other papers. Our contribution is simply to look at the *characteristics of workers* in these different occupations. Dingel and Neiman (2020) use the OES to ask the important question of what fraction of employment and income is accounted for by high work-from-home occupations. Leibovici et al. (2020) conduct a similar analysis, instead considering low personal-proximity occupations rather than high work-from-home occupations.[2] Both use the O*NET to classify occupations, and then employment and income data from the OES to study the geographic distribution of jobs and the fraction of employment and income accounted for by types of jobs. We add the CPS to understand how workers differ across these jobs.

Result. Our results can be broadly summarized as follows: workers in occupations that—according to these two metrics—are more likely to be affected by social distancing policies are workers that we would consider economically more vulnerable. Workers in these occupations are less likely to have a college degree and are less likely to have health insurance provided by their employer. They are less likely to be white, less likely to work at a large firm, and less likely to be born in the USA.

[1] See https://bfi.uchicago.edu/wp-content/uploads/BFI_White-Paper_Dingel_Neiman_3.2020.pdf

[2] St Louis Federal Reserve, *On the Economy* blog: https://www.stlouisfed.org/on-the-economy/2020/march/social-distancing-contact-intensive-occupations

These results are monotonic in both measures work-from-home measure. When we compare the top quartile of occupations by this measure, to the bottom quartile of occupations by this measure, we find that these patterns are more severe. When we compare the third quartile of occupations by this measure, to the second quartile of occupations by this measure, we find that these patterns are slightly less severe but in all cases still statistically significant. Occupations that score relatively lower (higher) in terms of the work-from-home (personal-proximity) measure, are even more economically vulnerable.

These relationships are in general stronger when we split occupations by the work-from-home measure. With regards to personal-proximity in the workplace, these patterns still exist but are weaker. Relative to occupations that have a high work-from-home measure, there is greater economic diversity among occupations associated with low levels of personal proximity. For example, both sales assistants and dentists work in high personal-proximity environments. This is suggestive that economic costs of social distancing may be more tightly related to pre-Coronavirus economic status, while the economic costs of a slow return to work that starts with low personal-proximity occupations may be more broadly distributed.

Overview. Section 2 describes the data sources, our construction of measures of work-from-home and personal-proximity using the O*NET and OES data, and how we summarize differences in worker characteristics across occupations using the CPS. Section 3 gives our main results, which are summarized in Figure 1 (work-from-home) and Figure 2 (personal-proximity). Section 4 concludes.

2 Method

We describe how we construct measures of work-from-home and personal-proximity, and then how we merge these with data at the worker-level to understand how individual characteristics vary across these occupations.

2.1 Characteristics of jobs

We merge three sets of data:

O*NET - Occupation-level data on work activities by occupation, where occupations are defined at the fine SOC level.

CPS - Individual-level data containing information on various individual characteristics as well as occupation, coded using coarser Census OCC codes. Throughout we use the IPUMS harmonized (*occ2010*) codes. We use the same sample selection criteria as Heathcote et al. (2010).[3]

[3] In particular we apply the same criteria as used to obtain their Sample C.

OES - Contains economy-wide employment by occupation at the SOC level, which allows us to apply employment weights when aggregating skills across SOC level occupations, within OCC level occupations.

Rather than pooling data over time we take the most recent snapshot of the US economy available to us concurrently from these data, which is 2018. We use O*NET database 24, OES data from 2018, and the 2019 March CPS which asks questions regarding occupation and income in the prior year.

Using the O*NET and OES data we construct two measures for each occupation that are signed in terms of expected negative economic impacts of the crisis: (i) *low work-from-home* ($LWFH$), and (ii) *high personal-proximity* (HPP). We assign these to occupations based whether an occupation ranks high or low on a measure that we construct from the O*NET, and then use the CPS to describe the average attributes of workers in occupations that rank high or low on these measures.

We detail how we construct $LWFH$ and then describe the construction of HPP which follows many of the same steps. Let $j \in \{1, \ldots, J\}$ denote a 3-digit OCC-code occupation, which is the measure available in the CPS. Let $l \in \{1, \ldots, L\}$ denote the fine SOC-code categorization of occupations in O*NET and OES. We differ from Dingel and Neiman (2020) in how we aggregate skills, but use their set of O*NET job characteristics.

1. We take the following 18 measures of occupation attributes in the O*NET data from Dingel and Neiman (2020), which we index by $k = 1, \ldots, K$, and take on values $m_{lk} \in [1, 5]$:

    - **Measures from the Work Activities module**: Performing General Physical Activities, Inspecting Equipment, Structures, or Material; Handling and Moving Objects; Controlling Machines and Processes; Operating Vehicles, Mechanized Devices, or Equipment; Performing for or Working Directly with the Public; Inspecting Equipment, Structures, or Material; Repairing and Maintaining Electronic Equipment; Repairing and Maintaining Mechanical Equipment.

    - **Measures from the Work Contexts module** Electronic Mail; Outdoors, Exposed to Weather; Outdoors, Under Cover; Deal With Physically Aggressive People; Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection; Wear Common Protective or Safety Equipment such as Breathing Apparatus Safety Harness, Full Protection Suits, or Radiation Protection; Spend Time Walking and Running; Exposed to Minor Burns, Cuts, Bites, or Stings; Exposed to Disease or Infections.

    In some cases we reverse the values of the variable such that a *high value* implies that the occupation would be better suited to working from home. We convert these into binary variables $\widetilde{m}_{lk} \in \{0, 1\}$ based on whether $m_{kl} \geq 4$.

2. Within each 3-digit OCC occupation j, we take the weighted average of $\widetilde{m}_{kl}$ measures across SOC occupations l. We use employment from the OES at the SOC level (n_l) to construct the employment weighted average. This gives a measure for each occupation-attribute pair: $x_{jk} = \sum_{l \in j} \omega_l \widetilde{m}_{lk}$, where $\omega_l = n_l / \sum_{l' \in j} n_{l'}$. To map SOC code occupations into OCC code occupations we use a cross-walk that is initially from US Census, but with substantial editing and verification on our part.[4]

3. For each of these measures k we rank occupations $j = \{1, \ldots, J\}$ on x_{jk} and assign the dummy variable $z_{jk} = 1$ if the occupation j is above the employment weighted median value of x_{jk}.

4. We then construct a single measure for each occupation $\overline{z}_j$ by taking the unweighted mean of z_{jk}: $\overline{z}_j = K^{-1} \sum_{k=1}^{K} z_{jk}$. This gives the fraction of the 18 measures for which occupation j ranks above the employment weighted median.

5. We assign the dummy variable $LWFH_j = 1$ (*low work-from-home*) if occupation j is *below* the employment weighted median value of $\overline{z}_j$. We later compare occupations by quantiles of $\overline{z}_j$.

This produces a binary variable $LWFH_j$ for each occupation that we can map into the occupational codes contained in the CPS.

The procedure to construct HPP_j is similar to the above. We start with a measure from O*NET at the SOC level that reflects personal-proximity at work and takes on a value in $[1, 5]$. We then convert this into a binary variable by assigning a value of 1 if the O*NET measure is at least 4, which reflects proximity of arm's length or less; the appropriate level to consider here.[5] We take employment weighted means of this across SOC-level occupations to construct an OCC-level occupation measure. We then assign the dummy $HPP_j = 1$ (*high personal-proximity*) if the occupation is *above* the employment weighted median of this variable.

By construction HPP_j and $LWFH_j$ are binary variables that equal 1 for the occupations that are *most* likely to be effected by the epidemic and ensuing policies. Half of employment is in $HPP_j = 1$ jobs and half of employment is in $LWFH_j = 1$ jobs.

2.2 Characteristics of workers

With $LWFH_j$ and HPP_j measured for occupations that are consistent with the CPS, we can compare the characteristics of workers in occupations for which these measures are high and for

[4] The basic cross-walk from Census is available here: https://www2.census.gov/programs-surveys/demo/guidance/industry-occupation/2010-occ-codes-with-crosswalk-from-2002-2011.xls

[5] Workers that respond to the survey administered by O*NET choose one of: 1 = 'I don't work near other people (beyond 100ft)', 2 = 'I work with others but not closely (e.g. private office)', 3 = 'Slightly close (e.g. shared office)', 4 = 'Moderately close (at arm's length)', 5 = 'Very close (near touching)'. Publicly available O*NET data consists of an average of these responses. In assigning $HPP_j = 1$ to occupations for which the average is at least 4. Our high personal-proximity occupations therefore represent occupations for which the average respondent said they worked at arms length or less away from others. For additional information regarding this question: https://www.onetonline.org/find/descriptor/result/4.C.2.a.3.

which these measures are low.

Our approach is simple. Let y_{ij} be a characteristic of a worker i that reports that they mostly worked in occupation j last year.[6] We only consider binary variables in the CPS; for example we construct a variable $y_{ij} = 1$ if the continuous variable 'annual income' is above the median. We then run the following simple regression for each of our observables, using $LWFH_j$ as an example:

$$y_{ij} = \alpha_y + \beta_y LWFH_j + \varepsilon_{ij}. \tag{1}$$

We then plot the values for $\widehat{\beta}_y$. This sample moment gives

$$\widehat{\beta}_y = \left[y_{ij} \middle| LWFH_j = 1\right] - \mathbb{E}\left[y_{ij} \middle| LWFH_j = 0\right] \quad ,$$

where $\mathbb{E}$ is the sample mean. Given that y_{ij} is binary, $\widehat{\beta}_y$ simply gives the fraction of workers for which $y_{ij} = 1$ in *low* work-from-home occupations, relative to the fraction of workers for which $y_{ij} = 1$ in *high* work-from-home occupations. Clearly $\widehat{\beta}_y$ is between -1 and 1, taking the value of 1 when $y_{ij} = 1$ for all individuals for which $LWFH_j = 1$, and $y_{ij} = 0$ for all individuals for which $LWFH_j = 0$.

Comparing estimates across measures y and y', a higher value of $\widehat{\beta}_y > \widehat{\beta}_{y'}$ can be interpreted as

> *"Workers in occupations for which $LWFH_j = 1$ are relatively more different from workers in occupations for which $LWFH_j = 0$ along dimension y than along dimension y'".*

3 Results

We estimate (1) for a number of individual characteristics. In each case we assign $y_{ij} = 1$ to the individuals with the characteristic that we find to be most related to being in a low work-from-home occupation. This gives $\widehat{\beta}_y \in [0, 1]$. With this approach, we have the following characteristics of workers, all of which take on a value $y_{ij} = 1$:

- **Demographics.** (i) Non-white, (ii) No college degree, (iii) Age below 50, (iv) Male, (v) Single, (vi) Born outside USA, (vii) Non-US citizen, (viii) Rent their home

- **Work.** (i) No healthcare provided by employer,[7] (ii) Works at a small firm (< 500 employees), (iii) Part-time employed

- **Income.** (i) Annual income is below the median individual income, (ii) Experienced a spell of unemployment in the last year.

Work and income characteristics are associated with the job at which the worker was employed for the longest period of time in 2018.

[6] In the IPUMS coding of the CPS this is *occ2010ly*.

[7] We set the indicator for employer provided healthcare to 1 if the employer pays for part or all of the individual's health insurance premiums.

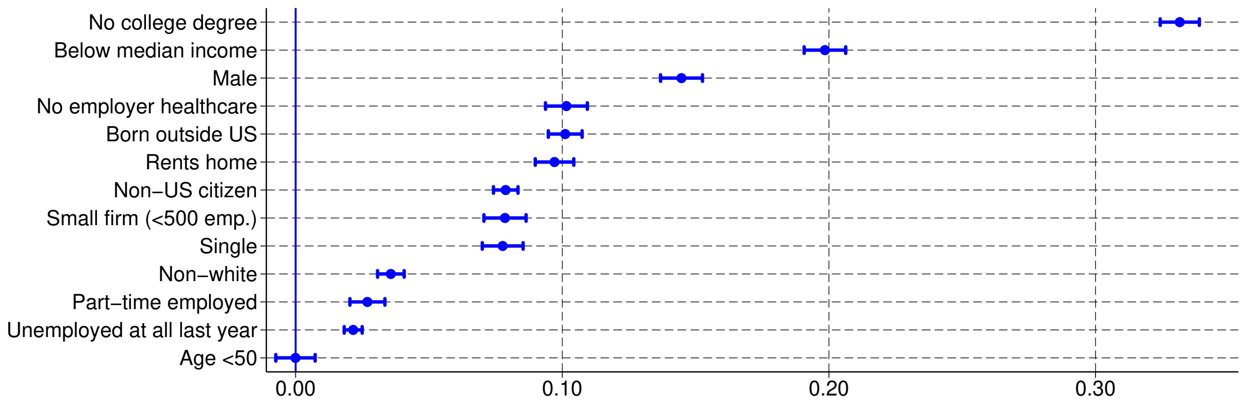

Figure 1: Characteristics of workers in *Low Work-from-home* occupations

Notes This figure plots estimates of $\widehat{\beta}_y$ for 10 characteristics y from regressions in which $WFH_j \in \{0,1\}$ is the independent variable. A high value means that workers in *low work-from-home occupations* are more likely than workers in *high work-from-home occupations* to be in the category listed on the vertical axis. Point estimates are given by the circle markers, and 95 percent confidence intervals are given by the lines through each marker.

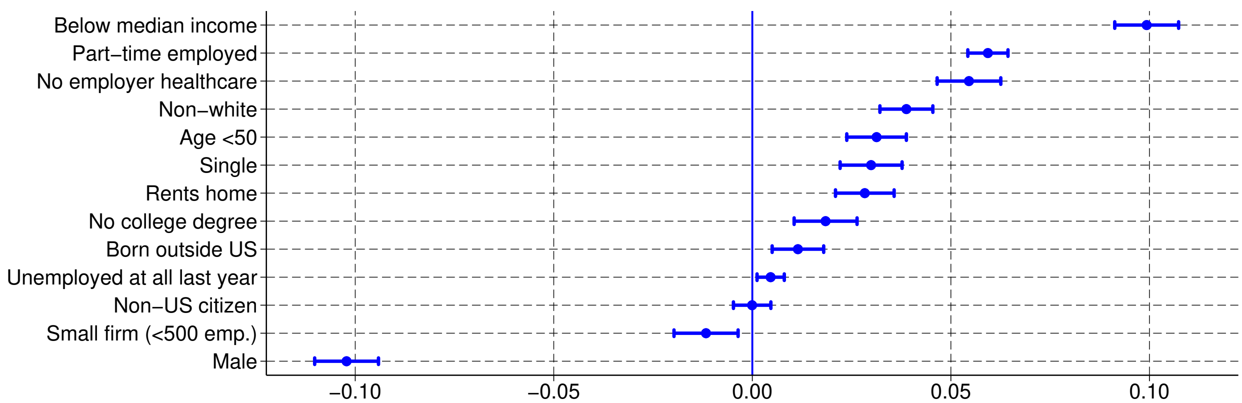

Figure 2: Characteristics of workers in *High Personal-proximity* occupations

Notes This figure plots estimates of $\widehat{\beta}_y$ for 10 characteristics y from regressions in which $PP_j \in \{0,1\}$ is the independent variable. A high value means that workers in *high personal-proximity occupations* are more likely than workers in *low personal-proximity occupations* to be in the category listed on the vertical axis. Point estimates are given by the circle markers, and 95 percent confidence intervals are given by the lines through each marker.

We plot the estimates for each of these characteristics for the low work-from-home regression in Figure 1, ordering these attributes from the highest to the lowest point estimate. Figure 2 repeats the exercise for the high personal-proximity regression. We discuss low work-from-home first.

Economic and demographic. Occupations that score low in terms of the work-from-home measure feature workers that by all measures are economically more vulnerable. Workers in these occupations are less likely to be white or to have a college degree, which relate to the fact that they are also more likely to have below median income. They are more likely to work in smaller firms, which are on average less financially robust and so less likely to remain in operation after the crisis.

They are more likely to rent rather than own their homes.

Workers in these occupations are also less likely to have access to informal insurance channels that may help them weather the crisis. They are less likely to be married, which diversifies household income against individual income risk. They are less likely to be US citizens or born in the US, which may lead to less family support, as well as restricted access to emergency government programs.

Workers in low work-from-home occupations are more likely to have unstable employment. They are less likely to be employed full-time and more likely to have recently experienced unemployment.

Healthcare. Availability of healthcare is a key insurance mechanism in a health crisis. Workers in occupations that are less readily able to be performed from home are less likely to have employer-provided healthcare. Meanwhile those in jobs that are more readily able to be performed from home are more likely to have employer provided healthcare.

Age. The mortality rate for those with COVID-19 is significantly higher for older individuals. However, we find that the age of workers across these high- and low- work-from-home occupations does not systematically differ.

High physical proximity. We find that along most of the individual characteristics the results for high work-from-home occupations and low personal-proximity occupations are the same in terms of their sign. For example, workers in both high personal-proximity occupations and low work-from-home occupations are less likely to have a college degree than workers in low personal-proximity and high work-from-home occupations, respectively. The results, however, are less stark, as evidenced by the magnitudes of the coefficients. Differences in workers across high and low personal-proximity occupations are less pronounced than the differences between workers in low and high work-from-home occupations. If we consider high personal proximity occupations to be slower to be brought back as social distancing policies unwind, then this suggests that the slow recovery may be more broadly felt than the more concentrated effects of broad social distancing.

Sex. The results differ across these two measures most sharply for sex. Individuals in occupations that score highly in terms of work-from-home are more likely to be women, but this is also true for occupations that have high personal-proximity. Taking these results at face value, female workers may be less affected by the universal social distancing measures currently in place, but could be more affected in the future as these restrictions are targeted toward occupations with higher personal-proximity.

Robustness - Alternative splits. If instead of splitting occupations above and below the median we consider more or less extreme comparisons of occupations, we find that our results vary monotonically. In constructing Figure 1 we assigned the dummy variable $LWFH_j$ based on whether

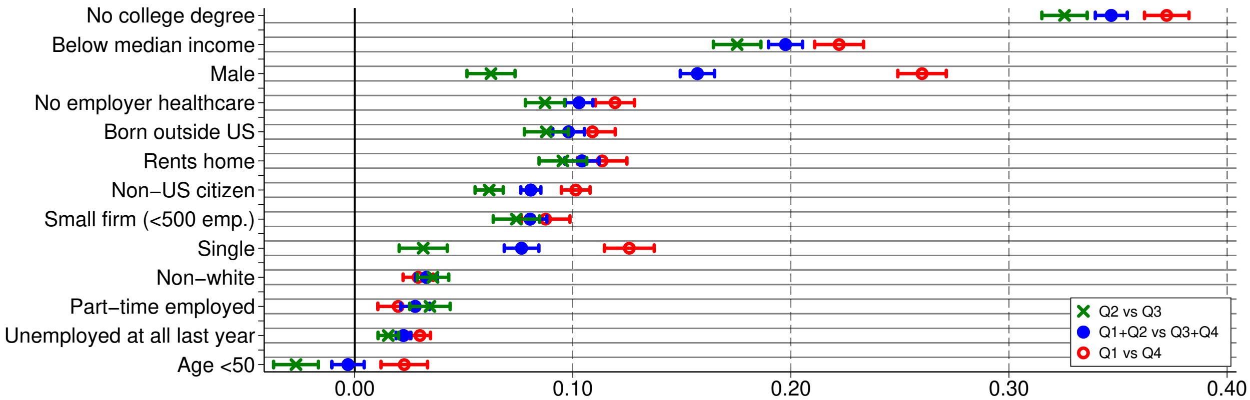

Figure 3: Comparing different groups of occupations on the Work-from-home measure

Notes This figure extends Figure 1. The red markers replicate Figure 1. In constructing the estimates plotted in green, we set $WFH_j = 0$ for the second quartile of our continuous measure, and $WFH_j = 1$ for the third quartile of our measure. In constructing the estimates plotted in blue, we set $WFH_j = 0$ for the first quartile of our continuous measure, and $WFH_j = 1$ for the fourth quartile of our measure.

the occupation was below or above our the employment weighted median of our index $\overline{z}_j$. Figure 3 compares the lower quartile to the upper quartile (dropping the middle quartile, in red), and the second quartile to the third quartile (dropping the upper and lower quartile, in green). When we compare the very low work-from-home occupations to the very high work-from-home occupations (in red), coefficients are uniformly larger in magnitude, and uniformly smaller when comparing moderately low work-from home occupations to moderately high work-from-home occupations (in green).

4 Conclusion

We show that workers systematically differ across the types of occupations that are most likely to be hit by the social distancing and stay-at-home policies required to stop the spread of the Coronavirus. Workers in occupations that are *most likely* to be affected—those with a low score in the work-from-home measure proposed by Dingel and Neiman (2020), or a high score in the O*NET measure of personal-proximity—are predominantly characterized by traits associated with the more economically vulnerable in the US economy. These workers are more likely to be less educated, have limited healthcare, and be toward the bottom of the income distribution.

This simple approach can be extended to individual economic indicators in any microdata that records occupation. An obvious example would be individual level data on wealth and asset portfolios. We therefore conclude by appealing to those at the Federal Reserve Board with access to the confidential *Survey of Consumer Finance* (SCF) files to extend this analysis. The publicly available SCF contains data on individual wealth and asset portfolios but does not include *occupation*, while the confidential records include both.

References

Dingel, Jonathan and Brent Neiman, "How Many Jobs Can Be Done At Home?," BFI White Paper, Becker Friedman Institute 2020.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante, "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006," *Review of Economic Dynamics*, January 2010, *13* (1), 15–51.

Leibovici, Fernando, Ana Maria Santacrue, and Matthew Famiglietti, "Social Distancing and Contact-Intensive Occupations," Technical Report, St. Louis Federal Reserve Bank - On the Economy Blog March 2020.