

An alternate to survey methods to measure work from home

Mohit Sharma, Sargam Gupta and Xavier Estupinan



**Indira Gandhi Institute of Development Research, Mumbai
September 2020**

An alternate to survey methods to measure work from home

Mohit Sharma, Sargam Gupta and Xavier Estupinan

[Email\(corresponding author\): mohit.mse@gmail.com](mailto:mohit.mse@gmail.com)

Abstract

*During unprecedented times like COVID-19, when 'social distancing' is a new normal, knowing the jobs that can be performed from home is useful for policymakers and researchers. A most common way to estimate jobs that can be performed from home is by carrying out surveys, but this method is time and cost-intensive. Moreover, using US-based O*NET surveys to estimate work from home (WFH) index in a home country, other than the US, can potentially lead to large measurement errors. In this paper, we propose a rating based methodology to estimate the WFH index. The ratings are provided for occupation-wise work activities according to their likelihood of getting performed from home by experts. This method is time and cost-effective and can easily be replicated in any country accounting for indigenous reality. Using tools like Inter-Rater Reliability (IRR), which is extensively used in psychology literature, we attempt to create a robust index of WFH for India.*

Keywords: Covid-19 pandemic, Work from home; Inter-rater reliability; Agreement Coefficients; Ratings

JEL Code: C1; J2; J6; J8

An alternate to *survey methods* to measure work from home*

Mohit Sharma¹, Sargam Gupta² and Xavier Estupinan³

September 2020

Abstract

During unprecedented times like COVID-19, when ‘social distancing’ is a new normal, knowing the jobs that can be performed from home is useful for policymakers and researchers. A most common way to estimate jobs that can be performed from home is by carrying out *surveys*, but this method is time and cost-intensive. Moreover, using US-based O*NET surveys to estimate work from home (WFH) index in a home country, other than the US, can potentially lead to large measurement errors. In this paper, we propose a *rating* based methodology to estimate the WFH index. The ratings are provided for occupation-wise work activities according to their likelihood of getting performed from home by experts. This method is time and cost-effective and can easily be replicated in any country accounting for indigenous reality. Using tools like Inter-Rater Reliability (IRR), which is extensively used in psychology literature, we attempt to create a robust index of WFH for India.

JEL Codes: C1; J2; J6; J8

Keywords: Covid-19 pandemic, Work from home; Inter-rater reliability; Agreement Coefficients; Ratings

* We thank Alankar Gupta for his excellent research assistance.

¹ Corresponding Author: Collaborative Research and Dissemination, Delhi, India. Email: mohit.mse@gmail.com

² Indira Gandhi Institute of Development and Research, Mumbai, India. E-mail: sargam@igidr.ac.in

³ International Labour Organization, New Delhi, India. E-mail: estupinan@ilo.org

1 Introduction

In unprecedented times like the COVID-19 pandemic, social distancing has become the need of the hour. To ensure social distancing, measures like working from home (WFH) have emerged as a cope up mechanism in certain jobs to ensure that work is not affected. In such jobs the worker can perform some/ all work activities related to their occupation from home, ensuring that physical interaction is minimized. To make informed policy decisions it would be useful for a policymaker to know about the jobs that can be fully/ partly performed from home. In the literature, US worker based O*NET surveys have been extensively used to estimate jobs that can be performed from home (see Baker, 2020; Dingel and Neiman, 2020; Koren and Peto, 2020; Mongey and Weinberg, 2020). The study by Dingel and Neiman (2020) proposed a methodology to estimate jobs that can be performed from home for countries other than US based on O*NET surveys. However, estimates based on O*NET surveys cannot be generalized as the *task content* of the occupation may vary significantly in other countries (Dicarlo et al., 2016; Lo Bello et al., 2019; Saltiel, 2019; Salitel 2020). Using STEP survey Salitel (2020) estimates workers who can WFH in 10 developing countries. In case of unavailability of a reliable WFH indicator or survey, researchers have relied on O*NET surveys to estimate jobs that can be performed from home in their respective countries.⁴ We show in subsequent sections that WFH estimates obtained using O*NET surveys have an upward bias in case of a developing country like India and can lead to misleading conclusions.

Given this background, we propose a *rating* based robust approach to measure WFH which is a cost and time-effective alternative to countries where indigenous O*NET like surveys do not exist. To check the reliability and robustness of ratings we borrow tools like Inter-Rater Reliability (IRR) from psychology literature. We believe, WFH thus derived provides a reliable estimate in the absence of surveys and that this methodology can easily be extrapolated to other countries given the availability of basic occupational dictionary and data on employment share. Further, the rating scale can be employed as an instrument to capture the country-specific intra/ inter occupation heterogeneity.

⁴ Chatterjee et al. (2020) have used O*NET surveys for India. The authors estimated that 16 percent of the jobs can be performed from home using NSS 2011-12.

2 Methodology

We take India as a case study to provide details on the methodology but the same can be replicated for any country. There are three steps to our methodology to generate a WFH measure; first, providing ratings to occupation-wise work activities as a likelihood to be performed from home; second, is to check the robustness of the ratings and measure the extent of agreement among the raters using inter-rater reliability (IRR), and third, is to generate WFH indicator, which measures the extent to which jobs can be performed from home. We have applied this methodology on the recently released Periodic Labor Force Survey (PLFS) 2018-19 in India.⁵

2.1 Step 1: Providing Ratings

The manual ratings have already been used in the literature to generate WFH. Rio-Chanona et al. (2020) use them to estimate first-order supply shocks in the US. We also provide estimates for first-order supply shocks for India using similar rating based WFH index in Estupinan et al. (2020). Dingel and Neiman (2020) used a manual assignment to test the reliability of their O*NET based WFH. Interestingly, WFH scores generated by them using manual assignment closely correlated with their WFH scores using O*NET surveys. This further corroborated our belief that if a significant agreement exists among experts than WFH measure based on those *subjective* ratings can provide a robust estimate of the jobs that can be performed from home. Use of IRR can help to formalize this method and make it more reliable.

We use indigenous occupation classification framework, National Classification of Occupation (NCO 2004 and 2015, India), to rate occupation-wise 511 work activities at 4 digit level 436 occupations (Figure 1).⁶ The crosswalk is used to map occupation-wise work activities available in NCO 2015 to occupations at a 4-digit level in NCO 2004.⁷ For further analysis, we consider a total of 434 occupation-wise work activities for 301 occupations as they are common in NCO

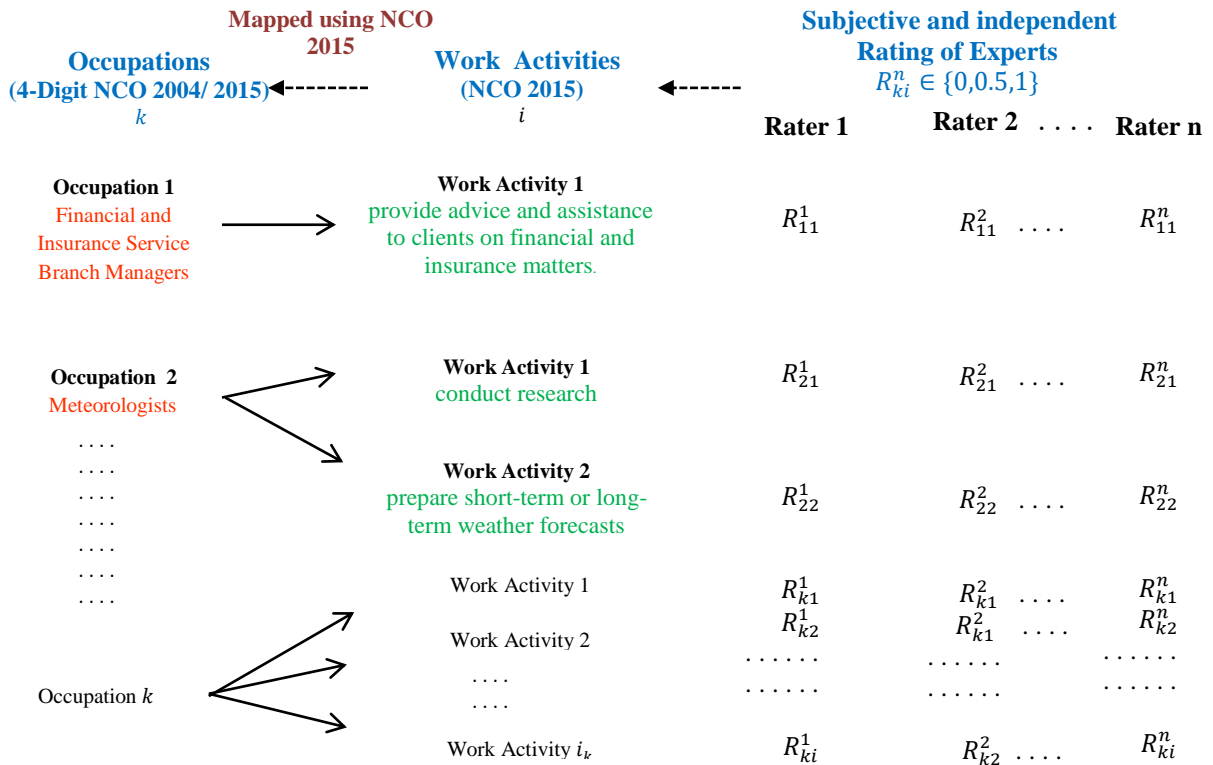
⁵ PLFS provides information on key indices such as employment and unemployment. Further it contains occupation and industrial group wise employment share. 1, 01,579 households were surveyed and 4, 20,757 individuals were surveyed in 2018-19.

⁶ The benefit of rating work activity is that it provides more clarity about the nature of occupation and helps account for more variability within an occupational category. However, in case of non-availability of data on work-activities experts can also rate occupation category. The limitation of rating work activity is the need to assume that all work activities, pertaining to same occupation, are equally important.

⁷ The need for mapping arises as data on work activities is available only in NCO 2015 and PLFS 2018-19 have information at 3-digit level NCO 2004.

2015 and NCO 2004. We used a Likert scale of three $\{0, 0.5, 1\}$ to rate work activities. Rating of 0, 0.5 and 1 imply, work activity is not likely, somewhat likely and is very much likely to be performed from home, respectively. The ratings are given by 4 economists; three of whom are the authors of the paper. Note that two aspects of ratings are crucial here, first rating work activity *within* each occupation; second, a Likert scale of three. These two aspects allow us to capture the varying likelihood of WFH for similar work activity in different occupations and the intra-occupation heterogeneity within the country. For instance, ‘research’ as a work activity can easily be conducted by a mathematician or an economist from home, however, biologist or chemist might need to access laboratory to conduct full-blown research. Moreover, the lack of internet connectivity in rural India precludes primary or secondary school teachers to teach students online. The same primary teacher can work from home in many parts of urban India due to better internet connectivity. We believe the rating of work-activity *within* each occupation in conjunction with Likert scale of three helps us to grasp the essence of such scenarios.

Figure 1: The NCO framework used to rate work activities within occupations for India



2.2 Step 2: Inter-Rater Reliability (IRR)

Inter-Rater Reliability (IRR) is a widely used tool in the field of psychology, medical and social research to estimate the reliability of the rating system (Gwet, 2008b). For example, psychologists use Agreement Coefficients (ACs), based on their independent and subjective ratings, to generalize diagnoses and treatment related to personality orders. If agreement among raters is *substantial* then it can be treated as evidence in favour or rejection of a particular diagnosis or treatment. We aim to extrapolate this methodology in this paper. We believe that if a *substantial* agreement exists among experts about likelihood of a particular work activity to be performed from home, in India, then there is a high likelihood that it can be performed from home and vice versa.

IRR framework involves *subjects* and *raters*. In our research framework, *subjects* would be occupation-wise work activities. We consider independent ratings of 4 raters on 434 subjects. There are broadly three steps to evaluate ratings using IRR. First, various ACs are calculated based on the ratings, which are point estimates.⁸ Second, we use an appropriate benchmark scale to interpret ACs. Landis and Koch (1977) were the first one to suggest a deterministic benchmark scale, however, Gwet (2014) proposed a probabilistic benchmark scale which accounts for the uncertainty associated with estimation (for details see Gwet, 2014; Klein, 2018). In this paper, we use probabilistic benchmark scale to interpret our ACs. Finally, for statistical inference and to check the robustness of point estimates of ACs to the varying number of raters and the varying number of the subjects we provide estimates for conditional and unconditional standard errors. The estimates related to the first and second step are covered in Table 1 and the third step is covered in Table 2.

Table 1 shows results for basic percentage agreement, chance corrected ACs - Fleiss' Kappa and Gwet's AC₁. Among various ACs, Gwet's AC₁ and kappa statistics are the most popular ACs and have been extensively used in literature to provide estimates on IRR (Cohen 1960; Fleiss 1971; Cicchetti 1990; Gwet 2008a; Gwet 2014). Further, to account for the partial agreement between raters we show weighted ACs (column (2) & (3)) along with the unweighted ACs (column (1)) in Table 1. We believe in our case, weighted rather than unweighted ACs provides

⁸ The data on ratings, STATA codes for IRR analysis, WFH for India at 4-digit NCO for 2004 and 2015 is available at: [Data Repository: Alternate to survey method to estimate WFH](#)

a more realistic point estimate given our rating scale is designed to capture close calls. To illustrate, let us assume there are two raters, rater 1 and rater 2. Rater 1 is certain that a particular work activity can be performed from home and gives a rating of 1, but rater 2 is uncertain and therefore gives a rating of 0.5. Ideally, this should not be considered a perfect disagreement between rater 1 and rater 2 as is done in the case of unweighted ACs. The perfect disagreement would entail if rater 2 would have been certain that the work activity can't be performed from home and therefore provides a rating of 0. Hence, a downward bias of ACs can result if we don't account for partial agreement between raters. For this purpose, we make use of two of the most commonly used weights - quadratic and linear, to arrive at weighted ACs (Klein, 2018). As can be seen, all the observed ACs are significant at 1% significance level.

Table 1: Agreement Coefficients to test Inter-Rater Reliability

Agreement Coefficients (ACs)	Weights Type (Probabilistic benchmark interval)		
	Unweighted (1)	Quadratic (2)	Linear (3)
Percent Agreement	0.793*** (Yes)	0.938*** (Yes)	0.890*** (Yes)
Scott/Fleiss' Kappa	0.612*** (No)	0.773*** (Yes)	0.693*** (Yes)
Gwet's AC₁	0.717*** (Yes)	0.868*** (Yes)	0.801*** (No)

Number of subjects = 434, Ratings per subject = 4, Number of rating categories = 3

*** p<0.01, ** p<0.05, *p<0.1; Null hypothesis H₀; AC_{observed} = 0

Benchmark interval and scale [Landis and Koch (1977)]

<0.0000	Poor
0.0000-0.2000	Slight
0.2000-0.4000	Fair
0.4000-0.6000	Moderate
0.6000-0.8000	Substantial
0.8000-1.0000	Almost Perfect

Source: Authors estimations

According to the benchmark interval and scale (mentioned below Table 1), all the unweighted and weighted ACs show at least a *substantial* agreement. The weighted percentage agreement and Gwet AC₁ shows the agreement to be *almost perfect* among raters. However, the benchmark interval scale is deterministic and does not account for the probability distribution and the error margins associated with point estimates. Thus, we consider probabilistic benchmark intervals

which are calculated using standard errors discussed later. In Table 1, whenever the probabilistic benchmark interval coincides with the deterministic benchmark scale we mark it as ‘(Yes)’ otherwise ‘(No)’. Using probabilistic benchmark interval, we find that all the unweighted and weighted ACs show a *substantial* agreement except the unweighted Fleiss' Kappa which indicates *moderate* agreement. Linear weighted Gwet AC₁ is also marked as ‘No’ as probabilistic benchmark interval indicates *substantial* agreement while deterministic benchmark interval suggests *almost perfect* agreement. As weighted ACs give more appropriate point estimates in our case and all weighted ACs show substantial agreement, we observe *substantial* to an *almost perfect* level of inter-rater reliability.

Table 2 summarizes the statistical inference for ACs when the subjects are sampled from the subject universe and raters are drawn from a larger rater’s population. Gwet (2008a, 2008b and 2014) provides the *variance estimators* to derive standard errors for all the three feasible cases; first, when sampling variance occurs due to sampling of *subjects* conditional on rater population, second, when sampling variance occurs due to sampling of *rater population* conditional on a specific sample of subjects and third when we account for sampling variance due to the sampling of both the subject from the subject universe and raters from the rater population. The standard errors associated with the third point are known as unconditional standard errors. All these precision estimates help to generalize our findings to the subject universe and rater population. Following this, Table 2 provides conditional standard errors which are varying in the number of subjects and number of raters, respectively and unconditional standard errors with varying subjects and raters together. Note that the default standard errors used in Table 1 to check the significance level are the standard errors conditional on the raters. Also, standard errors conditioned on the subjects use a jackknife estimator as proposed by Gwet (2014) and thus provides a robustness check to the small rater sample considered here.

Now, to test whether the observed ACs are significantly different from 60% which is the minimum required value for *substantial* agreement, we set the null hypothesis to be $H_0: AC_{\text{observed}} = 60\%$. We find that we can reject the null for all cases except, (i) for unweighted kappa for all precision measures; (ii) for linear weighted kappa and unweighted Gwet AC₁ when sampling of rater population conditional on a specific sample. Given that the weighted point estimate is more realistic in our case, both Gwet AC₁ and kappa estimate provide overwhelming

evidence towards the substantial agreement among raters. Moreover, all the observed weighted ACs is statistically significant using unconditional standard errors and thus we observe there exists at least *substantial* agreement for a larger rater population and the subject universe.

Table 2: ACs and associated precision measures

Unweighted ACs			
Statistics	Per cent agreement	Kappa	Gwet AC ₁
Agreement coefficient	79.3%	61.2%	71.7%
S.E. ^a due to the sampling of subjects	1.4%	2.2%	2.1%
S.E. ^a due to the sampling of raters	3.1%	4.8%	4.6%
Unconditional standard error	3.4%	5.2%	5.0%
95% C.I. ^b conditionally upon the rater sample	(76.6%; 81.9%)	(57.0%; 65.4%)	(67.7%; 75.8%)
95% C.I. ^b conditionally upon the subject sample	(69.3%; 89.3%)	(46.0%; 76.4%)	(57.1%; 86.3%)
Unconditional 95% confidence interval	(72.6%; 86.0%)	(50.9%; 71.5%)	(61.8%; 81.6%)
Weighted ACs: Quadratic Weights			
Statistics	Per cent agreement	Kappa	Gwet AC ₁
Agreement coefficient	93.8%	77.3%	86.8%
S.E. ^a due to the sampling of subjects	0.5%	1.9%	1.3%
S.E. ^a due to the sampling of raters	1.6%	4.0%	3.8%
Unconditional standard error	1.6%	4.4%	4.0%
95% C.I. ^b conditionally upon the rater sample	(92.9%; 94.8%)	(73.6%; 81.0%)	(84.2%; 89.4%)
95% C.I. ^b conditionally upon the subject sample	(88.9%; 98.8%)	(64.6%; 90.0%)	(74.7%; 98.9%)
Unconditional 95% confidence interval	(90.7%; 97.0%)	(68.6%; 85.9%)	(78.9%; 94.7%)
Weighted ACs: Linear Weights			
Categories	Per cent agreement	Kappa	Gwet AC ₁
Agreement coefficient	89.0%	69.3%	80.1%
S.E. ^a due to the sampling of subjects	0.8%	2.0%	1.7%
S.E. ^a due to the sampling of raters	2.1%	4.2%	4.2%
Unconditional standard error	2.2%	4.6%	4.5%
95% C.I. ^b conditionally upon the rater sample	(87.5%; 90.5%)	(65.4%; 73.2%)	(76.8%; 83.4%)
95% C.I. ^b conditionally upon the subject sample	(82.4%; 95.6%)	(56.1%; 82.5%)	(66.7%; 93.5%)
Unconditional 95% confidence interval	(84.7%; 93.3%)	(60.3%; 78.3%)	(71.2%; 89.0%)

Source: Authors estimations; ^aStandard error; ^bConfidence interval

While the acceptable benchmark interval for agreement depends on the underlying research question and discretion of the researchers, we believe for the present study establishing *substantial* agreement is sufficient. Moreover, as pointed by McHugh (2012), when the raters are trained and little guessing is likely to exist, per cent agreement is also a sensible choice to determine inter-rater reliability. For the present study, the percentage agreement estimates indicate *almost perfect* agreement among the raters for all precision measures.

2.3 Step 3: WFH Indicator and results

After establishing substantial agreement among raters we provide a WFH score to each occupation-wise work activities to ascertain the likelihood of that activity to be performed from home. First, we use the median rater's value to derive the scores at the work activity level. Then we average the work activity scores at the level of occupation to get WFH scores at NCO 4-digit level and NCO 3-digit level.⁹ Crosswalk is run between NCO 3-digit level and major occupation level according to O*NET derived classification (2 digit level) which Dingel-Neiman (DN) have used in their paper to draw comparisons between the two methodologies.

Table 3 shows the share of jobs that can be done from home according to DN and WFH measure. For most of the occupational groups, DN overestimates the WFH measure. To illustrate, about 97 per cent and 58 per cent of jobs in Legal occupation can be performed from home in India using DN approach and our WFH approach, respectively. We believe our estimate is more realistic given that a significant difference in production technologies and telecom infrastructure exists between developed and developing countries. For instance, in 2016, 82 per cent of households in the US had access to the internet (US Census, 2016), in comparison to 23.8 per cent of households in India (NSSO, 2019). Using employment-based weights we find that 20 per cent of the jobs can be done from home in India using DN approach in comparison to 13.5 per cent according to our WFH estimates (Table 4). Overall these results show that DN estimates may have an upward bias and it would be better to rely on methodologies which account for indigenous realities to estimate the jobs that can be performed from home.

⁹ Note that the information on employment is available at 3-digit NCO level in PLFS 2018-19.

Table 3: Share of jobs that can be done from home according to DN and WFH

Occupation	DN	WFH
Computer and Mathematical Occupations	1.00	0.98
Education, Training, and Library Occupations	0.98	0.50
Legal Occupations	0.97	0.58
Business and Financial Operations Occupations	0.88	0.66
Management Occupations	0.87	0.56
Arts, Design, Entertainment, Sports, and Media Occupations	0.76	0.42
Office and Administrative Support Occupations	0.65	0.46
Architecture and Engineering Occupations	0.61	0.28
Life, Physical, and Social Science Occupations	0.54	0.46
Community and Social Service Occupations	0.37	0.50
Sales and Related Occupations	0.28	0.18
Personal Care and Service Occupations	0.26	0.10
Protective Service Occupations	0.06	0.00
Healthcare Practitioners and Technical Occupations	0.05	0.20
Transportation and Material Moving Occupations	0.03	0.00
Healthcare Support Occupations	0.02	0.00
Farming, Fishing, and Forestry Occupations	0.01	0.00
Installation, Maintenance, and Repair Occupations	0.01	0.01
Production Occupations	0.01	0.04
Building and Grounds Cleaning and Maintenance Occupations	0.00	0.00
Food Preparation and Serving Related Occupations	0.00	0.00
Construction and Extraction Occupations	0.00	0.00
Misc	0.00	0.00

Source: Dingel and Neiman (2020) and authors WFH estimations

Table 4: Share of Work-from-Home at NCO 1-digit level according to DN and WFH

Occupation at NCO 1-Digit Level	All India (per cent)	
	DN	WFH
Legislators	87.00	72.56
Professionals	77.72	64.81
Technicians	76.33	39.51
Clerks	62.65	38.96
Service workers	26.10	12.58
Skilled agricultural workers	1.00	0.00
Craft workers	0.50	2.24
Machine operator	2.32	0.00
Elementary occupations	4.12	0.00
Occupations not classified	0.00	0.00
Total	20.03	13.49

Source: Authors estimations using PLFS 2018-19

4 Conclusion

We provide an alternate approach to estimate the work from home (WFH) index for countries where relevant surveys are absent. Using tools like inter-rater reliability we try to formalize a methodology that can be used to get WFH index in a time and cost-effective way bypassing the lengthy route of surveys. We believe our methodology can readily be applied to any country, to get estimates for WFH, given suitable occupation codes and employment data is available. Further, this can aid home countries to minimize measurement errors usually associated with the estimation of WFH index using other country's available occupational surveys.

References

- Baker, M. G. (2020). Characterizing occupations that cannot work from home: a means to identify susceptible worker groups during the COVID-19 pandemic. medRxiv
- Chatterjee, P., Dey, S., & Jain, S. (2020). Lives and Livelihood: An Exit Strategy from Lockdown. *Economic and Political Weekly*, vol IV no. 22.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37–46.
- Dicarlo, E., S. L. Bello, S. Monroy-Taborda, A. M. Oviedo, M. L. Sanchez-Puerta & I. Santos. (2016). The Skill Content of Occupations Across Low and Middle Income Countries: Evidence from Harmonized Data. IZA DP No. 10224.
- Dingel, J., & Neiman, B. (2020). Who Can Work at Home? *Journal of Public Economics* 189 (2020) 104235.
- Feinstein, A. R. & D. V. Cicchetti. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology* 43: 543–549.
- Estupinan, X., Sharma, M., Gupta, S., & Birla, B. (2020). Impact of COVID-19 Pandemic on Labor Supply and Gross Value Added in India. Available at SSRN 3628761.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76: 378–382.
- Fleiss, J. L., J. Cohen & B. S. Everitt. (1969). Large sample standard errors for kappa and weighted kappa. *Psychological Bulletin* 72: 323–327.
- Gwet, K. L. (2008a). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology* 61: 29–48.

- Gwet, K. L. (2008b). Variance estimation of nominal-scale inter-rater reliability with random selection of raters. *Psychometrika* 73: 407–430.
- Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters*. 4th ed. Gaithersburg, MD: Advanced Analytics.
- Klein, D. (2018). Implementing a general framework for assessing interrater agreement in Stata. *The Stata Journal*, 18(4), 871-901.
- Koren, Miklós, Petö, Rita. (2020). Business disruptions from social distancing. *COVID Economics: Vetted and Real-Time Papers* 1 (2), 13–31.
- Lo Bello, S., Sanchez Puerta, M. L., & Winkler, H. (2019). *From Ghana to America: The Skill Content of Jobs and Economic Development*. World Bank.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3), 276-282.
- Mongey, Simon, Pilossoph, Laura, Weinberg, Alex. (2020). Which Workers Bear the Burden of Social Distancing Policies? Becker Friedman Institute white paper
- NSSO (2019). Key Indicators of Household Social Consumption on Education in India. NSS 75th Round (July 2017- June 2018).
- Rio-Chanona, R. M., Mealy, P., Pichler, A., Lafond, F., & Farmer, J. (2020). Supply and demand shocks in the COVID-19 pandemic: An industry and occupation perspective. INET Oxford Working Paper No. 2020-05.
- Saltiel, F. (2019). Comparative Evidence on the Returns of Tasks in Developing Countries. mimeo.
- Saltiel, F. (2020). Who Can Work From Home in Developing Countries? 6 April Mimeo.
- US Census (2016). *Computer and Internet Use in the United States: 2016 American Community Survey Reports*. Issued in August 2018. ACS-39
- <https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf>