Working from home: estimations without surveys *

Henning Holgersen *Statistics Norway*
Zhiyang Jia *Statistics Norway*
Simen Svenkerud *Statistics Norway*

The COVID-19 crisis has forced great societal changes, including forcing many to work remotely (work from home) in an effort to increase social distancing. The ability to work from home has long been considered a perk, but we have few estimates of how many are actually able to work from home. Social media has been quick to argue that the people who are able to work at home are already privileged, highly educated and highly paid, while those who still have to go to work are more often in low-paid but critical roles such as drivers and grocers. This paper attempts to estimate the share of the Norwegian workforce able to work remotely by combining register-based labor statistics, annotated job ads and the official documentation of the ISCO-08 standard which is used both in labor statistics and in the job ads. We find that approximately 36 % of Norwegian jobs can be performed at home.

Keywords: Covid-19, Working from home, Job advertisements, Unconventional data, Norway

Introduction

Covid-19 pandemic hit the world hard and unprepared. In a study of the Spanish flu, (**?**) shows that non-pharmaceutical interventions known as "social distancing" during a pandemic can significantly reduce the disease transmission and lower both the peak and cumulative excess mortality. Learning from the historical lessons, many countries, including Norway, implemented measures to limit physical contacts between people. Encouraging working remotely (working from home) is one important part of these measures. However, concerns are raised quickly among experts and ordinary people on both its effectiveness and impact on social fairness. They argue that such policy will help only those who are high educated, highly paid and have comfortable jobs. The people who are much more susceptible benefit little from such measures. These concerns are well founded, but they ignored the externality of this measure. In fact, the measure will reduce the risk of infection indirectly for those who cannot work from home as well, since it will reduce both the potential infection risk and frequency of exposures to such risk. Here, the prevalence of remote feasible jobs matters. If there is only a very small fraction of working force can work from home, the effect may be ignorable to people others than those who work from home. Large prevalence of such jobs will imply higher effectiveness and lower negative consequence on social equality. However, there is very limited knowledge on prevalence of remote feasible jobs. In this paper, we try to answer the question: who and how many can work from home in Norway. Our analysis is based on three different types of data: a) the information of tasks described in the ISCO-08 standard (**?**). b) Job advertisements published by the Norwegian welfare administration (NAV) between January 2012 and march 2019, in which there are mentions of remote possibilities. c) official statistics on occupational employment published by Statistics Norway. The feasibility of an occupation is evaluated using ISCO descriptions of tasks to be performed. Detailed explanation of such classification is explained in the following in section sec:methodology [!TODO]. The classifications are then linked to the job advertisement data for consistency check and potential

*TBD

bias connection. Combing the feasibility classifications with occupational employment statistics, we then obtain information on the prevalence of remote feasible jobs and some characteristics of workers with such jobs. With help of the employment register, this can be done not only on national level but also on smaller geographical areas: for example, on municipalities.

We find that approximately 36 % of Norwegian jobs can be performed at home. As we expected, remote friendly jobs are often paid better than non-remote friendly jobs. There are large geographical differences in the prevalence of remote friendly jobs. There are larger share of remote friendly jobs in urban areas than rural areas. To some extent, this is a good news given that urban areas often facing a large challenges in containing the spread of Covid-19 given their high population density.

The method we applied here is similar to a very recent study by (**?**). They explores the same question in the United States using the Occupational Information Network (O*net) surveys covering "work context" and "generalized work activities". Unlike their study, we reply on much less conventional types of data: the description of tasks and contexts from job advertisements, which in turn requires new techniques for information extraction: manual labelling and natural language processing (NLP) techniques. Our study shows that combining conventional and unconventional data sources can lead to novel information that are useful for both official statistics, research and social policy making.

There were two surveys in norway in which some remote work feasibility is asked: The Norwegian labor force survey and a recent survey by the Norwegian Institute of Transport Economics(TØI). The Norwegian labor force survey has asked about remote-possibilities, covered in a report by (**?**). The question was whether the respondent had the opportunity to work from home at times, which is not to say that the job could be performed remotely in its entirety, and neither to say that those who weren't given the opportunity couldn't have worked from home if they had the option. The survey by TØI is designed specially for the Covid-19 situation with the main focus on the effective of remote work (**?**). While the survey does provide an overall estimate of the prevalence of remote jobs, information asked is rather limited. Neither occupational nor geographical aspects of the jobs are collected. Nevertheless the results we get were broadly similar to these surveys, while our analysis provides much more detailed information.

Method

The International Standard Classification

of Occupations organizes jobs into a set of groups according to the tasks and duties undertaken in the job. Using the detailed task descriptions listed in the ISCO-08 documentation, we try to provide a assertion of whether an occupation is likely able to be performed from home. To do this, we created a public labeling job through Amazon Mechanical Turk (**?**). Each occupation was presented together with a brief description. The exact question formulation was "Can this type of job likely be performed from a home office?", and an example of a job description could be:

> Social work and counselling professionals provide advice and guidance to individuals, families, groups, communities and organizations in response to social and personal difficulties. They assist clients to develop skills and access resources and support services needed to respond to issues arising from unemployment, poverty, disability, addiction, criminal and delinquent behaviour, marital and other problems.

The respondent was asked to evaluate wether it was likely that the job could be performed primarily from a private home. The alternatives were "Yes", "No" and "Unknown", which were

provided with the following description:

1. *Yes: This job can be performed primarily from an office in a private home*
2. *No: Substatantial parts of this job must be performed outside the employees home*
3. *Unknown: There is not enough information to decide*

In order to reduce the serendipity in the labels, we acquired five labels from different respondents for each occupation, and we provided an `uncertain` option in addition to the `yes/no` options in order to reduce arbitrary responses to uninformative occupation descriptions. The final labels include an uncertainty measure which shows that some of the occupations were evaluated differently by different annotators, but no occupation was given a final label of "Unknown" which means we can treat the remote-friendly annotation as a binary variable.

Since the job was on Mechanical Turk, there respondents were not subject matter experts, and likely reside in different countries. This adds to the importance of obtaining more than one label per occupation, but the number of labels does not correct for possible cultural differences - it is possible that some jobs that cannot be performed remotely in other countries can be performed remotely in Norway. We should consider the annotations as `international`, which is also true for the ISCO-08 standard itself.

Consistency check using the job announcements data

In order to evaluate the annotations from Mechanical Turk, we use job advertisements from the Norwegian welfare administration (NAV). The job advertisements have been published as open data by NAV, and contain the text, title, employer information, and annotations made by subject matter experts at NAV including the occupational code (ISCO) of the job. The dataset covers January 2002 through march 2019 (a total of 2.6 million ads), but due to changes in the ISCO structure (a switch from ISCO-98 to ISCO-08 in 2012), we only use ads from 2012 and onward. Furthermore, the volume of NAV ads increased sharply in 2018 due to new sources of data. From 2018 the NAV data is close to complete in covering formally advertised jobs in Norway. Some shops still only advertise in their shop window, but the vast majority of jobs posted online are now also posted on NAV.

Because the possibility to work from home is a perk for many, some employers mention it in their job ads in order to attract candidates. We search the texts for mentions of `hjemmekontor` and `heimekontor`, two distinctive words unlikely to mean anything other than the possibility of working from home. Since far from every employer advertise this possibility, it is difficult to say anything about the total number of remote-friendly jobs from these ads. It may however say something important about the relative frequency of remote-friendly jobs across broader occupational groups, which we can use to validate the results from Mechanical Turk.

We have also tried to combine the two data sources in a more rigorous way to correct possible bias in the annotation labels. However, this can only be done under rather strong assumptions on employee's behavior. Detailed discussion can be found in the appendix.

Evaluate Remote feasibility

With the estimated remote feasibility for each occupation, we use the Norwegian labor market data from Statistics Norway (SSB) to evaluate the remote feasibility in Norway. The data from

SSB covers employment, earnings and demographics. Employment data is register-based statistics which comes from two tables, covering the number of employed in Norway, divided by municipality and occupation respectively. Earnings data is also register based, and shows average monthly earnings by occupation. The demographics data shows population and density per municipality.

A note on nomenclature: For brevity, we sometimes refer to "remote" occupations rather than "occupations that can be performed remotely". We use the terms interchangeably, always referring to occupations that can be performed from home. This does not mean that such employees in actuality work from home either permanently or occasionally.

Remote feasibility using the annotation data

From the Mechanical Turk annotations, around 28 per cent of the occupations can likely be performed from home. Combining the annotations with labor statistics published per occupation, we find that 875 344 wage employers, 36 per cent of the workforce, are likely able to work from home.

From the Mechanical Turk annotations we see that 28 per cent of the occupations can likely be performed from home. Taken together with the labor statistics per occupation, we find that 875 344 jobs, 36 per cent of the workforce, are likely able to work from home.

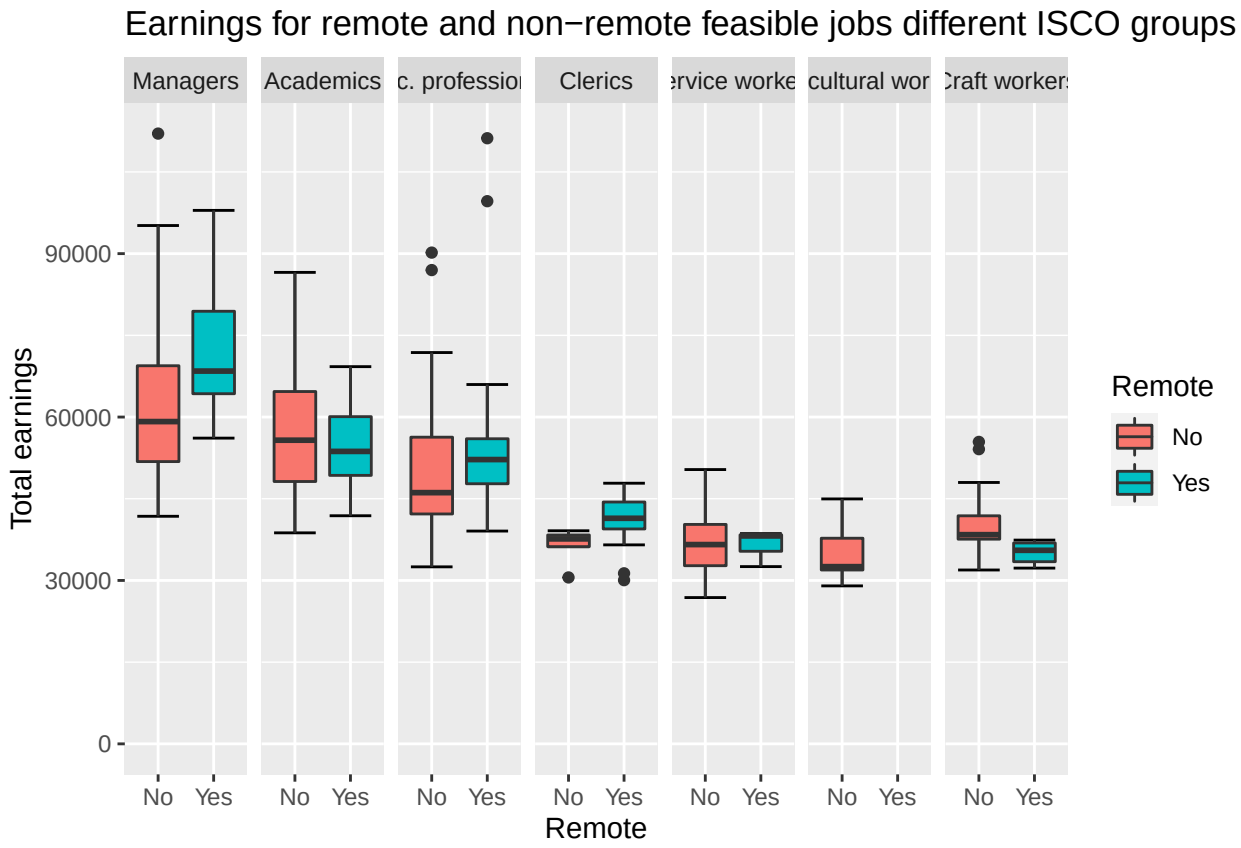
From the Mechanical Turk annotations, around 28 per cent of the occupations can likely be performed from home. Taken together with the labor statistics per occupation, we find that 875 344 wage employees, 36 per cent of the workforce, are likely able to work from home.

Splitting the annotated data into occupational groups, we estimate what percentage of occupations are remote-friendly across occupational groups. The results are presented in Table tab:Percent_occupation} [!TODO]. For each broad occupational group, the share of jobs that can be performed remotely varies from 0 to 67 per cent. Academics and managers are both groups where more than half of the employees can be done remotely, but an even higher share of clerical support workers are likely able to do their jobs from home.

	Number of jobs		
Occupational Group	Total	Remote-friendly	Percent remote friendly
Managers	183412	94959	51,8%
Academics	576136	297256	51,6%
Technicians and associate professionals	373065	158504	42,5%
clerical support workers	169230	112842	66,7%
Service and sales workers	570761	165493	29,0%
Skilled agricultural, forestry and fishery workers	21631	0	0,0%
Craft and related trades workers	219658	46290	21,1%
Plant and machine operators and assemblers	163197	0	0,0%
Elementary Occupations	134400	0	0,0%

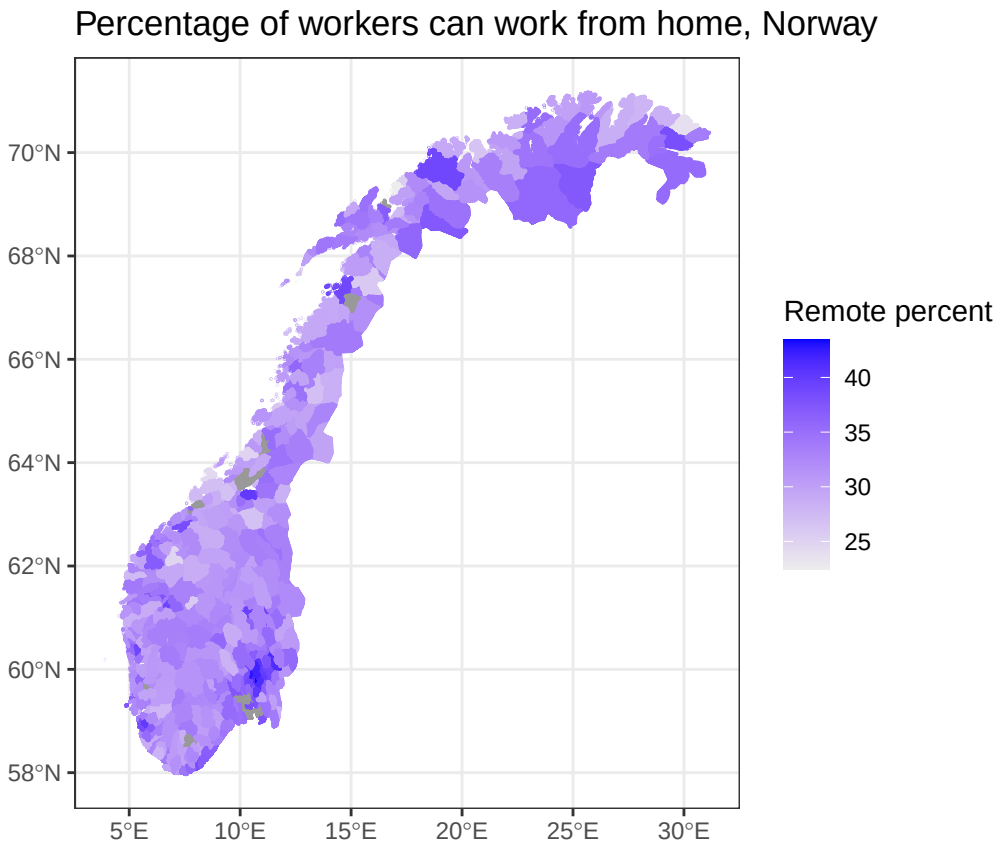
In general, occupations that can be performed remotely also pay better, as shown in table tab:Earning [TODO]. The same pattern is also found when we split the data by occupational group (Figure figure:earning [TODO]). However, we see two exceptions: for academic professions where a lot of the remote-friendly occupations occur, the average wage for non-remote workers is slightly higher. And the distribution of wages have a much larger spread as well. The wage differential among remote and non-remote Craft and related trades workers is much larger.

Remote	Average earnings	Median earnings
No	43274	40000
Yes	48929	47584



Variation of prevalence of remote feasibility across different regions

The geographic location of jobs have been a point of interest for years, amid both pressure for workers to centralize and specialize, and fears of increased inequality between cities and rural areas. Figure figure:geo[TODO] shows the percentage of workers can work from home in norway. As we expected, cities have a higher share of remote-friendly jobs, which may be fortunate given the need for social distancing. The pattern looks clear, especially in the area surrounding Oslo but also the other cities like Bergen, Trondheim and Stavanger seems to stand out on the map.

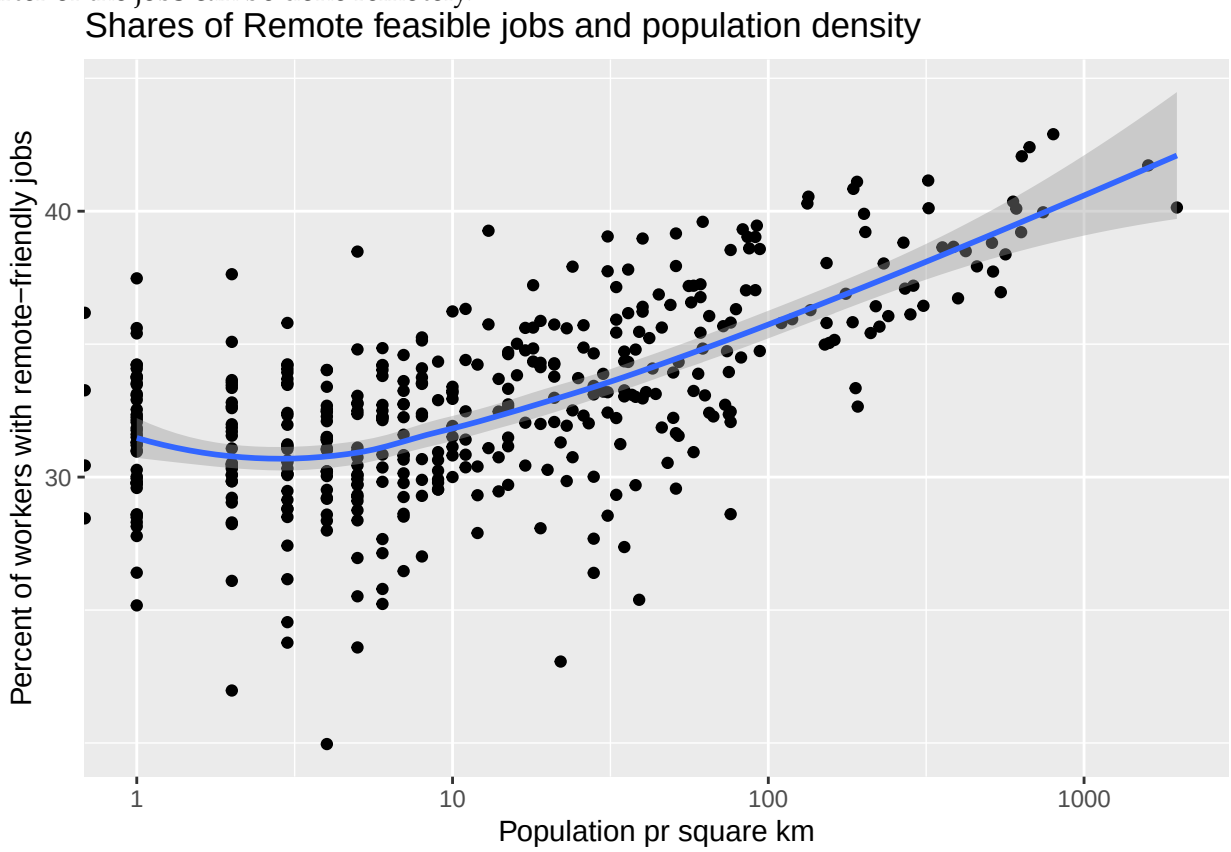


	Number of jobs		
Municipality	Total	Remote-friendly	Percent remote friendly
Oppegård	12964	5636	43,5%
Bærum	60596	25977	42,9%
Asker	28960	12294	42,5%
Oslo	344149	144657	42,0%
Nesodden	9027	3733	41,4%

	Number of jobs		
Municipality	Total	Remote-friendly	Percent remote friendly
Torsken	373	84	22,5%
Båtsfjord	1023	244	23,8%
Frøya	2494	613	24,6%
Stranda	2318	578	25,0%
Roan	426	107	25,1%

Table tab:5 most [TODO] and table tab:5 least[TODO] list the top and bottom 5 municipalities in terms of remote friendly jobs in Norway. We see that there are large heterogenity across different regions. Interpolating from the annotations we estimate that 42 percent of the jobs in Oslo can be done from home. On the other end of the spectrum, in municipalities like Båtsfjord less than a

quarter of the jobs can be done remotely.



By introducing a measure of urbanness, we can analyze the relationship more formally. We use population per square km as a proxy for urbanness. From figure [TODO]figure:pop_density, we can see a clear correlation between `urbanness`, or population pr km$^2, and the availability of remote-friendly jobs. Denser locations imply greater risks of Covid-19 spread, but this increased risk may be mitigated by better opportunities for remote work.

Validating results against job-ads

The job ads can not be compared directly to the annotated ISCO-08 data, but one of the comparisons we are able to do is the relative frequency of remote possibilities across occupational groups. Some job ads mention remote work, and we expect that remote possibilities are mentioned more often for occupations where remote possibilities are an option. We can compare this frequency with the number of remote-friendly occupations within each broader group (table tab:relative[TODO]). We also compute the composition of remote-friendly jobs across the occupational groups, giving us a measure that is directly comparable.

	Relative remote frequency		
Occupational group	Annotations	Job ads	Difference
Managers	4,9%	4,3%	0,5%
Academics	65,1%	49,7%	15,4%
Assc. professionals	14,9%	32,4%	−17,5%
Clerics	5,3%	5,5%	−0,2%

Service workers	4,6%	7,0%	−2,4%
Agricultural workers	0,0%	0,0%	0,0%
Craft workers	5,2%	0,9%	4,3%
NA	0,0%	0,1%	−0,1%
NA	0,0%	0,0%	0,0%

The notable discrepancies for the groups `Academics` and `Technicians and associate professionals` have a curious symmetry. Both of these groups require higher education or similar skill level, but Technicians and associate professionals are more vocational. There are many possible explanations for such differences, but for now it will suffice to conclude that one or more of the assumptions made is violated to some extent. Still the correlation is decent considering the spuriousness of the data.

In the appendix, we include some more discussions on this issue. It contains also a method we have developed to utilize job announcement data to correct for possible classification bias in the annotation labels. The magnitudes of the estimated biases are relatively small. Detailed estimates can also be found in appendix.

Related work

There are several very recent analyses that study the remote feasibility of jobs: (**?**), (**?**) and (**?**) for the United States, (**?**) for Germany, and (**?**) for Italy. Unlike our study, they mostly rely on different types of survey data. However, the results can be compared given the similar labor market institutions and labor force skill levels.

In Norway, as mentioned in section sec:introduction[TODO], there are two surveys related to our study. The results from the labor force survey was broadly similar to what we obtained with some differences. In the labor force survey, 71 per cent of managers responded that they had opportunity to work from home at times, much higher than the results here. This is likely attributable to the distinction mentioned above: Being able to perform *some* part of your job remotely does not mean that the job can be performed *primarily* remotely. The second survey by TØI was carried out after the Covid-19 outbreak in Norway. It reports an higher percentage of jobs (nearly 50%) are done from home these days. However, it is also reported considerable efficiency loss connoted to the home office arrangement. It may be that workers are stretching the limits of what can be effectively accomplished from home, or perhaps the annotations are slightly conservative. Note that although annotations were collected after the Covid-19 outbreak, it is hard to know whether the annotators had the pandemic in mind when annotating.

Conclusion

The sudden question of remote work highlights the need to expand our knowledge of occupations and their contents. In this paper we study the remote feasibility for different occupations in Norway. This analysis sheds more lights on a fundamental problem in the labor market. And more importantly it also provides useful knowledge for decision makers to evaluate potential social policies to combat the Covid-19 pandemic. Norwegian government will re-open primary schools and kindergartens soon. However, opening a school where majority of the parents can work from home would have rather different implications on potential virus spread, if compared with opening a school where majority of parents don't have this option. In addition, occupations with different remote feasibility will not be hit as hard, so these estimates would be useful when one want to assess potential economic impacts of the pandemic.

On the other hand, for National Statistical Institutions in Europe, the most natural option would be to expand the existing ISCO ontology with this data. This analysis is an attempt to combine the conventional and unconventional sources for statisical and research purpose. The results we have found also suggest that alternative approaches to collecting such information is feasible. The ISCO-08 documentation includes more detailed descriptions of the jobs than what we made use of in this paper. The complete list includes lists of tasks commonly performed, which may further nuance what parts of a job can actually be performed from home. The ISCO ontology may also serve this purpose.Although there were some discrepancies between the job ads and the annotations, there is clearly a pattern, and the differences between such sources should be explored. From an economic perspective this may tell us something about employer preferences related to hiring, and from the perspective of a National Statistical Institution such sources may, if calibrated correctly, provide further information about the labor market and working conditions. The possibilities are not limited to the question of working from home. Job ads are a great resource for describing actual jobs and what they entail, although the picture the advertisements paint might be a little rosy.

Appendix

Notations

Jobs are indexed using i. Each job belong to one occupation (indexed by j) which can be aggregated again to the ISCO group (indexed by g). The set of jobs for occupation j is denoted as $B(j)$ and the set of occupations for ISCO group g is denoted as $D(g)$.

The job ads data

Job ads data contains all new jobs announced through NAV.

Let Y_i be the variable indicating whether the job announcement for job i contains the phrase "working from home", i.e:

$$Y_i = \begin{cases} 1 & \text{contains "hjemmearbeid"} \\ 0 & \text{otherwise} \end{cases}.$$

Whether to list the phrase `working from home` in the announcement for job i depends on the feasibility that the job can be performed remotely. However, it depends also other factors, such as whether the employee consider the feature "working from home" is important for them to draw high quality potential candidates. That is, for a job i of occupation j, we can write

$$P(Y_i\& = 1) = q_a(i)q_f(i) = p_a(j)p_f(j)\epsilon(i)$$

where $p_a(i)$ is measures the mean probability that employee consider the working from home feature is important for job i, $p_f(i)$ is the probability that job i is a "working from home" job, $E_{i\in B(j)}(q_a(i)) = p_a(j)$, $E_{i\in B(j)}(q_f(i)) = p_f(j)$ and $E_{i\in B(j)}(\varepsilon(i)) = 1$. Using observed data, we have from (eq:prob_job[TODO])

$$\bar{Y}_j = \frac{\sum_{i\in B(j)} Y_i}{n_g(j)} = p_a(j)p_f(j)\frac{\sum_{i\in B(j)} \varepsilon(i)}{n_g(j)} \longrightarrow p_a(j)p_f(j).$$

Assume that $p_a(j)$ is a fixed number within the same ISCO group, then we see immediately that

Percentage of occupations are remote-friendly across occupational group

term	estimate	std.error	statistic	p.value
Intercept	-1.02	0.13	-8	0.0000
Mentions pr 1000	0.08	0.02	4	0.0001

$$\frac{\bar{Y}_j}{\bar{Y}_k} \longrightarrow \frac{p_f(j)}{p_f(k)}.$$

That is, the risk ratio between two occupations within the same group can be consistently estimated. If we are willing to assume in addition that $p_a(j)$ is the same for all occupations, we have

$$\bar{Y}_g = \sum_{j \in D(g)} S_g(j)\bar{Y}_j \longrightarrow p_a p_f(g).$$

where $p_f(g) = \sum_{j \in D(g)} S_g(j) p_f(j)$ is the feasibility measure for ISCO group g and $S_g(j)$ is the share of jobs of occupation j in ISCO group g. Similar to (eq:risk_ratio_occupation[TODO]), we have

$$\frac{\bar{Y}_g}{\bar{Y}_l} \to \frac{p_f(g)}{p_f(l)}.$$

The occupation annotation data

This data set contains information of working from home feasibility information based on the description of the tasks for each occupation. For occupation j it contains a dummy variable A_j which indicates whether this occupation is working from home friendly, and a variable (τ_j) which indicates how certain A_j is.

Logit regression: consistency between two data sources

$$latex In order to evaluate it statistically we would benefit from more detailed data. One approach is to use a logistic regression explaining the annotation of an occupation (a binary label) with the ratio of mentions of"hjemmekontor"{} in the ad. Since mentions are quite rare in the ads, we computed "mentions pr 1000 ads"{} as a more intuitive measure. $

$$latex

$

Mentioning "hjemmekontor" in the add is significantly correlated with the ad being labeled annotation. It remains significant when including occupational group (dummy), and number of ads for the occupation as further explanatory variables. As for the goodness of fit, the model including occupational groups and number of ads achieve an accuracy of 75%, which is only modestly above the baseline at 68%. Given the spuriousness of the data sources, we consider a clearly significant correlation between mentions of "hjemmekontor" and the ISCO annotations a good result.