# Does Feasibility Explain the Unequal Development of Working From Home?

Thomas Breda, Paul Dutronc-Postel and Vladimir Pecheu*

Friday 25$^{\text{th}}$ October, 2024

**Abstract**

Using rich historical surveys on job tasks and advanced machine learning techniques, we study which jobs can be moved from the office to home over three decades in France. The share of jobs that can be done from home has increased steadily from 14% in 1991 to 45% in 2021. At the same time, actual Working From Home (WFH) remained limited to less than one fifth of its full potential before the Covid-19 crisis and is still below it in 2021. The growth of WFH is largely unrelated to the evolution of job tasks, implying that the main obstacles to WFH have not been technical constraints. Low-skilled employees in particular have jobs that have long been largely teleworkable but they were barely teleworking before the Covid-19 crisis and remained still below 50% of their full potential during it. This pattern is not explained by differences in workers' desire to telework. It implies that the well-known large inequality in access to WFH along the earnings distribution cannot be attributed only to feasibility constraints and is potentially inefficient.

JEL codes: J22, J81, M54.

Keywords: Working from home, Feasibility, Inequality, Historical series

*Breda: CNRS, Paris School of Economics; e-mail: thomas.breda@ens.fr. Dutronc-Postel: Paris School of Economics -Institut des politiques publiques; e-mail:paul.dutronc@ipp.eu. Pecheu: Paris School of Economics; e-mail:vladimir.pecheu@psemail.eu. This work was supported by a public funding by the French National Research Agency (ANR).

# 1 Introduction

The COVID-19 pandemic has triggered a massive shift toward working from home (WFH) as well as intense debates regarding the quantitative importance of this shift and whether it will be permanent or temporary. Recent research has tried to evaluate both the share of jobs that can be done from home in the modern economy (e.g. Dingel and Neiman, 2020) and the likelihood that the high levels of WFH observed during the pandemic will persist (e.g. Barrero et al., 2021).

We contribute to these debates by examining the long-run evolution of the working practices that make WFH feasible in France with an analysis that combines rich survey data and advanced machine learning techniques. Doing so, we can assess how the task composition of jobs has changed over the past 30 years and how this change has increased the share of jobs that can be done from home. This allows us to study gaps between actual WFH and its feasibility over time and across occupations, and to assess the extent to which WFH is and has been mostly limited by technical feasibility constraints. This historical approach sheds light not only on the barriers to the development of WFH just after the Covid-19 crisis, but also on the potential for its expansion at an even larger scale in the long run. Doing so, it informs the debate on the likelihood that WFH could be widespread in the future. Our comparisons across occupations then allow us to study overall inequality in access to WFH conditional on its feasibility.

Our first contribution is to study the evolution of working practices (or tasks) that make WFH feasible. Following Hatayama et al. (2020), we consider four types of tasks or factors that can facilitate or impair WFH: (1) information and communication technology (ICT) use, (2) face-to-face interactions, (3) physical work, and (4) the quality of Internet connection at home. We present the evolution of the prevalence of these various fundamental aspects of work since 1991. We then use machine learning (ML) techniques to predict which job tasks or working practices make WFH feasible and validate the relevance of the factors considered by the existing literature. This is done using a working conditions survey administered in early 2021 in France—a period with strong COVID-19 prevalence—and containing information on WFH feasibility of jobs (self-assessed by survey respondents) and the task content of jobs. We predict WFH feasibility by the task content of jobs and apply this predictor to former surveys that include similar questions, allowing us to examine the evolution of predicted teleworking feasibility since 1991. A strength of this approach based on ML is to be agnostic regarding the factors that should be considered to assess which jobs can be done from home.

The second and main contribution of the paper is to provide an analysis of the gaps between actual WFH and WFH feasibility over time and across occupations. This allows us to identify practical obstacles to WFH, and how they may have been lifted over time. Specifically, we study the evolution of WFH at the occupational-level using monthly information from the French labor force survey, and compare it to our measures of WFH feasibility aggregated at the occupational level and their evolution over time.

Our analysis yields two main results: first, before the COVID-19 crisis, WFH remained very limited and increased much more slowly than its feasibility. The share of jobs that can be done from home increased steadily from 14% in 1991 to 35% in 2019 and 45% in 2021. At the same time, the share of the workforce teleworking increased from virtually 0 in the early 1990s to 6% in 2019 and 23% in early 2021. Even during the Covid-19 crisis, the recourse to teleworking remained far below its full potential. These results clearly show that the main obstacles to teleworking are not feasibility constraints that were progressively alleviated while turning to the digital economy.

Second, there is a large gap in actual WFH across occupations that cannot be explained by differences in WFH feasibility: office employees that were already spending more than 80 % of their time on a connected computer in the 2000s are much less likely to work from home than white-collars with similar working practices. This is remarkably true both before and after the Covid-19 crisis. It implies that the large inequality in access to WFH along the earnings distribution cannot be attributed only to feasibility constraints: we estimate that in 2019, the share of workers who actually telework among those who could do so increases from around 15% at the bottom half of the earnings distribution to 50% at the top decile.

To explain these gaps for the recent period, we use information on workers' desire to work from home and management practices such as workers' degree of autonomy and hierarchical control. We find that in early 2021, among individuals who could telework but do not, low-skill workers are more likely to desire to telework than their higher-skill counterparts. This shows that the inequality in access to telework conditional on its feasibility that we document could be inefficient from an economic point of view if the effect of teleworking for low-skilled jobs on firms' productivity is not negative. Hence, there is a strong margin for increasing teleworking at the bottom of the earnings distribution and this could be welfare-improving for both workers and firms.

We also find that both in 2019 and 2021 Workers' autonomy and extent of hierarchical control play a role beyond pure feasibility constraints as they are positively associated with the probability to telework (conditional on feasibility). This suggests that managers' willingness to maintain control over their workers could be a main factor limiting teleworking for low-skilled workers today. We conclude that the evolution of teleworking in the future will depend more on the evolution of management practices than on changing the task content of jobs to make WFH feasible.

**Related literature** Our study relates to a rapidly growing literature on WFH. A first strand of this literature examines which jobs can be done from home. Dingel and Neiman (2020) have developed an axiomatic approach that consists in identifying the types of tasks that make jobs (not) amenable to WFH. Mongey and Weinberg (2020); Gottlieb et al. (2020); Saltiel (2020); Hatayama et al. (2020); Garrote Sanchez et al. (2021), among others, have built on the work of DN2020. Mongey and Weinberg (2020) propose a continuous measure of WFH feasibility. Gottlieb et al. (2020) apply DN2020 to 57 countries and focus on biases induced by categorizing farmers as not WFH in developing countries. Saltiel (2020) and Hatayama et al. (2020) use skills data in several countries to examine if the task content of different occupations vary across countries, making US-based measures à la DN2020 not applicable to other countries. Garrote Sanchez et al. (2021) develop a measure of WFH feasibility that also takes into account Internet access and apply it to 107 countries.

A last group of papers departs from the task-based approach and uses directly individual-level data on WFH (e.g. Hensvik et al., 2020), or on WFH feasibility directly self-assessed by workers (e.g. Adams-Prassl et al., 2022; Alipour et al., 2023). Alipour et al. (2023) and Adams-Prassl et al. (2022) highlight the strong heterogeneity of WFH feasibility within occupations or industries, while Alipour et al. (2023) study directly the link between tasks and WFH feasibility, showing in particular that a single indicator of PC usage does a good job in predicting WFH capacity. Finally, independently of the method they use, all papers focusing on WFH amenability conclude that it increases with workers' earnings and education levels.

Our contribution relative to these papers is twofold. First, we analyze the evolution over 30 years of the tasks and working practices that make WFH feasible in a developed country. This allows us to estimate how the share of jobs that can be done from home has evolved over time. This approach can be seen as similar in spirit to the work on job polarization that has examined the evolution over time of employment in occupations depending on their task content (e.g. routine versus non-routine tasks, see for example Autor et al. (2003) or Autor and Dorn (2013)). The difference is that we analyze tasks that are amenable to WFH, and we have access to direct information on the prevalence of these tasks at work at different points in time.

Second, we rely on two rich working conditions surveys administered just before and during the pandemic to propose alternative, more agnostic, approaches to assess the task content of jobs that are associated with both actual WFH (before and during the COVID crisis) and WFH feasibility. Doing so, we follow and extend the work of Alipour et al. (2023). While they use a linear probability model to link job tasks to WFH feasibility, we use Random forests, allowing us to take into account interactions between predictors and nested effects of the tasks that make WFH feasible. The ML approach is also key to get an accurate predictor of WFH or WFH feasibility that we can apply retrospectively without fearing that we are over-fitting the data.

We then relate to the papers that try to assess if WFH will stick durably after the COVID-19 crisis (Barrero et al., 2021; Bartik et al., 2020; Bick et al., 2021). These papers rely on theoretical arguments typically combined with survey data measuring, during the crisis, workers' desire to keep WFH and employers' plans to maintain it. Our retrospective analysis of the development of working practices that make WFH feasible complement these papers as it can shed light not only on the immediate persistence of WFH but also on the potential for its expansion at an even larger scale in the long run.

Finally, we indirectly relate to literature that attempts to evaluate how workers and employers value WFH and are willing to pay for it (Mas and Pallais, 2017; Oettinger, 2011; Nagler et al., 2022; Lewandowski, Piotr et al., 2023). We do so by analyzing the gaps between WFH feasibility and actual WFH, combined with information on workers' desire to WFH. We show in particular that gaps between actual WFH and WFH feasibility or workers' desire to work from home vary strongly across occupations.

**Organization** The rest of the paper is organized as follows. Section 2 briefly describes the institutional context and presents the data and methods. Section 3.1 analyses the tasks that predict WFH feasibility and actual WFH in the recent period. Section 3.2 shows the evolution of these tasks and more boradly of WFH feasibility in the labor market since 1991. Section 4 analyzes gaps between actual WFH and WFH feasibility before and during the COVID-19 crisis and their determinants. The evolution of the gaps over time and their differences across skill groups are highlighted. Section 5 provides a brief analysis of the variation in the take-up of teleworking given its feasibility along the earnings distribution. Section 6 summarises the main results and discusses the possibility of a widespread usage of WFH in the economy in the long-run.

# 2 Context, Data and Methods

In this section, we provide contextual information on the development of WFH in France and in other countries, on the difference between WFH and teleworking, and on the regulations governing these practices. We then describe our data and the methods used to analyze it.

## 2.1 Context

Systematic quantitative evidence on teleworking and WFH arrangements prior to the Covid-19 crisis is remarkably scarce. Two notable exceptions are Barrero et al. (2023) and de Vos et al. (2019) on the USA and the Netherlands, respectively.

The former documents a 7.2 % take up of "working from home" in terms of days worked in the USA in 2019, while the latter reports that 22 % of Dutch workers declared having a home working day in 2018. This large cross-country variation may come from cultural and institutional differences, but also from differences in the definition and measurement of teleworking and WFH arrangements. These definitions matter, both for assessing international or regional differences, and the temporal evolution of the take-up of these practices. So far, we have used the two terms interchangeably but a distinction is typically made between *working from home* and *telework*. The former term refers to the fact that a worker employed in a firm is not working within the firm's premises, but rather at her home. The latter also implies remoteness from the firm's premises, albeit not necessarily at a worker's home, but also the use of information and communication technologies. French law defines telework more restrictively, by considering only work that could have otherwise been executed within the firm's premises and that relies on formal arrangements with voluntary workers. ICT use also directly enters in the French legal definition of teleworking since the worker is required to be reachable by ICT in order to be considered to be teleworking.[1]

[1]Whether the definition of teleworking also implies some degree of regularity is important, in particular because of its legal implications. "Regular" teleworking is subject to more regulations—notably in terms of rights for

Figure 1 shows the evolution of teleworking and WFH arrangements in France since the 1990s (see section 2.2 for a description of the data sources). The share of workers who telework at least sometimes—telework being defined as the combination of remoteness from the employing firm's premises and usage of ICT—has been increasing in France since the early nineties, with a more rapid increase in the 2010s, prior to the pandemic. It reached about 6 % of the labour force before the Covid-19 crisis and then jumped by more than 15 percentage points, reaching 24 % in 2021 before decreasing slightly to 21 % in 2023. The share of workers who declare working at least partly from home over the last month when surveyed increased by a similar magnitude in absolute terms during the Covid-19 crisis, but relative to a much higher baseline, at around 17 % in 2019. It is 10 percentage points higher than the share of workers who telework throughout the studied period.[2] Finally, the large gap between the purple and red series illustrates that WFH outside regular hours, that is, extra time worked at home, represents a large share of the WFH prevalence before the Covid-19 crisis.

The *potential* for teleworking and WFH arrangements has received regular bursts of media and policy attention. The inception of the word "teleworking" dates back to the 1970s, and originates from the engineering literature. Ever since, periods of renewed interest have alternated with suggestive evidence that actual teleworking practices are less prevalent than anticipated. In France, the first instance of teleworking receiving official attention from policymakers dates back to the early 1990s (Breton, 1993).[3] It includes a study by the French statistical office (the Insee) that evaluates the take-up of telework at the time at 16 000 workers, or 0.1 % of the labour force.[4] Using survey data and projections on the evolution of sectoral and occupational structure of the workforce, it also projects the potential for an increase in telework in the future, with an

employees—than occasional or exceptional teleworking.

[2]These figures can be put into perspective with those of Barrero et al. (2023) and de Vos et al. (2019) who document increases of "working from home" to 28 % in the US in 2022 and to 53 % in the Netherlands in 2021.

[3]The report originated from the ministry in charge of land use planning. Historically, one of the main sources of academic and political interest for teleworking development, both in the US and France, has been connected to road congestion and transport infrastructure.

[4]In comparison Barrero et al. (2023) estimates 0.4 % of days worked from home in 1965.

estimated 200 000 to 300 000 workers practicing teleworking in 2005. In contrast to most other predictions, this one actually turned out remarkably close to the observed take-up of telework in 2006 at 2 % (Figure 1). More generally, how close “realized” teleworking follows its “potential” has been debated for decades in non-academic work. In 1992, 54 % of French workers declared being willing to stay at home to work at least part of their working time if given the chance to do so, illustrating that there has been a strong desire to work from home on the workers’ side for a very long time. In this paper, we explore the historical discrepancy between teleworking potential and teleworking practices in a data-driven way to see if workers’ desires have been fulfilled when technical barriers to teleworking started to be removed in their jobs.

## 2.2 Data

**Working Conditions Surveys 1991-2021** Working Conditions Surveys (*Enquêtes “Conditions de Travail”* in French) are administered by the statistical office (the *Dares*) of the French Ministry of Labor every 6 to 8 years to samples of 15,000 to 25,000 workers since 1991. Survey microdata are available for years 1991, 1998, 2005, 2013 and 2019. The surveys include sampling weights allowing us to produce statistics representative for the whole workforce. They include an extensive questionnaire which is aimed to capture working conditions in a broad sense: family status, occupation, former career path, work schedule, work intensity, family-life balance, working practices and tasks executed at work, subjective health, psycho-social risks (since 2016) and WFH (since 2019). We have excluded occupations 31 ”Independent Workers” 35 ”Artists and Entertainment Occupations”, 44 ”Clergy” and 42 ”Teachers” because they do not correspond to typical salaried jobs (see details in Appendix A).

In addition, the Ministry of Labor administered an adapted version of the survey (“Tracov”) to 19,953 workers between January 25th and March 7th 2021 to capture the evolution of working conditions during the COVID-crisis. While former regular surveys contained little information on telework, Tracov includes several questions allowing us to measure both telework, telework feasibility and workers' desire to Telework. Table 1 provides basic information on each survey (sample size, response rate, main topics included, measures of WFH and teleworking). A challenge for the purpose of this paper is that survey questions are not fully consistent across the whole period 1991-2021. We discuss how we address this in Section 2.3.

**Labor Force Surveys 2003-2023** The Labor Force Survey (LFS) is a household survey run on a quarterly basis by the national statistical institute (Insee) since 1950. We keep only employed individuals and use sampling weights in all analyses except regressions to get representative statistics. The LFS includes information on whether a worker has worked at least some time from their home in a consistent way since 2003. We use it to compute the evolution of *WFH* since 2003.

**Archive data on teleworking** Data on teleworking practices are scarce. We use data from Breton (1993), which surveyed the French private sector extensively on teleworking practices in 1993. We also use survey data from the 1997 and 2006 waves of the Digitization and Organisational Changes survey (*Enquête sur les Changements organisationnels et l'informatisation*) ran by the Insee. The survey samples workers from the private sector and collects information about how often they work outside of their firms' premises, and their use of ICT.

**Internet coverage** We use data from the *ARCEP*, a public entity which depends on the Ministry of Economic and Financial Affairs. The data contains the coverage rate of Internet connections with a given download speed at the communal level from 2013 to 2019, broken down in five speed categories: ADSL (Asymmetric Digital Subscriber Line, with a speed of 512 kb per sec when it was introduced in 1999) and 4 broadband Internet connection speed levels: 3, 8 , 30, and 100 Mb per second.[5]

[5]For earlier periods, we interpolate these series with an earlier snapshot of data in 2005 (introduction date of connection speeds faster than 30 Mb/sec), and with the knowledge that virtually no commercial household connection existed in 1998 other than via the phone network. For later periods, we extrapolate based on the 2013-2019

## 2.3 Methods

**Measures of WFH, telework, and its feasibility** As pointed out, the different surveys we use in our analysis encompass different concepts of working outside firm premises. Telework is reported in Working Conditions surveys of 2019 and 2021 (WC2019 and Tracov2021 thereafter), while WFH is reported in LFSs. We focus exclusively on the takeup of these practices at the extensive margin. In the WC2019 and Tracov2021 surveys, we construct a telework dummy variable that identifies individuals who telework at least "a few days per month". In the LFS, we consider as our main measure of WFH, an indicator that identifies individuals that have worked from home at least partly during the month preceding the survey. Our main measure is obtained after excluding workers who report working during evenings or weekends. We argue that this latter measure is closer to telework because it excludes situations where WFH is not a substitution to working at the office (which telework is).

Tracov2021 asks an additional question that identifies individuals who report having a job that "is not concerned by telework because its tasks are not compatible with it". We use it to construct a dummy variable that identifies telework feasibility: whether the individual is employed in a job that is teleworkable.

**Machine learning prediction** To analyze the evolution of telework and its feasibility as a function of job tasks, we use the Random Forest machine learning technique as it appears particularly well-suited to predict outcomes in our particular case. Formally, a Random Forest is a non-linear prediction technique based on a set of decision-tree classifiers optimized for the prediction of an outcome (see Appendix Figure A.2 for an example for teleworking in 2021). Each tree is constructed with a number of randomly drawn explanatory variables among a set of potential predictors. Its advantage, common to other machine learning techniques, is that it is an agnostic

trends. For 2013-2019 we do not have the data for ADSL access only, but for access to ADSL or other forms of cable connection. This change in the variable should not be problematic if we reconsider it as a measure of the prevalence of low speed Internet connectivity over this latter period.

method to rank the importance of predictors, and thus prevent issues related to cherry picking the explanatory variables of the model, or making other arbitrary choices. As opposed to linear models like LASSO or Ridge Regression, the Random Forest is well suited to take into account hierarchies and nested effects between predictors, which are likely to be important in our case because teleworking feasibility depends on interacted conditions (e.g. it is not feasible if workers exert physical tasks or if they do not but have contact with the public). We provide further details on Random Forests in Appendix A.

The most important constraint we face in our analysis regarding the choice of our predictors comes from the fact that we cannot use all information available in the recent surveys (WC2019 and Tracov2021), to predict telework and its feasibility backwards using historical survey waves, which contain fewer variables. Nonetheless, to be fully transparent and informative about the performance of our models, we provide two sets of predictors. The first-best models use as predictors *all* survey variables in WC2019 and Tracov2021 that measure the nature of tasks at work or working practices. In contrast, the second-best predictors use the subset of these variables that are available in WC2019 and Tracov2021, as well as in all historical waves of WC surveys. Only these second-best models can be used to apply our predictors backwards. To establish the list of eligible predictors, we identify all variables in each survey that are directly reflecting the nature of the job, and exclude those that relate to management or organizational practices. Indeed, these latter variables reflect individual and/or collective choices regarding the way the job is accomplished rather than immediate feasibility constraints directly related to the nature of the job.

The variables retained as predictors are detailed in Table A.1 which provides the exhaustive list of survey questions capturing task features and the yearly surveys in which they are included. Most variables mainly characterize contacts at work, physical activities, and the use of ICT. We are left with 67 predictors in CT2019 and 13 predictors in Tracov2021 (see Table 2). However, we have only 5 variables common to these two surveys and all historical working conditions surveys. These variables are the share of hours spent using connected computers, whether one's job requires contacts with the public, and whether one's job is sometimes, often, or always physically demanding (3 variables).[6] While this set of predictors is limited, it covers the three main aspects of work likely to affect workers' ability to telework.

Finally, in an additional exercise, we also add as predictors to our second-best model five measures capturing Internet coverage in the department or region of residence of survey respondents at the time of the survey (i.e. access to ADSL and four variables for access to broadband Internet with various speeds—from 3 Mb to 100 Mb per second).[7] The reasons for doing this separately from the rest are that (i) we can only match Internet coverage to Tracov or WC surveys at a a very aggregated level, and (ii) Internet coverage being already high in 2019 and 2021, these aggregated variables of coverage yield limited variations across regions and only have limited predictive power for teleworking feasibility.

**Long-run evolution of predicted teleworking and teleworking feasibility** We construct historical series of predicted teleworking feasibility using the ML models. Let us denote $f_{2021}^{WFH feasibility}(X)$, $f_{2021}^{WFH}(X)$ and $f_{2019}^{WFH}(X)$ the "second-best" predictors of teleworking feasibility and teleworking depending on individual job task features $X$. We apply these functions to measures of tasks obtained backward in time. Formally, to illustrate the evolution of predicted teleworking feasibility, we show for years $t = 1991, 1998, 2005, 2013, 2019$ the empirical counterpart of:

$$\mathbb{E}_{2021}[Feasibility_t] = \int f_{2021}^{WFH feasibility}(X_t)\, dX_t$$

[6]Categorical variables often take multiple modalities. In order to increase the performance of ML predictions, we split all categorical variables into dummies for each modality but the reference one. Therefore a single job feature (e.g. having a physically demanding job) can be accounted for by multiple predictors provided to the the ML algorithm.

[7]In Tracov2021, WC2019 and WC2013, we observe respondents' *département* of residence (100 in total in France). In WC2005, we only observe the region (22 in total) while in earlier surveys we do not have any information and adopt a conservative scenario assuming that Internet coverage is 0 for all types of connections and all respondents.

where $X_t$ are task features observed among individuals surveyed at date $t$. The two predictors of teleworking applied backward are obtained similarly. They do not directly provide an estimation of the share of jobs that could be teleworked at each examined date. Instead, they inform on the share of jobs that would have been teleworkable at these dates if the likelihood to work from home depending on the content of a job had been similar to what it was in 2019 and 2021. The results should only be seen as summary measures of the evolution of labor market amenability to telework. They are then useful to compare this evolution to the actual evolution of telework in the past decades.

**Time series on teleworking** We reconstruct a time series of teleworking since 1993 by interpolating historical data points gathered from the various sources described above. Starting 2003, we also use a yearly measure of WFH prevalence from the Labor Force Survey.

**Gaps between teleworking and its feasibility** Finally, we compare the actual evolution of teleworking over the past three decades to the evolution of its feasibility. The objective is to understand if the increased implementation of teleworking in the economy has closely followed the removal of feasibility constraints, or if instead it should be attributed to other factors (e.g. cultural barriers, management practices, HR opposition, institutional context). More precisely, we compare the evolution of the actual share of jobs done from home to the following three statistics:

1. The evolution of the share of jobs that *can be teleworked* according to our Random Forest predictor fitted in 2021
2. The evolution of the share of jobs that *would be teleworked* if the adoption of telework was determined by the organizational constraints and practices during the peak of the COVID-19 pandemic.

3. The evolution of the share of jobs that *would be teleworked* if the adoption of telework was determined by the organizational constraints and practices in 2019, that is just before the peak of the COVID-19 crisis.

These comparisons allow us to discuss whether the main factors behind the evolution of teleworking before the COVID-19 crisis are technical or related to the nature of the work (e.g. Internet coverage, ICT use at work, face-to-face interactions, physical intensity of work). Since our Random Forest predictors of teleworking and its feasibility are only available for years $t = 1991, 1998, 2005, 2013, 2019$, we sometimes use simple linear interpolations to recover a measure for each year since 1991. To refine the analysis, we also present evidence for different occupations separately using the LFS since 2003.

# 3 The Tasks that Make Teleworking Feasible and their Evolution

## 3.1 Predicting Teleworking and its Feasibility in the Recent Period

Table 2 reports the results of the ML predictions for our three variables of interest: telework feasibility in 2021, and actual telework in 2021 and 2019. The first column displays the results for the first-best models (using as predictors all variables describing tasks in each survey) while the second one displays those for the second-best models (using as predictors only the five variables available in each survey).

The first-best models have a MSE smaller than 0.15 and a $Q^2$ (which is a similar to a $R^2$ in a non-linear model except that it is computed on a testing sample not used for predicting, see Appendix A) larger than 0.4. Based on 67 predictors in total, the first-best model based on the 2019 Working Conditions survey is exceptionally good at predicting teleworking, with a MSE of only 1.3 % and more than 80% of the variance in teleworking explained. Table 2 also shows that it is harder to predict teleworking ($Q^2 = 0.43$) than its feasibility ($Q^2 = 0.55$) when using as predictors the same 13 variables describing the nature of the job. This indirectly validates our choice of predictors as teleworking feasibility should indeed be more immediately impacted by the content of the job than is actual teleworking which also depends on decisions taken by firms and their employees.

The performance of second-best models decreases compared to that of first-best models but remains relatively large given the small number of predictors available. For example, with only 5 predictors in Tracov2021, the ML-model explains 36% of the variance in teleworking feasibility, which is the main outcome of interest in the rest of the paper. As we use ML predictors only to predict the evolution of the average feasibility over large groups of workers, this level of performance seems sufficient.

Last, Table 2 shows that the most important predictors in first-best and second-best models are largely overlapping. Even in models that include 13 or 67 predictors, the share of working hours spent in front of connected computers, the contact with the public or the physical intensity of the job are systematically among the main predictors.[8] This shows that, while the second-best models do not rely on a large number of predictors, they do not miss the most important dimensions underlying teleworking and its feasibility. This is a key step to move forward with the historical analysis based on the second-best predictors. Finally, when adding measures of Internet coverage at home to potential second-best predictors, we find that these measures rank systematically lower in terms of variable importance relative to our main predictors.

## 3.2 The Long-Run Evolution of Teleworking Feasibility

We now study how the predicted feasibility of teleworking has increased since 1991 given the *task mix* observed on the labor market at different points in time. Figure 2, panel (a), shows

[8]They are among the five main predictors except job's physical intensity in the model with 67 predictors as it only enters within the first ten best predictors.

that the share of jobs that can be teleworked has more than tripled over the past 3 decades, from around 14 % in 1991 to more than 45 % in 2021. The increase is close to linear between 1991 and 2019. Feasibility then jumps by more than 10 percentage points between 2019 and 2021. This jump is driven by an adaptation of working practices during the Covid-19 crisis to make teleworking feasible, showing that the task composition of jobs is endogenous and can be quickly adapted. The scope for adaptation in the short run remains however limited: even in a context of very strong incentives to switch to teleworking during the Covid-19 crisis, teleworking could be made possible for only an additional 10 % of the economy.

Most of the overall increase in teleworking feasibility occurred within occupations. In order to show this, we perform an empirical exercise akin to an Oaxaca-Blinder decomposition. We maintain constant the share of each 2-digit occupation at its 1991 level by weighting the 2-digit occupation level measures of our variables by the 1991 occupation weights, and consider the reweighted average of predicted teleworking feasibility (see the dotted line on Figure 2, panel a).[9] If the occupational mix of the labor market had remained that observed in 1991, 31% and 39% of jobs would be teleworkable in 2019 and 2021, respectively ; compared to 35% and 45% when taking into account the evolution of the occupational structure. Interestingly, this conclusion is obtained despite large changes in the occupational mix in the past 3 decades: for example the share of white collar workers almost doubled, from 11% to 21%, and the share of blue collar workers declined substantially, from 34% to 23%, between 1991 and 2021 (Figure A.3). This suggests that very large transformations of the nature of work have occurred within occupations, a phenomenon that we explore more directly in the next subsection.

**Accounting for Internet Coverage** Predicted feasibility is almost identical at all points in time when we include Internet coverage in workers' region of residence among the set of predictors.

[9] This reweighting is applied on top of individual sampling weights $w_{i,t}$ which are systematically used in order to get representative statistics. Formally, the weight applied to worker $i$ in occupation $o$ at time $t$ after reweighting is $w_{o,t}w_{i,t}$ where $w_{i,t}$ is the sampling weight available in the survey and $w_{o,t} = (\sum_{i \in o} w_{i,1991})/(\sum_{i \in o} w_{i,t})$.

This is despite the fact that the coverage of the Internet was much lower in the 2000s than it is now, and almost nonexistent in the 1990s (see Figure A.1). This result possibly reflects that the Internet is widespread across the French territory in 2021 and therefore does not represent anymore a significant constraint on teleworking. As a consequence, the ML predictors fitted in 2021 may not give a strong role to Internet coverage and they maynot be able to capture well the likely additional constraints induced by the limited coverage of (high-speed) access to the Internet in the 1990s and 2000s. In what follows, we will therefore use an ML predictor that does not include measures of Internet coverage. A disadvantage of this predictor is that it might overestimate teleworking feasibility in earlier years.[10] We show in Section 4.1 that this does not alter our conclusions.

## 3.3 The Long-Run Evolution of Tasks that Make Teleworking Feasible

Panel (b) of Figure 2 shows the evolution of the key tasks that make teleworking feasible. The most striking feature of this evolution is the continuous increase of the share of hours worked spent using connected computers, from 11 % in 1991 to more than 45 % in 2019. This means in particular that slightly less than half of labor input in the French economy was provided in front of a computer before the Covid-19 crisis. The share of hours spent in front of a computer has then jumped to almost 60 % in 2021.

Only 30 % of jobs are not physically demanding, and this share has remained remarkably stable over the past three decades. Finally, the share of jobs requiring to be in contact with the public (e.g. clients, customers, patients) is large and has been slowly increasing from just above 60 % in 1991 to above 70 % in 2019. This trend tends to make teleworking less feasible in the present than in the past. We see however that the extent of contacts with the public has been

[10]The overestimation may not be very strong. According to our predictors, the share of jobs that could be teleworked was already quite limited by the task content of jobs and, most importantly, these jobs may not have required an Internet access at the time: a phone and possibly a non-connected computer at home may have been considered sufficient to perform them.

strongly reduced during the Covid-19 crisis, illustrating that this feature of jobs can be adjusted more rapidly than others when incentives to do so are strong.

Consistent with our analysis of teleworking feasibility, we also decompose in Panel (b) of Figure 2 the evolution of the task content of jobs in within- and between-occupations variations. Most of the changes occur within occupations. In particular, the dramatic increase in the time spent in front of connected computers is not primarily driven by the increase in the share of white-collar workers in the labor force.

# 4 Gaps Between Teleworking and Teleworking Feasibility

In this section, we compare actual teleworking with its feasibility. We do so both over time and across main occupational groups.

## 4.1 Comparison of Teleworking and Teleworking Feasibility since 1991

Figure 3, panel (a), provides a direct comparison of teleworking and its feasibility over time. Teleworking has remained very limited compared to its feasibility until the Covid-19 crisis. Even in the Covid-19 period (2021) and the beginning of the post-Covid-19 era (2022), teleworking remained significantly below its predicted (and for the later years, observed) feasibility. To further illustrate this point, we plot in panel (b) the ratio and the absolute gap between teleworking and its predicted feasibility. The ratio was close to 0 in the early 1990s when teleworking was almost non-existant according to our archive data sources. It then raised quickly to 10 % and remained remarkably stable at that level between 1997 and 2016. Hence, during almost two decades after the advent of teleworking, it was only one tenth of teleworkable jobs that were teleworked. During this period, the gap between teleworking and its feasibility steadily increased from 10 to 30 percentage points. A change in trend started to occur in 2016, with the ratio between teleworking and its feasibility rapidly increasing from 10 % to 20 % between 2016 and 2019, while the absolute gap stabilized. This increase coincides with a law passed in 2017 to facilitate firm-level collective agreements allowing for teleworking on a large scale, suggesting that the institutional environment can also play a role.

The wedge between teleworking and its feasibility has then been reduced further after the advent of the Covid-19 crisis, but it remained substantial: in early 2021, only 50 % of workers who can telework do it at least to some extent, and a gap of 20 percentage points remains between actual teleworking and its declared feasibility.

These comparisons suggest that technical and feasibility constraints were not the main drivers that prevented workers from teleworking before the Covid-19 crisis, and that even after the pandemic, teleworking remains largely under-used compared to its feasibility. This conclusion is reinforced by the fact that our measure of teleworking is broad: workers are counted as teleworkers as soon as they do telework at least half a day per week. Considering measures of teleworking at higher intensity would increase further the gap between teleworking and its feasibility.

Similarly, considering the additional constraints imposed on feasibility by the availability of (high-speed) Internet in the region of residence does not alter our main conclusions. First, in the recent period, these constraints have little impact on the feasibility of teleworking as almost all the French territory has access to an Internet connection (Figure A.1). Hence, they cannot explain the large gap between actual teleworking and its feasibility. Second, as discussed in Section 3.2, these constraints imply that we might overestimate teleworking feasibility in the past decades. Considering them would therefore lead us to conclude that the share of teleworkable jobs that were indeed teleworked decreased over the period 1997-2016 (instead of remaining stable), while the absolute gap between teleworking and its feasibility had increased even more than observed on Figure 3, panel b). Hence, independently of whether Internet access constraints are taken into account or not, we can conclude that the actual recourse to teleworking given its feasibility has remained very limited during the three decades that preceded the Covid-19 crisis.

Figure 3, panel (a), also compares the evolution of teleworking since 1991 to what it would have been if the recourse to teleworking given the task content of jobs had remained similar to (i) what is observed in 2019 (teleworking predicted from job tasks in 2019), and (ii) what is observed during the COVID-19 pandemic (teleworking predicted from job tasks in 2021). Actual teleworking practices have increased slightly more rapidly—i.e., from a lower level in 1991—than we predict based on the tasks and teleworking level observed in 2019. This is largely driven by the late increase in the 2010s, which is compatible with a gradual removal of non-technical barriers to teleworking during the few years before the pandemic. However, teleworking has increased much more slowly and was much lower than one would predict had the link between tasks and teleworking remained similar to the one observed in 2021. This illustrates that the prevalence of teleworking jumped during the Covid-19 crisis not only because jobs and their task content were adapted, but also, and mostly, because a much larger share of the jobs that could already be done remotely given their task content have started to be teleworked.

## 4.2 Evolution of WFH and WFH Feasibility by Occupation: 1991-2022

In this section, we study the gaps between actual teleworking and its feasibility for different occupations.

**Identifying Relevant Groups of Occupations** Figure 4 shows, based on Tracov2021, actual teleworking and teleworking feasibility for each 2-digit occupation in early 2021. We have split these occupations in four groups. The first includes all "non-connected employees", that is all occupations with a share of teleworking below 20% in the Covid-19 period. All those occupations have a teleworking feasibility below 30% and will not be our main focus hereafter. They include all blue-collar workers (occupations starting with digit "6"); most low-skilled clerical occupations (occupations starting with digit "5") and one mid-skill occupation (Health and social workers). None of the high-skilled occupations (so-called *Cadres* in French and starting with digit "3") falls into that category.

The second group, named “low-skilled connected employees”, corresponds to one large low-skilled occupation: administrative employees. More than 70% of employees can telework in this group, and about 45% do so. The third (“mid-skilled connected employees”) includes 4 mid-skilled occupations, all having a teleworking feasibility above 50%, and the fourth includes all 2-digit high-skilled occupations (also called “executives” hereafter) which all have a feasibility above 60%. Except for professors and science-related high-skilled occupations, the gaps between teleworking and its feasibility are smaller in the fourth group than in the second and the third ones.

**Teleworking Feasibility by Occupation Group since 1991** Figure 5, panel (a), shows the evolution of predicted teleworking feasibility for each of these four groups obtained using the historical working conditions surveys since 1991. A striking result is that feasibility used to be higher for connected low-skilled employees (i.e. office workers) than for executives. In 1991, 40 % of connected low-skilled employees could telework, while it was the case only for 23 % of executives. This is largely explained by the early and high prevalence of ICT tools amongst office employees. High-skilled workers saw a stronger transformation of their job task content, so that they progressively caught up with connected low-skilled employees in terms of teleworking feasibility. Just before the Covid-19 crisis, 60 % of executives and 65 % of connected low-skilled employees could telework. Teleworking feasibility was identical among mid-skilled connected occupations and high-skilled occupations in 1991, but it increased less over time in the former group. Finally, teleworking feasibility only increased slightly and it remained limited—below 20 %—for “non-connected” employees over the studied period.

During the Covid-19 crisis, the increase in tasks that make teleworking feasible was mostly observed for mid-skilled connected occupations and executives. The latter group eventually became the most teleworkable occupations with almost 85 % of jobs that could be teleworked. On the contrary, the absence of a large break in the time series for office employees (low-skilled connected workers) suggests that much of the factors that would make teleworking possible for them have remained relatively stable throughout the Covid-19 crisis.

Overall, these differences across occupations appear to be driven mostly by the use of connected computers and the amount of physically demanding tasks to perform (Figure A.4). The share of hours spent on a connected computer has been historically much larger for connected employees. In this group, 42 % of hours worked were already spent using computers in 1991 (60% in 1998), in comparison to only 18 % and 21% of hours worked among high-skilled occupations and mid-skilled connected occupations, respectively, and 5 % among non-connected employees. Executives saw the largest transformation of their job content with a quadrupling of their time spent in front of a computer between 1991 and 2019 (from 18 to 72 %). This is the main feature explaining why teleworking feasibility increased more for them than for the other groups, with other factors being more comparable across groups and much more stable over time.

**Gaps between Teleworking and its Feasibility by Occupation Group since 2003.** For years prior to 2019, only aggregated statistics on teleworking are available. In order to provide statistics by occupation, we therefore use the LFS and focus on WFH during regular hours that we can observe since 2003. Panel (b) of Figure 5 shows that, even if teleworking has been more feasible for connected employees than for executives, WFH during regular hours has been much more implemented for executives than for other groups since 2003. This remains true after the Covid-19 crisis in 2023. Consequently, the ratio between WFH and teleworking feasibility is much larger for high-skilled workers. It ranges between 5 and 10 % for connected low-skilled employees before the Covid-19 crisis. It is slightly larger for connected mid-skilled occupations and, strikingly, three to five times larger among executives (panel (c)).[11] This shows that conditional on its feasibility, the actual recourse to WFH varied substantially across occupations, with

[11]The ratio between WFH and teleworking feasibility is actually largest among non-connected workers but as feasibility remains low in this group, the ratio should be interpreted with more caution—to put it differently, the ratio is large but the gap in percentage point between WFH and its feasibility remains small.

the most skilled occupations being much more likely to use it.

In all four groups, the ratio strongly increased during the Covid-19 crisis. In relative terms, the most important increase is that of connected low-skilled employees, whose ratio increases almost fivefold over a two-year period. This large increase also illustrates that WFH was particularly under-used compared to its feasibility among connected low-skilled employees before the Covid-19 crisis, hence making possible the large increase observed.

## 4.3 Organizational Barriers to WFH

Why is the recourse to teleworking so limited relative to its feasibility? And why are there such large differences in the use of WFH across occupations given its feasibility? In this section, we provide some evidence on potential supply and demand side factors, namely workers' desire to WFH (i.e. to supply work done from home) and management practices.

**Workers' Desire to Telework by Occupation** Workers' desire to telework is directly observed in the Tracov2021. Figure 6 shows, for each occupation category, how workers that telework or can telework are split according to their desire to telework. There are two main observations. First, among workers that could telework but do not, the share of those who would like to telework is larger than the share of those who would not. This is true for all four skill groups. Second, also in all skill groups, the share of teleworkers who would prefer not teleworking is lower than the share of non-teleworkers who would like to do so (and could do it). In other words, if teleworking were perfectly aligned with workers' preferences in 2021, it would increase further. This shows that the remaining wedge between teleworking and its feasibility still observed in 2021 cannot be explained by supply-side factors. Even if the Covid-19 crisis may have boosted workers' desire to telework for health-related reasons, we think that this conclusion is likely to hold as well prior to the crisis. This is because much fewer workers were teleworking before the pandemic and only a few of the many workers that were pushed to telework during the Covid-19 crisis declare they would prefer not to telework.

Figure 6 also suggests that supply-side factors do not explain differences in teleworking *across skill groups*. Indeed, the share of workers who could telework and do not but would like to, is largest among connected employees, while the share of workers who telework but would prefer not to is largest among executives. If teleworking was fully aligned with workers' preferences (among workers that can telework), the share of teleworkers would increase from 7 to 11 % among non-connected employees, from 41 to 57 % among connected employees, from 51 to 64 % among connected mid-skilled, and from 66 to 74 % among executives. Hence, if workers' desire to telework was taken into account, differences across skill groups would decrease.[12]

**Subordination, Control and Teleworking** We focus on two main dimensions of the collective organization of work that might influence employers' willingness to allow their employees to telework: the degree of autonomy and the extent of hierarchical control at work. These variables are natural candidates because they differ between low-, mid- and high-skilled workers (Table A.3) and are likely to be related to managers' willingness to let their subordinates telework. Indeed, teleworking implies a more remote and difficult control of workers' effort at work. In settings where workers are closely monitored (e.g. because the management worries about shirking or because the job is highly collaborative), managers may be more reluctant to allow them to telework as this would mean a partial loss of their ability to control them, or a higher cost of interaction.[13]

We first test this hypothesis during the Covid-19 crisis in 2021 using a series of linear regres-

[12]This larger mismatch between desired and actual WFH down the earnings distribution is consistent with what Barrero et al. (2021) have predicted for the U.S. in the post-Covid-19 era.

[13]In some jobs, such as call centers (Bloom et al., 2015), individual productivity of remote workers can be directly monitored and this argument does not apply.

sions. Namely, we estimate the following models using OLS:

$$T_{ij} = \alpha Feas_{ij} + \beta Occ_i + \gamma Feas_{ij} * Occ_i + \delta X_{ij} + \delta' X'_{ij} + \varepsilon_{ij}$$

where $T_{ij}$ is an indicator variable equal to 1 if worker $i$ in firm $j$ is teleworking at least half a day per month, $Feas_{ij}$ an indicator variable equals to 1 if worker $i$ indicates that her job can be teleworked, $Occ_i$ three occupation indicators for connected low-skilled, connected mid-skilled and high-skilled workers (non-connected employees are the reference group), $X_{ij}$ a set of variables describing either worker $i$'s desire to work from home or her degree of autonomy and control in her job, and $X'_{ij}$ additional controls for workers' demographics and their firms' size and industry.[14]

Results in Table 3 show that teleworking feasibility strongly predicts actual teleworking, as do occupations. The correlation between teleworking feasibility and actual teleworking is much stronger for executives than it is for mid-skilled connected employees, and much stronger for the latter than it is for low-skilled connected employees. These differences remain largely unchanged even after controlling for a broad set of controls, including workers demographics or desire to work from home, workers' degree of autonomy and of hierarchical control. The latter do exhibit some correlation with the practice of teleworking, but of small economic magnitude.

These results largely carry over to the pre-Covid period, as we show in Table 4. Like in 2021, the prevalence of teleworking in 2019 strongly depends on its (predicted) feasibility, and on broad occupational groups: executives (connected mid-skilled workers) are 22 (4) percentage points more likely to telework than are connected low-skilled workers. The partial correlation between actual teleworking and teleworking feasibility is again much stronger for executives than

[14] $\alpha$ captures the link between teleworking and teleworking feasibility and $\beta$ the gaps in teleworking between the three occupational groups conditional on teleworking feasibility. $\gamma$ captures differences in the link between teleworking and its feasibility across occupations. $\delta$ captures the relationship between supply and demand factors and teleworking. We examine as well if controlling for these latter factors narrows the gaps between occupations conditional on feasibility, in the spirit of Oster (2019).

it is for connected mid-skilled or connected low-skilled workers (column 2). These differences across occupations are largely robust to controlling for basic workers' demographic characteristics (column 3) and management practices, workers' autonomy and hierarchical control and work organization (column 4). These latter variables affect teleworking beyond feasibility constraints: facing no deadline in one's daily work is associated with an almost 4 percentage points lower probability to telework, while having no hierarchy is associated with a 6 percentage point larger probability to telework. Given that less than 10 % of the workforce telework in 2019, these magnitudes are large, clearly showing that the organisational aspects of work and management practices were related to the probability of teleworking before the Covid-19 crisis. We further control for workers' education and firm size, that are observed in the 2019 working conditions survey. Their inclusion as controls actually reduces by 4 percentage points the premium of executives in terms of teleworking and changes the ordering between connected low-skilled and connected mid-skilled (column 5). These changes are driven by the control for education, revealing that workers' background, even for jobs that are similar in many other dimensions, may also influence their probability to work from home.

Two broad conclusions emerge from the analyses above.[15] First, management practices and the collective organization of work (e.g. giving workers autonomy, deadlines or having a lot of hierarchical links) are related to teleworking beyond feasibility constraints. Their link with teleworking used to be quantitatively large in 2019 but became less significant in 2021, when teleworking was much more determined by feasibility constraints. Second, there are very large differences in the prevalence of teleworking between high-skilled and low-skilled occupations, even when comparing occupations and jobs that can be done from home to the same extent. These differences cannot be accounted for by feasibility constraints, neither by supply-side factors (workers' desire to telework and their attitudes and risks with respect to the pandemic in 2021). They are also largely unexplained by management practices that directly relate to tele-

[15] See Appendix C for a detailed discussion of the regression results and additional robustness checks.

working and differ across occupations.

### 4.4 Remaining Drivers of the Gaps across Occupations

If observed management practices or workers' desire to telework do not account for much of the large gaps across occupations in teleworking conditional on its feasibility, what are the remaining potential explanations?

The first is akin to taste-based discrimination: managers may have a distaste or incur a larger cost for managing remotely low-skill workers, for example because it is more complicated to do so. It could be that providing explanations to lower-skill workers online is just more complicated and costly for managers (e.g. communication is hard). They may enjoy a direct utility from monitoring and supervizing their subordinates from lower-skill groups in face-to-face interactions, which is reduced with remote monitoring. Indeed, remote monitoring introduces frictions in managerial relationships which may be larger for lower-skill workers. Managers may find it harder to reach out to their subordinates as easily and directly as they would in an office setting. Walking over to an employee's desk can allow for quicker resolution of an issue, compared to communicating via chat or scheduling a remote call.

The second explanation is akin to statistical discrimination: employers may have more negative priors regarding the productivity of low-skilled workers. This could be because these workers are typically less work-oriented than their higher-skill counterparts and may be thought to be more likely to shirk. As a consequence, managers may be less willing to trust low-skilled workers to exert sufficient effort when working from home.

Both explanations are likely to play a role in explaining the remaining teleworking gap across occupations. Consistent with our findings, Lewandowski, Piotr et al. (2023) provide evidence that employers are less willing to let workers holding routine tasks telework than they are for non-routine workers. They further show that employers' unwillingness to let workers telework is both explained by negative priors regarding productivity and a personal cost of managing remote workers. Using Tracov2021, we find suggestive evidence of statistical discrimination against low-skilled connected workers: the gaps across occupations conditional on teleworking feasibility are much smaller among workers with higher tenure. For example, the gap between executives and office employees is reduced from 23 among workers with less than a year of tenure to 8 percentage points among workers having more than 5 years of tenure (Table A.4). These results are consistent with the idea that employers may initially hold negative priors regarding the productivity of low-skilled workers at home, and that these priors disappear to a large extent when they get more time to observe their employees and consider them trustworthy. [16]

The gaps across occupations are also smaller among workers whose main task is to supervize others and workers who often take initiatives or who do not need much help from their supervisors (Table A.5).[17] Supervision tasks and autonomy as measured by these variables are likely offered to workers that are considered trustworthy by their hierarchy. Hence, the lower gaps in teleworking among workers featuring such work practices suggest that trust of managers is likely to be an important determinant of teleworking among low-skilled workers

# 5 Inequality in Access to Teleworking

The gaps in teleworking across occupations conditional on its feasibility suggest that the well-documented inequality in access to teleworking along the earnings distribution (Barrero et al., 2021; Bonacini et al., 2021) is not only driven by feasibility constraints. To back-up this point and conclude the paper, we compare teleworking and its feasibility along the earnings distribution in 2019. Figure 7, panel a), first shows that the probability for a worker to hold a job in one of

[16]The fact that a gap remains among workers with more than 5 years of tenure could indicate that there still are asymmetries of information or moral hazard concerns that prevent managers from allowing some of their their low-skilled subordinates to telework. Alternatively, the gap could suggest that for a fraction of low-skilled workers, their productivity is lower when working from home. Finally, it could reflect that managers are engaging in taste-based discrimination or have a larger cost in supervizing low-skilled workers remotely as discussed above.

[17]Note that the existence of these heterogenous gaps across skill groups depending on work practices is not in contradiction with the results in the previous section showing that these practices cannot directly account for the gaps across occupations.

the three groups of “connected” occupations (connected low-skilled, connected mid-skilled, or executives) steadily increases along the earnings distribution, from 13% in the bottom decile to 90% in the top one.

Panel b) shows gaps between teleworking and its predicted feasibility *for these connected workers only*, along the deciles of the overall earnings distribution. We see that feasibility is almost flat, ranging between 50 and 60%: among connected workers, high wage-earners are not more likely to have a job that can be done from home. However, they are more likely to telework, implying that the share of jobs that are teleworked among those that are teleworkable increases along the earnings distribution, from around 10% in the bottom decile, to more than 30% in the top one. This increase takes place in the top half of the earnings distribution.

In panel c), we include all workers, so that variations in teleworking along the earnings distribution capture both the increase in teleworking feasibility and the increase in the intensity of teleworking given its feasibility. The key result is that there are large inequalities in access to teleworking given its feasibility: 50% of high-wage earners who can telework do so in 2019, while this is the case for only 10 to 20% of low wage earners.

# 6 Conclusion

Our analysis shows that the obstacles to teleworking over the past thirty years have been mostly non-technical and beyond feasibility constraints. Teleworking and WFH practices have increased in two phases. The first saw a gradual increase from the early nineties to the brink of the Covid-19 crisis. Over this period, work practices that are today the strongest predictors of telework feasibility, in particular the share of total working time that workers spend using ICT and computers, have increased continuously. If these were the only factors to determine teleworking, one would have expected it to increase much earlier, especially for office workers, who were already spending more than half of their working time on ICT tools by the late nineties.

During the Covid-19 crisis, teleworking increased dramatically for the workers for whom it was feasible. For executives, teleworking came closer to its full potential and their desired level, while their actual job tasks evolved substantially in the direction of making teleworking more feasible. For office employees, a larger gap remains between actual teleworking and its feasibility, implying that these workers are still far from their desired level of teleworking. Working practices that make teleworking feasible were also much less affected during the Covid-19 crisis for these workers. This shows that the Covid-19 crisis has alleviated some the non-technical obstacles to teleworking, but not all of them, and mostly for skilled workers, leaving suybstantial room for extending teleworking among their low-skilled counterparts.

The gaps in teleworking across occupations conditional on its feasibility have been observed since the early 2000s. They are very robust statistically. Some of them can be explained by organisational factors, such as autonomy or monitoring. However, a larger share remains unexplained and is likely related to employers' distaste or difficulty for managing remotely low-skilled workers and to their priors regarding the productivity of these workers at home.

This last explanation is crucial from a normative point of view. The gaps across occupations imply that the well-documented inequality in access to teleworking along the earnings distribution is not only driven by feasibility constraints. This source of inequality could therefore be alleviated by allowing low-skill workers to telework more. However, this would be efficient from an economic point of view only if employers' priors are incorrect and that low-skilled workers would be more or equally productive at home. Existing studies find contrasting results,[18] making it hard to draw general conclusions. This is likely because the productivity of (low-skilled) workers at home depends on the exact type of job considered, how teleworking is implemented and organized, and on workers' ability to adapt to a very novel work practice over time. Hence, there could be margins to efficiently extend WFH among low-skilled workers. Our results indeed

[18] See Bloom et al. (2015), Bergeaud et al. (2023), Atkin et al. (2023), Emanuel and Harrington (2024), Atkin et al. (2023) and Fenizia and Kirchmaier (2024).

suggest that independently of productivity considerations and feasibility constraints, managerial trust toward their employees and more broadly managerial discretion over the decision to work from home may limit its development.

# References


Adams-Prassl, A., T. Boneva, M. Golin, and C. Rauh (2022, January). Work that can be done from home: evidence on variation within and across occupations and industries. *Labour Economics 74*, 102083.

Alipour, J.-V., O. Falck, and S. Schüller (2023, January). Germany's capacity to work from home. *European Economic Review 151*, 104354.

Atkin, D., A. Schoar, and S. Shinde (2023). Working from home, worker sorting and development. Technical report, National Bureau of Economic Research.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review 103*(5), 1553–1597.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics 118*(4), 1279–1333.

Barrero, J. M., N. Bloom, and S. J. Davis (2021, April). Why Working from Home Will Stick.

Barrero, J. M., N. Bloom, and S. J. Davis (2023). The evolution of working from home. *Journal of Economic Perspectives 37*(4).

Bartik, A. W., Z. B. Cullen, E. L. Glaeser, M. Luca, and C. T. Stanton (2020, June). What Jobs are Being Done at Home During the Covid-19 Crisis? Evidence from Firm-Level Surveys.

Bergeaud, A., G. Cette, and S. Drapala (2023). Télétravail et productivité avant, pendant et après la pandémie de covid-19/telework and productivity before, during and after the covid-19 crisis. *Economie et Statistique 539*(1), 77–93.

Bick, A., A. Blandin, and K. Mertens (2021). Work from home before and after the covid-19 outbreak. *Available at SSRN 3786142*.

Bloom, N., J. Liang, J. Roberts, and Z. J. Ying (2015). Does working from home work? evidence from a chinese experiment. *The Quarterly journal of economics 130*(1), 165–218.

Bonacini, L., G. Gallo, and S. Scicchitano (2021). Working from home and income inequality: risks of a 'new normal'with covid-19. *Journal of population economics 34*(1), 303–360.

Breiman, L. (2001). Random forests. *Machine Learning 45*.

Breton, T. (1993). Le télétravail en france : situation actuelle, perspectives de développement et aspects juridiques. Technical report, Délégation interministérielle à l'aménagement du territoire et à l'attractivité régionale.

de Vos, D., M. van Ham, and E. Meijers (2019). Working from home and commuting: Heterogeneity over time, space, and occupations. *IZA Discussion paper series IZA DP No. 12578*.

Dingel, J. I. and B. Neiman (2020, September). How many jobs can be done at home? *Journal of Public Economics 189*, 104235.

Emanuel, N. and E. Harrington (2024). Working remotely? selection, treatment, and the market for remote work. *American Economic Journal: Applied Economics 16*(4), 528–559.

Fenizia, A. and T. Kirchmaier (2024). Not incentivized yet efficient: Working from home in the public sector. Technical report, Centre for Economic Performance, LSE.

Garrote Sanchez, D., N. Gomez Parra, C. Ozden, B. Rijkers, M. Viollaz, and H. Winkler (2021, February). Who on Earth Can Work from Home? *The World Bank Research Observer 36*(1), 67–100.

Gottlieb, C., J. Grobovsek, and M. Poschke (2020). Working from home across countries. *Covid Economics*. Publisher: CEPR.

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer.

Hatayama, M., M. Viollaz, and H. Winkler (2020, May). Jobs' Amenability to Working from Home: Evidence from Skills Surveys for 53 Countries.

Hensvik, L., T. Le Barbanchon, and R. Rathelot (2020, April). Which Jobs are Done from Home? Evidence from the American Time Use Survey.

Lewandowski, Piotr, Lipowska, Katarzyna, and Smoter, Mateusz (2023). Mismatch in preferences for working from home – evidence from discrete choice experiments with workers and employers.

Louppe, G. (2015). Understanding random forests: From theory to practice.

Mas, A. and A. Pallais (2017, December). Valuing Alternative Work Arrangements. *American Economic Review 107*(12), 3722–3759.

Mongey, S. and A. Weinberg (2020). Characteristics of workers in low work-from-home and high personal-proximity occupations. *Becker Friedman Institute for Economic White Paper*.

Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

Nagler, M., J. Rincke, and E. Winkler (2022). How much do workers actually value working from home?

Oettinger, G. S. (2011). The incidence and wage consequences of home-based work in the united states, 1980–2000. *Journal of Human Resources 46*(2), 237–260.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

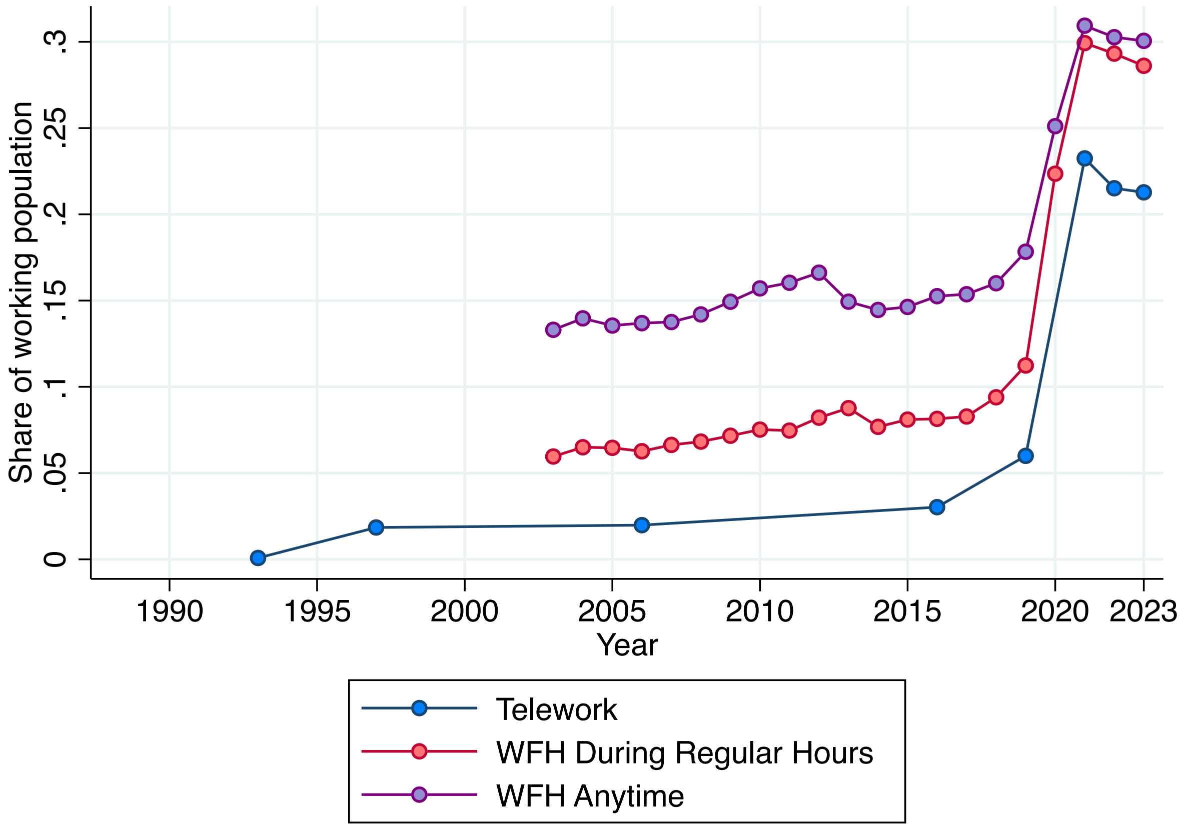
Quan, N. T. (1988). The prediction sum of squares as a general measure for regression diagnostics. *Journal of Business & Economic Statistics 6*(4), 501–504.

Saltiel, F. (2020). Who can work from home in developing countries. *Covid Economics 7*(2020), 104–118.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives 28*(2), 3–28.

# Figures and Tables

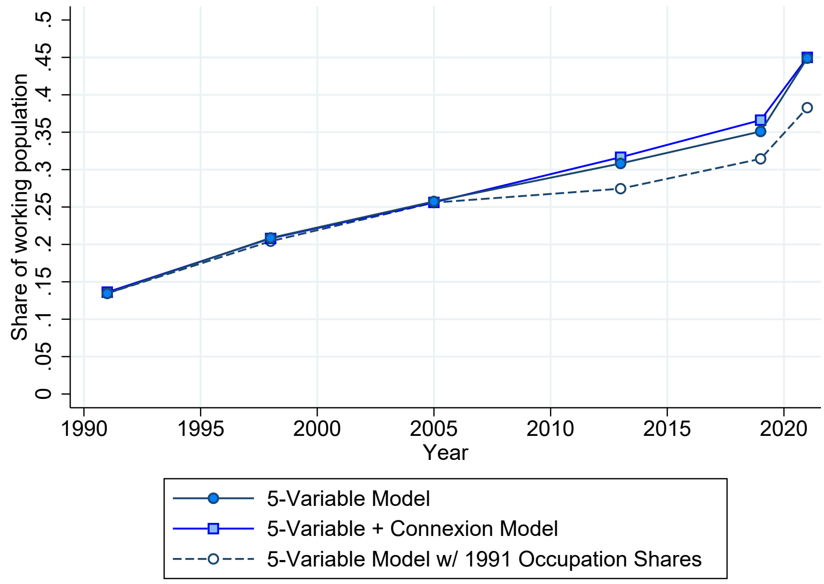
Figure 1: Evolution of teleworking and working from home since 1993



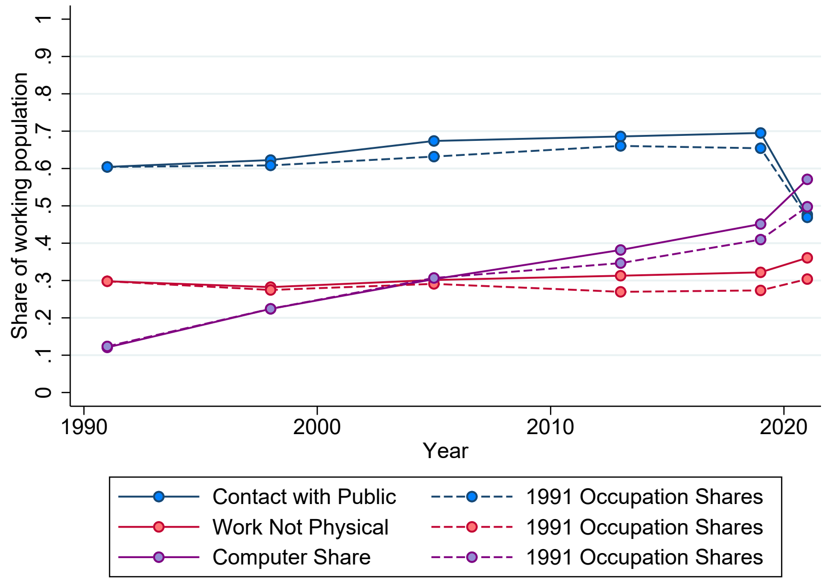

*Source:* Teleworking is obtained from various sources described in Section 2.2. WFH is measured using the Labor Force Surveys.

Figure 2: Evolution of predicted teleworking feasibility and tasks making teleworking feasible since 1991

a) Predicted teleworking feasibility




b) Evolution of the main tasks making teleworking feasible

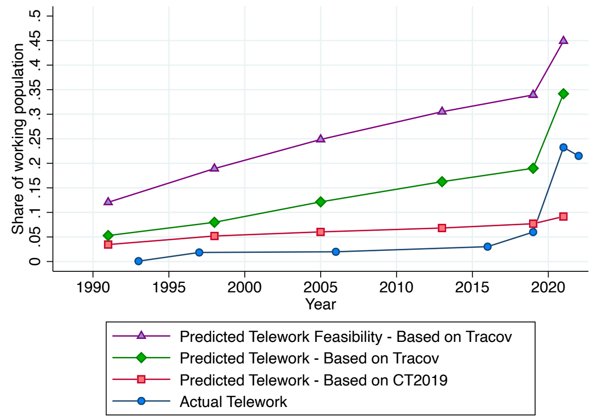



*Source:* Working Conditions surveys 1991 to 2019. Tracov survey 2021.

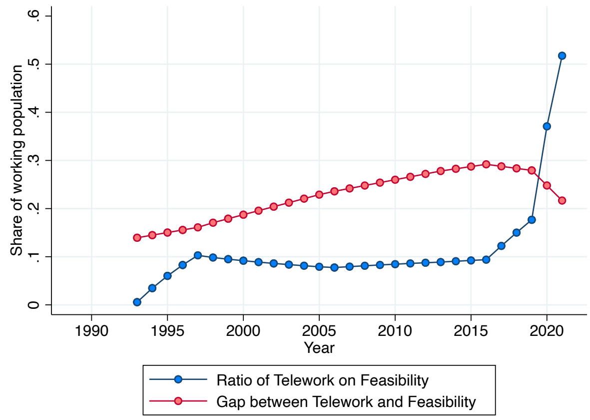
*Note:* Random forests are used to predict teleworking feasibility based on job tasks in early 2021. Before 2021, teleworking feasibility is no longer directly observed. In panel a), we use the same observed tasks as in 2021 in earlier years to predict the extent of ML feasibility in the labor market the corresponding year. In panel b), we show the evolution of the tasks that predict teleworking feasibility the most. The dotted lines show within-occupation variations, i.e. after reweighting individual observations at each period to maintain the share of each 2-digit occupation in the weighted population at its 1991 level.

Figure 3: Gaps between teleworking and teleworking feasibility since 1991

a) Actual teleworking, teleworking predicted in 2019 and teleworking feasibility



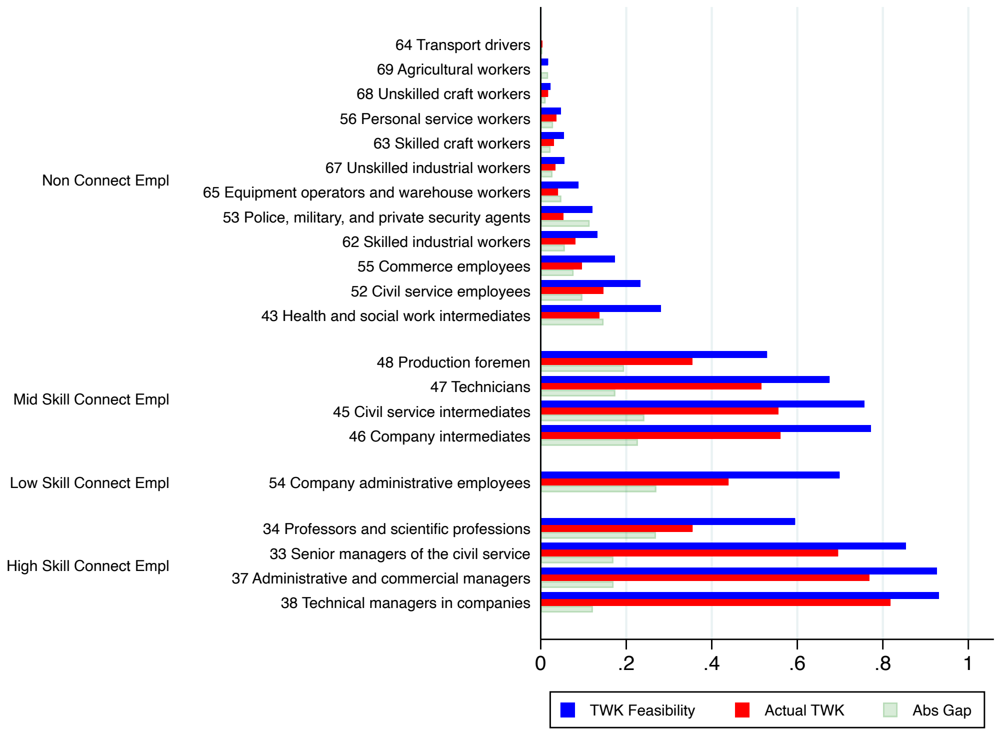

b) Ratio and gap between teleworking and predicted teleworking or teleworking feasibility




*Source:* Working Conditions surveys 1991 to 2021 for teleworking feasibility. Archive data for teleworking.

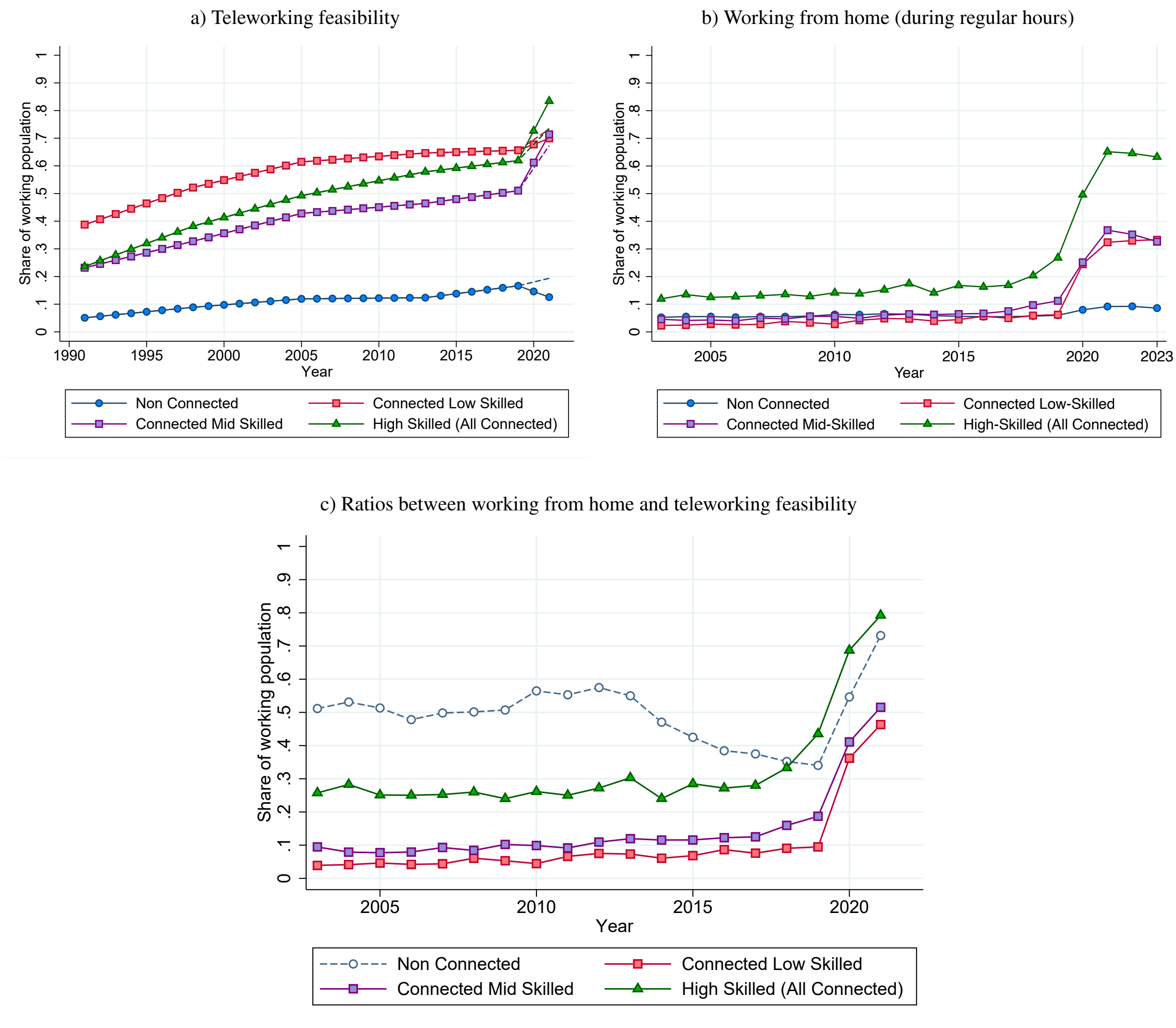
*Note:* In panel b), before computing the ratio and gap, we use a simple linear interpolation between to Working Conditions surveys to recover teleworking feasibility at each date.

Figure 4: Teleworking and teleworking feasibility by 2-digit occupation in early 2021.



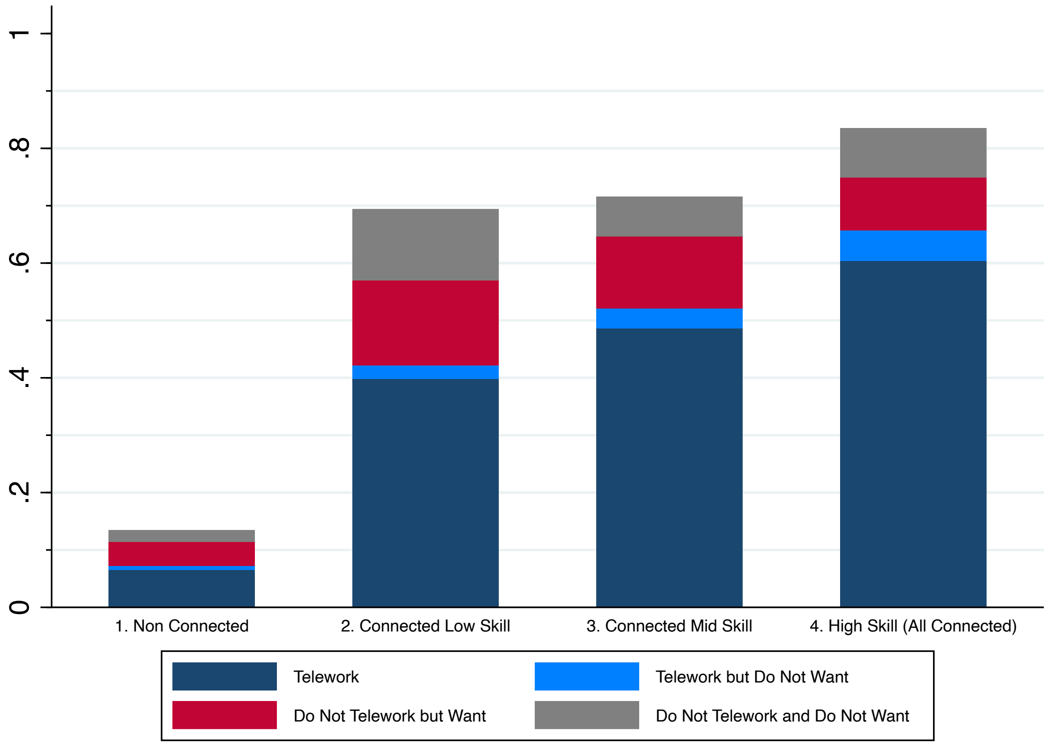

*Source:* Tracov 2021. *Note:* The Figure shows the share of jobs that are teleworked and could be teleworked in early 2021 in each 2-digit occupation. The gap between the two is also shown.

Figure 5: Evolution of working from home and teleworking feasibility by occupation




*Source:* Working Conditions surveys 1991 to 2021. Labor force survey surveys.

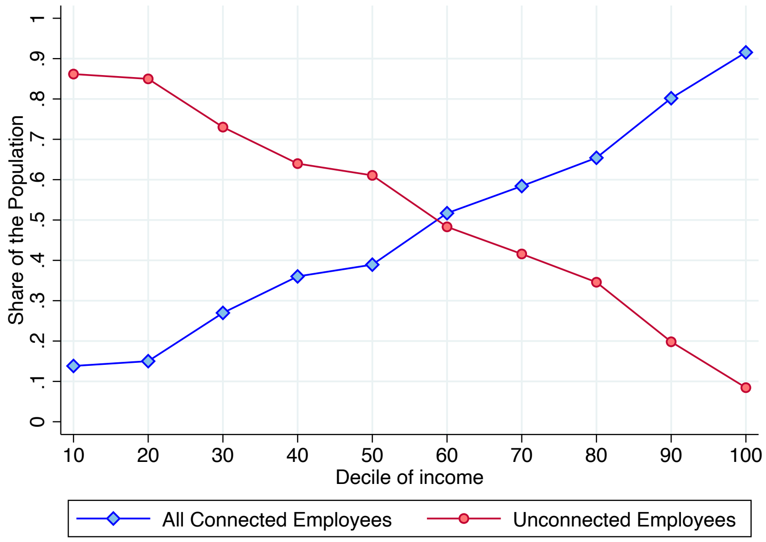
Figure 6: Mismatch between telework and desired telework by occupation categories



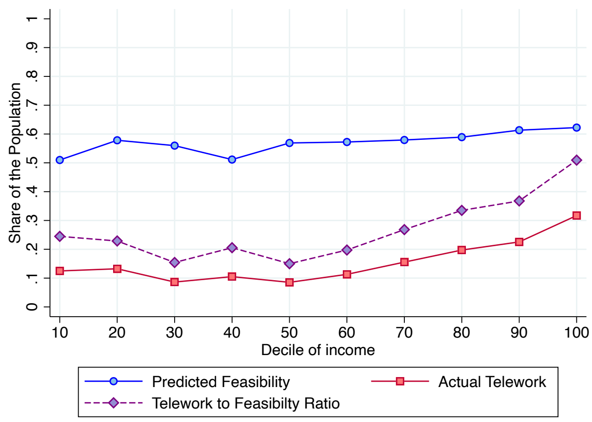

*Source:* Tracov 2021. *Note:* Weighted statistics.

Figure 7: Inequality in access to teleworking and teleworking feasibility along the earnings distribution in 2019

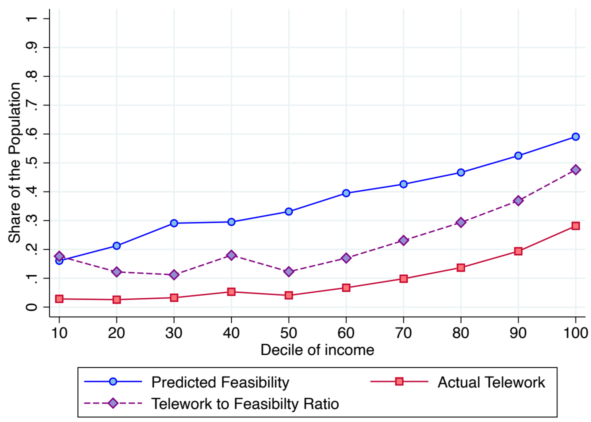
a) Share of connected workers by decile of labor income




b) Teleworking and its feasibility along the earnings distribution. Connected workers only.




c) Teleworking and its feasibility along the earnings distribution. All workers.




*Source:* WC2019. *Note:* Connected workers are those working in one of the three groups of connected occupations (i.e. occupations with a share of teleworkers larger than 30% in Tracov2021, see Figure 4). In panel (b) and (c), we compute average rate of teleworking and the average predicted teleworking feasibility in each earnings decile, as well as the ratio of the two. Panel (b) restricts the analysis to connected workers (still considering the deciles of the whole working population) while panel (c) focuses on the whole working population.

Table 1: Summary of micro-data sources

| | **Sample size (Nb workers)** | **Response rate** | **Measure of telework/WFH** | **Job task coverage** |
|---|---|---|---|---|
| *a) Working Conditions surveys* | | | | |
| Tracov 2021 | 19,953 | 39% | Telework + its Feasibility | Limited |
| WC2019 | 19,569 | 69% | Telework | Very large |
| WC2013 | 33,676 | 70% | None | Very large |
| WC2005 | 18,789 | 85% | None | Large |
| WC1998 | 21,380 | 93% | None | Large |
| WC1991 | 20,929 | NA | None | Large |
| *b) Labor Force Surveys* | | | | |
| 2003Q1 to 2022Q4 | 280,00 to 350,000 per quarterly survey | 80% to 90% | Working From Home | None |

Note: The Table describes the micro data-sources used in the paper. Details on the job tasks included in each working conditions survey are provided in Table A.1.

Table 2: Random Forest Prediction Results

| | | First Best<br>(all predictors) | Second Best<br>(common predictors) |
|---|---|---|---|
| a) Telework Feasibility 2021 | Nb Var | 13 | 5 |
| (mean= .451) | MSE | 0.111 | 0.158 |
| | $Q^2$ | 0.552 | 0.362 |
| | Five most important predictors (ordered) | Computer Share<br>Work always physical<br>Work often physical<br>Contact with public/clients<br>Work sometimes physical | Work always physical<br>Computer Share<br>Work often physical<br>Work sometimes physical<br>Contact with public |
| b) Telework 2021 | Nb Var | 13 | 5 |
| (mean= .321) | MSE | 0.142 | 0.183 |
| | $Q^2$ | 0.427 | 0.261 |
| | Five most important predictors (ordered) | Computer Share<br>Work always physical<br>Work often physical<br>Work close to other people<br>Contact with public/clients | Work often physical<br>Work always physical<br>Computer Share<br>Contact with public<br>Work sometimes physical |
| c) Telework 2019 | Nb Var | 67 | 5 |
| (mean = 0.078) | MSE | 0.013 | 0.055 |
| | $Q^2$ | 0.820 | 0.238 |
| | Five most important predictors (ordered) | Computer Share<br>Does not work alone<br>Work depends on external demand*<br>Contact with public<br>Reads small letters/numbers | Work always physical<br>Computer Share<br>Contact with public<br>Work sometimes physical<br>Work often physical |

Note: The Table summarises the performance of, and main predictors from Random Forests predictions of teleworking and teleworking feasibility in Tracov2021 and CT2019. The two columns provide results from the first-best predictors (when all items describing job tasks or the organization of work in a given survey are used as predictors) and second-best predictors (when only items common to all surveys since 1991 are used). The data sources (Tracov2021 and CT2019) are described in the main text. "Computer share" means the share of working time spent using connected computers (see details in Appendix A). In panel c), the exact predictor for "Work depends on external demand" is "Work depends on external demand that does not require immediate response". The $Q^2$ is akin to a $R^2$ in an OLS model except that it is computed on a separate test sample (not the sample used for prediction): $Q^2(Y) = 1 - \frac{\text{MSE}(Y)}{\text{Var}(Y)}$.

Table 3: Regressions of Teleworking on Actual Feasibility and Occupation Groups in 2021

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Dependent Variable: Teleworking Dummy | | | | | |
| TWK Feasible | 0.633*** | 0.601*** | 0.583*** | 0.213*** | 0.210*** | 0.161*** |
| | (0.008) | (0.008) | (0.008) | (0.012) | (0.012) | (0.012) |
| Executives | 0.149*** | 0.135*** | 0.115*** | 0.111*** | 0.106*** | 0.078*** |
| | (0.009) | (0.011) | (0.011) | (0.010) | (0.010) | (0.010) |
| Connected mid-skilled | 0.077*** | 0.092*** | 0.082*** | 0.071*** | 0.065*** | 0.037*** |
| | (0.009) | (0.009) | (0.009) | (0.008) | (0.008) | (0.009) |
| Connected low-skilled | 0.003 | 0.040*** | 0.039*** | 0.033*** | 0.030** | -0.016 |
| | (0.012) | (0.013) | (0.013) | (0.012) | (0.012) | (0.012) |
| TWK Feasible X Executives | | 0.206*** | 0.189*** | 0.164*** | 0.160*** | 0.165*** |
| | | (0.021) | (0.021) | (0.019) | (0.019) | (0.019) |
| TWK Feasible X Connected mid-skill | | 0.146*** | 0.135*** | 0.108*** | 0.108*** | 0.112*** |
| | | (0.018) | (0.018) | (0.017) | (0.017) | (0.017) |
| TWK Feasible X Connected low-skill | | 0.061** | 0.070*** | 0.069*** | 0.073*** | 0.096*** |
| | | (0.026) | (0.025) | (0.024) | (0.024) | (0.024) |
| Wants to TWK | | | | 0.458*** | 0.459*** | 0.444*** |
| | | | | (0.011) | (0.011) | (0.011) |
| No hierarchy | | | | | 0.004 | 0.002 |
| | | | | | (0.011) | (0.011) |
| Degree of Hierarchy Control | | | | | -0.006* | -0.006* |
| | | | | | (0.004) | (0.004) |
| Autonomy | | | | | 0.026*** | 0.024*** |
| | | | | | (0.003) | (0.003) |
| Computer Share | | | | | | 0.070*** |
| | | | | | | (0.010) |
| Contact with Public | | | | | | -0.049*** |
| | | | | | | (0.006) |
| Work Always Physical | | | | | | -0.046*** |
| | | | | | | (0.010) |
| Work Often Physical | | | | | | -0.071*** |
| | | | | | | (0.009) |
| Work Never Physical | | | | | | -0.054*** |
| | | | | | | (0.007) |
| Observations | 11,435 | 11,435 | 11,435 | 11,417 | 11,376 | 11,372 |
| R-squared | 0.571 | 0.576 | 0.592 | 0.652 | 0.655 | 0.663 |
| Covid-related Controls | No | No | No | Yes | Yes | Yes |
| Demographic Controls | No | No | Yes | Yes | Yes | Yes |

Note: Covid-related Controls capture the fear of catching Covid-19 while commuting, being particularly vulnerable to Covid-19, living with people particularly vulnerable to Covid-19, and working in an environment when catching Covid-19 is likely. Demographic controls include gender, age, seniority, continent of Birth, region of living. Firm size controls are a second order polynomial in the number of employees. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4: Regressions of Teleworking on Predicted Feasibility and Occupation Groups in 2019

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Dependent Variable: Teleworking Dummy | | | | | |
| Predicted TWK feasibility Prediction | 0.093*** | 0.088*** | 0.088*** | 0.080*** | 0.061*** | 0.027 |
| | (0.008) | (0.008) | (0.008) | (0.009) | (0.009) | (0.023) |
| Executives | 0.236*** | 0.204*** | 0.197*** | 0.177*** | 0.133*** | 0.111*** |
| | (0.006) | (0.008) | (0.008) | (0.008) | (0.010) | (0.010) |
| Connected mid-skilled | 0.061*** | 0.068*** | 0.064*** | 0.054*** | 0.033*** | 0.027*** |
| | (0.006) | (0.006) | (0.006) | (0.007) | (0.007) | (0.007) |
| Connected low-skilled | 0.019** | 0.054*** | 0.046*** | 0.044*** | 0.041*** | 0.029* |
| | (0.009) | (0.015) | (0.015) | (0.015) | (0.015) | (0.015) |
| Pred TWK feas. X Executives | | 0.127*** | 0.103*** | 0.096*** | 0.083*** | 0.114*** |
| | | (0.021) | (0.022) | (0.022) | (0.022) | (0.024) |
| Pred TWK feas. X Connected mid-skilled | | -0.010 | -0.014 | -0.023 | -0.037* | -0.037* |
| | | (0.019) | (0.019) | (0.019) | (0.019) | (0.020) |
| Pred TWK feas. X Connected low-skilled | | -0.078** | -0.069** | -0.078** | -0.097*** | -0.079** |
| | | (0.035) | (0.035) | (0.035) | (0.035) | (0.036) |
| Autonomy in Deadlines | | | | 0.002 | -0.002 | -0.004 |
| | | | | (0.005) | (0.005) | (0.005) |
| No Deadlines | | | | -0.038*** | -0.037*** | -0.036*** |
| | | | | (0.006) | (0.006) | (0.006) |
| Controlled by Hierarchy | | | | -0.019*** | -0.012*** | -0.010* |
| | | | | (0.005) | (0.005) | (0.005) |
| Can Interrupt Work | | | | 0.014*** | 0.012** | 0.007 |
| | | | | (0.005) | (0.005) | (0.005) |
| No orders | | | | 0.000 | 0.005 | 0.008 |
| | | | | (0.010) | (0.010) | (0.010) |
| No Hierarchy | | | | 0.059*** | 0.071*** | 0.062*** |
| | | | | (0.011) | (0.013) | (0.013) |
| Autonomy regarding orders | | | | 0.013*** | 0.009** | 0.006 |
| | | | | (0.004) | (0.004) | (0.004) |
| Autonomy regarding workload | | | | -0.001 | -0.001 | -0.001 |
| | | | | (0.002) | (0.002) | (0.002) |
| Works autonomously | | | | 0.008*** | 0.010*** | 0.006*** |
| | | | | (0.002) | (0.002) | (0.002) |
| Degree of supervision of other | | | | 0.004* | 0.006*** | 0.006*** |
| | | | | (0.002) | (0.002) | (0.002) |
| Observations | 17,583 | 17,583 | 17,530 | 17,342 | 16,946 | 16,458 |
| R-squared | 0.142 | 0.144 | 0.155 | 0.164 | 0.209 | 0.224 |
| Demographic Controls | No | No | Yes | Yes | Yes | Yes |
| Education and firm size | No | No | No | No | Yes | Yes |
| Complete Task Variables | No | No | No | No | No | Yes |

Note: Demographic controls include gender, age, seniority, continent of Birth, region of living. Firm size controls are a second order polynomial in the number of employees. "Complete task variables" corresponds to the 67 identified task variables used for the first-best predictor of teleworking in 2019. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Online Appendix to

# Gaps between working from home and its feasibility over time and across occupations

Thomas Breda, Paul Dutronc-Postel and Vladimir Pecheu

Friday 25th October, 2024

## List of Appendices

# A Details on Machine Learning Prediction

**The Random Forest Method** As mentioned in the main text, Random Forest is a non linear technique that predicts an outcome based on a set of optimized tree classifiers. Each tree is constructed to predict the outcome variable with a number of randomly drawn explanatory variables among a set of potential predictors. Figure A.2 shows an example of a tree classifier of telework in 2021 on three variables: computer share, a dummy for whether the job is physically demanding, and a dummy for whether the individual interacts with the public. The tree predicts that an individual would telework if she work on a computer more than 86 percent of the time and her job is not physically demanding, or if her computer share is between 50 percent and 86 percent and her job is not physically demanding nor implying interaction with the public.

The outcome of the tree can be either continuous (the tree predicts the probability of teleworking in a given leaf as the share of individuals in the sample who telework in that leaf) or categorical (the prediction is 0 or 1 depending on the share being lower or larger than 50%). Because our objective is to best predict the average share of the population that telework, or can telework within large subgroup of workers, we favor continuous models. In that case, the outcome of the tree classifier can be interpreted as the probability that the individual teleworks.

A forest is a collection of tree classifiers, whose outcome for a given individual is the average prediction over those of its trees. Randomly drawing and limiting the set of explanatory variables used to predict the outcome prevents over-fitting the model, as opposed to the simple tree and forest models. This ensures that out-of-sample performance is maximized.

In practice we tune the number of explanatory variables randomly drawn using cross-validation. For each model with a given number of explanatory variables, we divide our training sample into 10 subsamples, and repeatedly compute the cross-validation error given a number of features randomly drawn. We do that by alternatively estimating the Random Forest model on 9 subsamples pooled together, and then predicting its mean square error on the remaining one. We then average these 10 mean square errors (MSE) and repeat the procedure for all possible numbers of explanatory variables to be randomly drawn. We eventually pick the model with the number of randomly drawn explanatory variables that gives the minimal MSE. The formula of the MSE is the following:

$$MSE(Y) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2 \tag{A.1}$$

where $Y_i$ is the outcome variable for $i$ , and $\tilde{Y}_i$ is the prediction of this variable given the characteristics of the job of individual $i$.

The other tuning parameter that we can adjust is the number of trees that constitute the forest. In principle the larger the forest, the better the prediction. We therefore set this number to be very large: 500. For more on the details of the random forest technique see Breiman (2001) and Louppe (2015).

**Construction of Predictors for First- and Second-Best Predictions and Evaluation of their Performance** In order to evaluate the quality of our predictions, we proceed with the cross-validation method described above on the training sample of our data, as already mentioned. This training sample is a randomly drawn subsample of 80 percent of observations of our base data set. After having fine tuned our Random Forest model, we then estimate it and compute its mean square error on the 20 percent remaining observations, which is the prediction sample. Our first-best models are trained using as potential predictors all variables describing tasks or the nature of the job. The second-best models only use as predictors the variables that are common across all surveys in different years.

Note that we split non-continuous variables with multiple modalities into separate dummies, e.g. work never physically demanding, sometimes physically demanding, often physically demanding, always physically demanding. We do that for three reasons. First, the initial coding of these variables (usually a series of integers) is not necessarily the right scale to order modalities, especially when NA-type modalities arise. Second, because questions do not always have the same modalities in different surveys, this allows a better matching of the variables across surveys. Third, this improves the flexibility and therefore performance of the model. Hence, all predictors are either continuous or dummy variables.

We evaluate the performance of the models using two criteria: the Mean Square Error, which has already been presented above, and the Q-Squared, as in Quan (1988), defined for outcome variable $Y$ as follows:

$$Q^2(Y) = 1 - \frac{MSE(Y)}{Var(Y)} \tag{A.2}$$

The Q-Squared measures the predictive relevance of a model. The higher its value, the more relevant the model is to predict the outcome variable. The difference with $R^2$ is that it is computed using the MSE on the prediction sample, which is not used for the estimation of the model.

**Performance of Alternative ML Methods** Alternative popular ML methods are for example LASSO, Ridge Regressions, Trees, Forests, Gradient Boosting and Neural Networks. Random forests are better suited than LASSO and Ridge Regressions in our case (and indeed offer better predictions) thanks to the non-linear flexibility they offer. Note that we have experimented with LASSO and Ridge Regression models and their predictions systematically under-performed those of Random Forests. Trees and forests are less sophisticated than random forests and more prone to over-fitting as mentioned above, which is also the case of gradient boosting (an extension to random forests). Last, neural networks perform potentially better than random forests, but only in large data samples, which is not the case in our analysis where the number of observations varies between 12,000 and 20,000. See Hastie et al. (2009) for a general introduction to these methods and Varian (2014) or Mullainathan and Spiess (2017) for surveys of applications in economics.

**Sample Selection in Working Conditions Surveys** We restrict the sample to salaried workers and exclude occupations 31 “Independent Workers” 35 ”Artists and Entertainment Occupations”, 44 “Clergy” and 42 “Teachers”. We do so because their jobs do not correspond to typical salaried work and their relation to working from home is different from those of other occupations. Many independent workers, artists, and members of the clergy, would declare working from home all the time because their home and their workplace can be the same environment. On their side, teachers’ occupation requires systematic working from home in the form of preparing lessons, or grading test and exams.

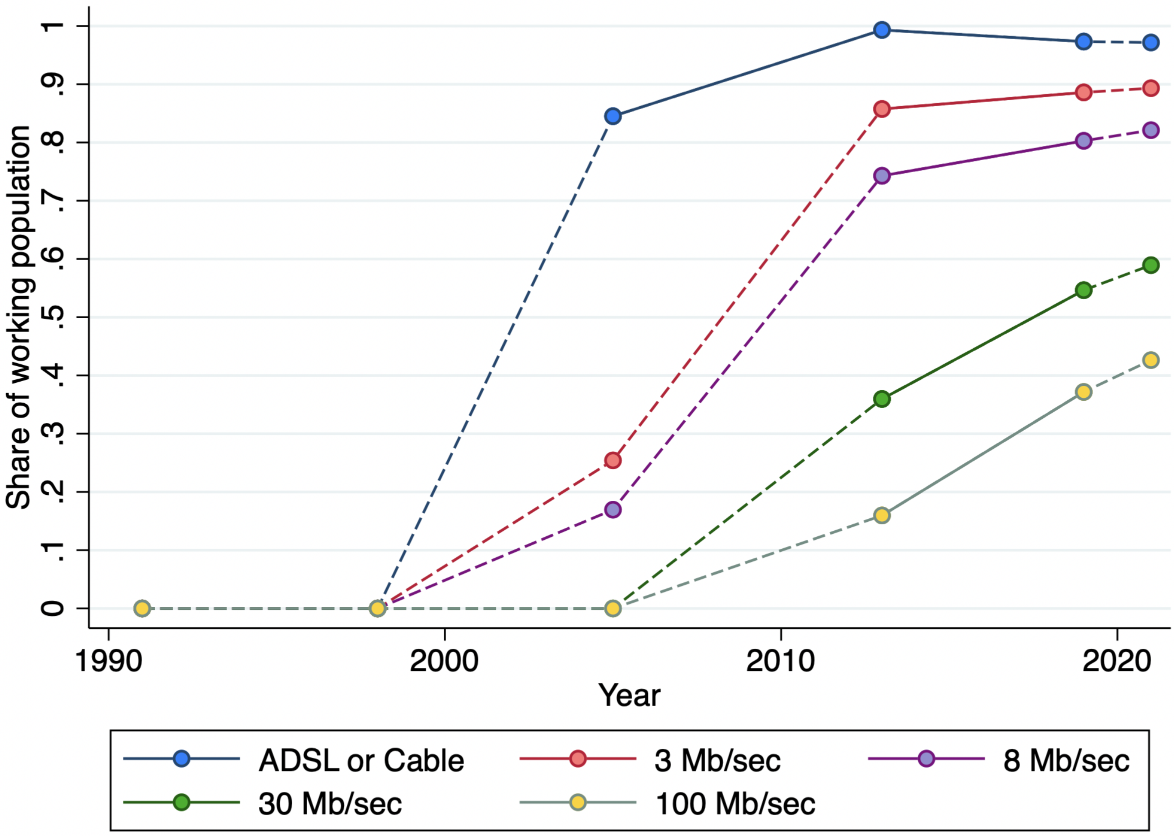
**Sample Selection in Tracov2021** For predictions based on Tracov2021, we have excluded workers who worked less than 15 hours during their reference week (i.e. short part-time work, representing 8.9% of the sample). We do that to construct a measure of the share of hours worked spent using computers that is comparable to that obtained from the other working conditions surveys. Indeed, similarly to other working conditions surveys, Tracov2021 provides the number of hours worked in front of a connected computer during a typical day. However, and in contrast to other working conditions surveys, it does not provide the total number of hours worked in a typical day, but only the number of hours worked per week. In order to avoid over-estimating the share of hours spent in front of a computer, we have therefore removed short part-time work and then dividing the number of hours an individual declares to work on a computer by seven (the legal daily working time in France) to get a reasonable measure of the computer share.[A.1]

Since the other working conditions surveys do report the typical daily hours worked, we have not applied this restriction in order to keep a sample representative of the whole working population. We have however performed robustness analyses with this sample restriction applied to all our yearly samples, and results are almost identical.

[A.1]Over-estimation is less of a concern as we winsorize the share at one.

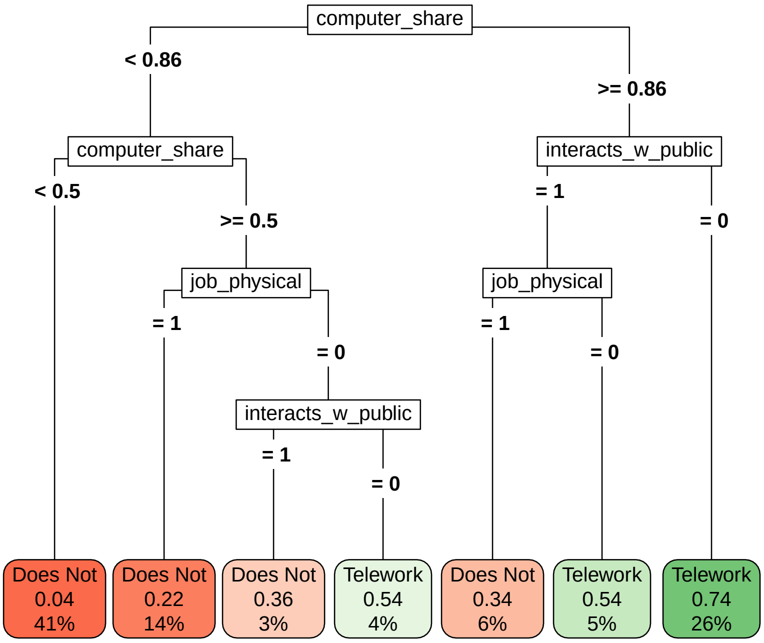
## B Additional Figures and Tables

Figure A.1: Evolution of Internet Coverage



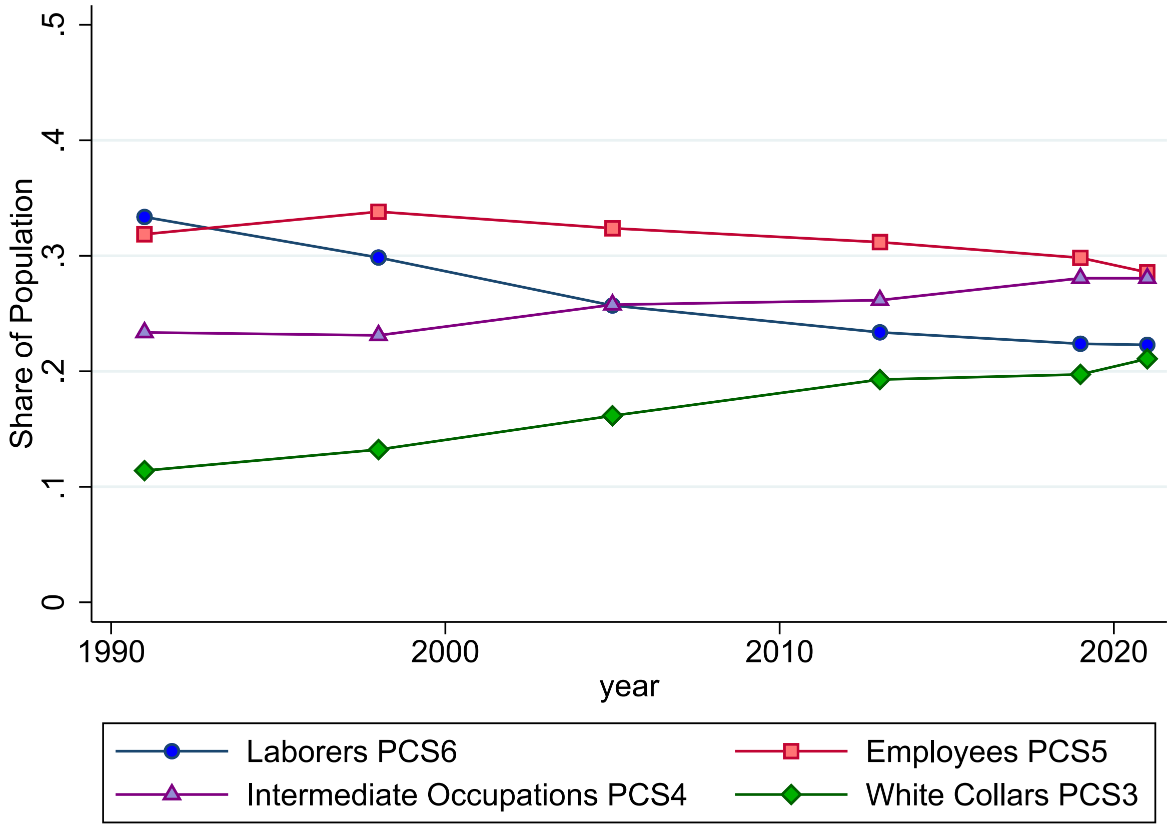

Note: The figure displays the percentage of the French population with access to different Internet connection types over time: ADSL (Asymmetric Digital Subscriber Line, with a speed of 512 kb/sec when it was introduced in 1999) until 2005 and ADSL or cable afterwards, and 4 broadband Internet connection speed levels: 3, 8 , 30, and 100 Mb/sec. Dashed lines are based on linear interpolations for earlier periods (we know that in 1998 none of the connection types were existing, and that in 2005 there was no connection speed equal or larger than 30 Mb/sec) and extrapolations for the period of 2019-2021.

Figure A.2: Example of an Optimized Decision Tree for Teleworking in 2021




Note: The first row of results indicates the prediction regarding teleworking or not, the second row, the average proportion of individuals teleworking in the corresponding class, the third one the percentage of individuals the class represents in the total sample.

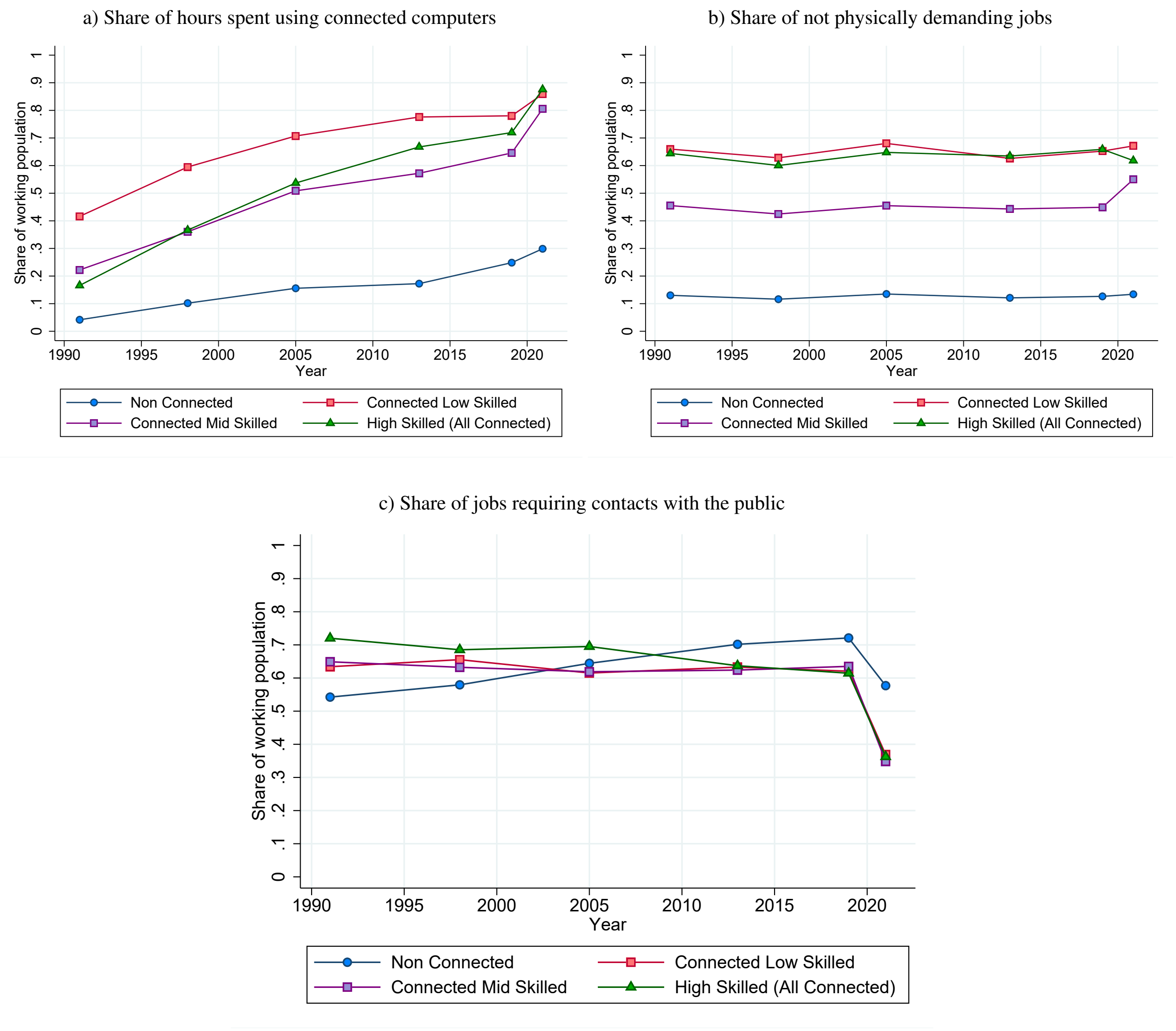
Figure A.3: Evolution of 1-digit occupation shares. 1991-2021




*Source:* .

*Note:*

Figure A.4: Evolution of tasks that make teleworking feasible by occupation




*Source:* Working Conditions surveys 1991 to 2021.

Table A.1: Questions of the Working Conditions survey

| Question | Working conditions survey 1991 | 1998 | 2005 | 2013 | 2019 | Tracov 2021 |
|---|---|---|---|---|---|---|
| **Intensity of ICT use at work** | | | | | | |
| During your work, do you use (even occasionally)<br>a)... a desktop computer? | | | | ✓ | ✓ | |
| During your work, do you use (even occasionally)<br>a)... a microcomputer connected to a network or to other computers? | | ✓ | ✓ | | | |
| During your work, do you use (even occasionally)<br>a)... a standalone microcomputer? | | ✓ | ✓ | | | |
| During your work, do you use (even occasionally)<br>b)... a laptop computer? | | ✓ | ✓ | ✓ | ✓ | |
| During your work, do you use (even occasionally)<br>c)... a tablet, a PDA, a mobile or embedded terminal?<br>*c)... a terminal or computer console* (1998, 2005) | | ✓ | ✓ | | ✓ | |
| Do you have a professional email inbox? | | | ✓ | ✓ | ✓ | |
| Aside from messaging, do you use the Internet for professional purposes? | | ✓ | ✓ | ✓ | ✓ | |
| Do you use an Intranet or internal communication network? | | | ✓ | ✓ | ✓ | |
| When you are not at your workplace, can you access...<br>b)... your organization's or institution's computer system? | | | | ✓ | ✓ | |
| In total, how much time do you spend professionally using<br>the computer equipment we have just discussed? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Intensity of face-to-face interactions with either colleagues or the public** | | | | | | |
| *...interactions with the public* | | | | | | |
| Are you in direct contact with the public?<br>(users, patients, students, travelers, clients, suppliers, etc.) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ... face-to-face? | | | ✓ | ✓ | ✓ | |
| ... by phone? | | | ✓ | ✓ | ✓ | |

| | | | | | | |
|---|---|---|---|---|---|---|
| ... electronically (email, forum, chat, social networks)? | | | ✓ | ✓ | ✓ | |
| Is your work pace dictated by external demands (clients, patients, public) requiring an immediate response? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Do you experience tense situations in your interactions with the public (users, patients, students, travelers, clients, suppliers, etc.)? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| *...interactions with colleagues* | | | | | | |
| Do you work alone? | | | | ✓ | ✓ | ✓ |
| Is your work pace dictated by immediate dependency on the work of one or more colleagues? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Do you experience tense situations in your interactions with your colleagues? 3. (no colleagues) | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Is your work pace dictated by constant monitoring or surveillance (permanent or at least daily) exercised by management? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| **Measures of physical intensity and manual work** | | | | | | |
| Does your work involve ... | | | | | | |
| ... long periods of standing? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... long periods in another uncomfortable posture or tiring in the long run? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... frequent and long walks? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... lifting or moving heavy loads? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... other significant physical efforts? | ✓ | ✓ | | | | |
| ... painful or tiring movements? | | ✓ | ✓ | ✓ | ✓ | |
| ... exposure to shocks or vibrations? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... not being able to take your eyes off your work? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... reading small letters or numbers (or poorly printed, or poorly written) ? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... examining very small objects, fine details? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... paying attention to brief visual or auditory signals, unpredictable or difficult to detect? | ✓ | ✓ | ✓ | ✓ | ✓ | |
| ... wearing work clothes? | ✓ | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| When you work, if a person, standing 2 or 3 meters away from you, speaks to you ... 1. you hear them if they speak normally 2. you hear them if they raise their voice 3. you cannot hear them | ✓ | ✓ | ✓ | ✓ | ✓ |
| Is your work pace dictated by ... a) ... the automatic movement of a product or a part? | ✓ | ✓ | ✓ | ✓ | ✓ |
| Is your work pace dictated by ... b) ... the automatic rhythm of a machine? | ✓ | ✓ | ✓ | ✓ | ✓ |
| Where do you spend most of your working time? 5. Traveling (e.g., flight attendant, sales representative, truck driver, driver...) | | | | ✓ | ✓ |
| 8. On one or more construction sites | | | | ✓ | ✓ |
| Is this mainly because ... ? 3 - you are on a mission, working on a construction site, or intervening at clients' premises | | | ✓ | | |
| Do you use a vehicle for work or for your professional needs, outside of commuting? | | | | ✓ | ✓ |
| **Health and Safety** | | | | | |
| At your workplace, are you required to ... | | | | | |
| a) ... breathe fumes or dust? | ✓ | ✓ | ✓ | ✓ | ✓ |
| b) ... be in contact with hazardous substances? | ✓ | ✓ | ✓ | ✓ | ✓ |
| c) ... be exposed to infectious risks? | ✓ | ✓ | ✓ | ✓ | ✓ |
| d) ... risk being injured or having an accident? | ✓ | ✓ | ✓ | ✓ | ✓ |
| e) ... risk traffic accidents during work? | ✓ | ✓ | ✓ | ✓ | ✓ |
| Does your employer provide you with personal protective equipment such as gloves, goggles, safety shoes, harnesses, ... ? | | | | ✓ | ✓ |
| Do you use them? | | | | ✓ | ✓ |
| **Quality of internet connection at home** | | | | | |
| Residential municipality | | | | | ✓ |

Table A.2: Regressions of Teleworking on ***Predicted*** Feasibility and Occupation Groups in 2021

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Dependent Variable: Teleworking Dummy | | | | | |
| TWK feasibility Prediction | 0.664*** | 0.620*** | 0.592*** | 0.214*** | 0.209*** | 0.281*** |
| | (0.012) | (0.012) | (0.012) | (0.011) | (0.011) | (0.036) |
| Executives | 0.238*** | 0.114*** | 0.098*** | 0.034*** | 0.033*** | 0.052*** |
| | (0.010) | (0.015) | (0.015) | (0.012) | (0.012) | (0.013) |
| Connected mid-skilled | 0.132*** | 0.092*** | 0.089*** | 0.019* | 0.015 | 0.022** |
| | (0.010) | (0.012) | (0.012) | (0.010) | (0.010) | (0.010) |
| Connected low-skilled | 0.009 | 0.044* | 0.036 | -0.030 | -0.033* | -0.021 |
| | (0.014) | (0.023) | (0.023) | (0.019) | (0.020) | (0.020) |
| Pred. TWK feas. X Executives | | | 0.487*** | 0.324*** | 0.311*** | 0.248*** |
| | | | (0.032) | (0.027) | (0.027) | (0.029) |
| Pred. TWK feas. X Connected mid-skill | | | 0.324*** | 0.204*** | 0.201*** | 0.160*** |
| | | | (0.028) | (0.023) | (0.023) | (0.024) |
| Pred. TWK feas. X Connected low-skill | | | 0.176*** | 0.165*** | 0.166*** | 0.113*** |
| | | | (0.049) | (0.040) | (0.041) | (0.042) |
| Wants to TWK | | | | 0.550*** | 0.552*** | 0.550*** |
| | | | | (0.008) | (0.008) | (0.008) |
| No hierarchy | | | | | 0.003 | 0.001 |
| | | | | | (0.011) | (0.011) |
| Degree of Hierarchy Control | | | | | -0.005 | -0.006 |
| | | | | | (0.004) | (0.004) |
| Autonomy | | | | | 0.024*** | 0.024*** |
| | | | | | (0.003) | (0.003) |
| Computer Share | | | | | | -0.060** |
| | | | | | | (0.024) |
| Contact with Public | | | | | | -0.026*** |
| | | | | | | (0.006) |
| Work Always Physical | | | | | | 0.033** |
| | | | | | | (0.014) |
| Work Often Physical | | | | | | 0.003 |
| | | | | | | (0.013) |
| Work Never Physical | | | | | | -0.003 |
| | | | | | | (0.010) |
| Observations | 11,454 | 11,454 | 11,454 | 11,434 | 11,389 | 11,385 |
| R-squared | 0.458 | 0.473 | 0.494 | 0.654 | 0.657 | 0.658 |
| Covid-related Controls | No | No | No | Yes | Yes | Yes |
| Demographic Controls | No | No | Yes | Yes | Yes | Yes |

Note: Covid-related Controls capture the fear of catching Covid-19 while commuting, being particularly vulnerable to Covid-19, living with people particularly vulnerable to Covid-19, and working in an environment when catching Covid-19 is likely. Demographic controls include gender, age, seniority, continent of Birth, region of living. Firm size controls are a second order polynomial in the number of employees. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.3: Hierarchical Control and Autonomy by Groups of Employees

| | Unconnected Employees | Connected Low-Skill | Connected Mid-Skill | Executives |
|---|---|---|---|---|
| *Standardized variables:* | | | | |
| Degree of Control by Hierarchy | .01 | .047 | .08 | -.084 |
| Autonomy | -.146 | .044 | .152 | .14 |
| *Indicator variables:* | | | | |
| No Hierarchy | .138 | .126 | .083 | .09 |
| Supervision is Main Task | .063 | .036 | .194 | .244 |
| Supervision is Secondary Task | .237 | .199 | .316 | .381 |
| No Supervision | .7 | .765 | .491 | .374 |
| Takes Initiatives | .591 | .496 | .751 | .87 |
| Needs Help of Supervisor | .238 | .222 | .23 | .251 |

Note: The table displays weighted averages by occupation groups. The variable "Degree of Control by Hierarchy" is a standardized measure of the response to the question "at what frequency are you controlled or monitored by your supervisors?" (5 modalities). The variable Autonomy is a standardized measure of the degree to which employees agree with the statements "Can organize myself" (4 modalities).

Table A.4: Teleworking gaps across occupations for different tenure groups

| | (1) Tenure $\leq$ 1 Year | (2) 1 Year $<$ Tenure $\leq$ 5 Years | (3) Tenure $>$ 5 Years |
|---|---|---|---|
| TWK Feasible | 0.505*** | 0.617*** | 0.633*** |
| | (0.026) | (0.015) | (0.010) |
| Executives | 0.203*** | 0.183*** | 0.084*** |
| | (0.032) | (0.017) | (0.012) |
| Connected mid-skilled | 0.063** | 0.083*** | 0.063*** |
| | (0.029) | (0.017) | (0.011) |
| Connected low-skilled | -0.027 | 0.049** | 0.004 |
| | (0.037) | (0.022) | (0.016) |
| Observations | 1,240 | 3,265 | 6,917 |
| R-squared | 0.582 | 0.623 | 0.590 |

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Teleworking gaps across occupations depending on supervision and autonomy

| | Supervisory function | | | Has to take initatives | | Help from supervisor | |
|---|---|---|---|---|---|---|---|
| | Supervize is Main Task | Supervize is Secondary Task | Does Not Supervize | Often or Always | Sometimes or Never | Does not Need Help of Supervizer | Needs Help of Supervizer |
| TWK Feasible | 0.586*** | 0.595*** | 0.627*** | 0.611*** | 0.617*** | 0.565*** | 0.638*** |
| | (0.022) | (0.016) | (0.010) | (0.010) | (0.013) | (0.018) | (0.009) |
| Executives | 0.114*** | 0.147*** | 0.153*** | 0.122*** | 0.152*** | 0.111*** | 0.134*** |
| | (0.028) | (0.018) | (0.013) | (0.011) | (0.019) | (0.022) | (0.012) |
| Connected mid-skilled | 0.089*** | 0.077*** | 0.073*** | 0.072*** | 0.052*** | 0.085*** | 0.061*** |
| | (0.027) | (0.017) | (0.011) | (0.011) | (0.015) | (0.021) | (0.011) |
| Connected low-skilled | 0.158** | -0.002 | -0.000 | 0.039** | -0.024 | 0.067** | -0.004 |
| | (0.069) | (0.028) | (0.013) | (0.017) | (0.016) | (0.030) | (0.015) |
| Observations | 1,450 | 3,182 | 6,799 | 7,707 | 3,725 | 2,314 | 7,370 |
| R-squared | 0.517 | 0.592 | 0.623 | 0.578 | 0.609 | 0.548 | 0.606 |

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

# C Additional results related to regression analyses

**Differences Across Occupations and Role of Hierarchical Control and Autonomy in 2021: Detailed Results** Table 3, column (1), confirms that there is a strong link between teleworking feasibility and actual teleworking: jobs that can be teleworked are 63.3 % more likely to be indeed teleworked. It also confirms that there are large gaps across occupations in teleworking conditional on its feasibility: executives and mid-skilled connected employees are respectively 15 and 8 percentage points more likely to telework than are low-skilled connected and no-connected employees for jobs that are equally teleworkable. Having a job feasible from home also increases actual teleworking more among higher-skilled occupations than among lower-skilled ones and among non-connected employees (see the interaction terms in column 2). The gaps between skill groups in the relationship between teleworking feasibility and actual teleworking are also always large and statistically significant. The differences (both in the average recourse to teleworking and the link between feasibility and teleworking) across broad occupational groups are only slightly reduced when detailed controls for workers' demographics are added (column 3). Controlling further for workers' desire to work from home—as well as detailed Covid-related controls such as fear of the Covid-19 or being particularly at risk—leaves the gaps across occupations largely unaffected (column 4). In column (5), we include individual-level measures of the degree of hierarchical control (a dummy for having at least some control and a categorical variable from very light to very intensive control) and autonomy at work (also categorical). Both categorical variables have been standardized to have a mean of 0 and a standard deviation of 1. Autonomy is positively associated with a higher likelihood of having at least half a day of teleworking. The relationship between hierarchical control and teleworking goes in the opposite direction, as expected. However, these associations remain quantitatively small (a one standard deviation variation in any of these variables is associated with variations in teleworking lower than 3 percentage points). Hence, these factors have little incidence on the gaps in teleworking across occupation during the Covid-19 crisis. In column (6), we control directly for the job tasks most strongly associated with teleworking on top of teleworking feasibility in order to control even more finely for the role of job tasks. Gaps between skill groups become a bit smaller but largely remain. Together, these results show that there remains large gaps across occupations in the recourse to teleworking during the Covid-19 crisis, and that these gaps cannot be easily explained by teleworking feasibility, workers' desire to telework, or by managerial practices.

**Robustness Checks for the Results Presented for 2019** To study if these conclusions can be extended to the pre-Covid period, we estimate similar regression models using the 2019 working conditions survey. Teleworking feasibility is however not observed directly, and we therefore replace it by its prediction, so that differences across occupations are conditional on predicted feasibility. To suggest that this change has no incidence on the conclusions, we have replicated the analysis in 2021 after also replacing teleworking feasibility by its prediction (see Table A.2 to be compared to Table 3). The gaps across occupations conditional on predicted feasibility are larger than those conditional on actual feasibility, but the influence of demand-side factors on these gaps remains close to 0. Similarly, the gaps across occupations in the link between predicted feasibility and teleworking are also larger and unaffected by demand-side controls. Not surprising, the link of teleworking with predicted feasibility is also lower than its link with actual feasibility. However, the differences in this link across occupations are comparable in both specifications. The only variable that has a significant influence on the gaps across occupations in the replication exercise is workers' desire to telework. This is partly mechanical, as this variable is mechanically correlated with actual feasibility (it is observed only for jobs that are feasible from home and we impute it to be 0 for jobs that cannot be done from home). Hence, its estimated effect in regressions that do not also control for actual feasibility cannot be easily interpreted.

The results for 2019 are then presented in Table 4 and largely described in the main text. A potential concern with the results in this Table is that our ability to predict teleworking precisely for a given worker might be poor.[A.2] To alleviate this concern, we directly add as controls in column (6) the 67 task variables used to predict teleworking in 2019. When we do this, differences across occupations are reduced but only slightly. They remain quantitatively very large given the average level of teleworking in 2019. The relationships between the presence of a hierarchy or deadlines and teleworking also remain large and largely unaffected.

[A.2] Note that this criticism does not apply to our historical time series in which we were considering averages in predicted feasibility over large groups of workers.