# Does Working from Home Boost Productivity?

by

SANJANA KHONDAKER
22375018

**THESIS**

Submitted to the department of Economics and Social Science in partial fulfillment of the Requirements for the degree of Master of Science in Applied Economics.

Department of Economics and Social Science
Brac University
November 2024

## Approval

The thesis titled "Does Working from Home Boost Productivity?" submitted by Sanjana Khondaker (22375018) has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Master of Science in Applied Economics on November, 2024.

**Examining Committee:**

Supervisor:
(Member)

Dr. Muhammad Shafiullah
Associate Professor
Department of Economics and Social Sciences
Brac University

Program Coordinator:
(Member)

Dr. Wasiqur Rahman Khan
Professor & Chairperson
Department of Economics and Social Sciences
Brac University

External Expert Examiner:

(Member)

22/10/24

Dr. Shuddha Rafiq
Senior Lecturer
Department of Economics
Faculty of Business and Law
Deakin Business School

Departmental Head:

(Chair)

Dr. Wasiqur Rahman Khan
Professor & Chairperson
Department of Economics and Social Sciences
Brac University

# Declaration

It is hereby declared that

1. This thesis represents my original work conducted while pursuing my degree at Brac University.
2. The thesis does not include materials previously published or authored by others. All external sources have been properly cited with complete and accurate references.
3. The thesis does not include material that has been accepted or submitted for any other degree or diploma at any university or institution.
4. I have duly acknowledged all significant sources of help.

**Student's Full Name & Signature**

Sanjana Khondaker

# Acknowledgements


I would like to express my sincere gratitude to everyone who has helped me to complete this thesis.

First and foremost, I would like to thank my supervisor, Dr. Muhammad Shafiullah, for his constant guidance, support, and invaluable feedback throughout the process. His insights and expertise have been instrumental in completion of this work.

I am also grateful to Brac University for providing access to all the necessary resources and facilities which helped in the execution of this project.

My heartfelt appreciation goes to my family, especially my spouse and mother, for their constant support, care and understanding, which has been the source of my strength and determination during this academic endeavor.

Thank you to everyone who has played a part in shaping and enriching this work.

Sanjana Khondaker
November 2024

# ABSTRACT

This study investigates the impact of remote work on employee productivity, exploring the roles of autonomy, training, and individual characteristics in dictating employee performance. Using cross-sectional data from the Chinese General Social Survey (CGSS), I employ various methodological approaches to address selectivity bias and endogeneity concerns inherent in the analysis. Initially, ordinary least squares (OLS) regression is utilized to estimate associations between remote work, autonomy, training, and productivity. Subsequently, a multinomial logit model is employed to identify which individuals are more likely to do remote work. To mitigate potential bias, propensity score matching estimator is used, considering observed heterogeneity between remote and non-remote workers. Furthermore, instrumental variable (IV)-free Gaussian Copula and Lewbel methods are also utilized to assess the causal effect of remote work on productivity. The findings reveal that remote work positively impacts employee performance, however, long hours or frequent remote work negatively affect productivity. Notably, a significant causal relationship is identified. Additionally, greater autonomy and training are associated with enhanced productivity. Moreover, demographic factors such as gender, marital status, and parental responsibilities influence remote work tendencies, with educated individuals exhibiting lower likelihoods of remote work, while those receiving training and receiving greater independence tend to opt for remote work arrangements.

## Table of Content:

# 1. Introduction

The terms work-from-home (WFH), teleworking, or remote working, tends to hold the same meaning, an accommodation in which employees have the freedom of working in a non-official environment. Although work from home has been around for some time, its instances have definitely increased with an increase in technological advancement. The prevalence of remote work has increased dramatically, a recent survey published in Statista, stated that the share of employees working remotely worldwide has increased from 20 percent in 2020 to 28 percent in 2023 (Sherif, 2024). The push from the lockdown of COVID-19 pandemic demonstrated that the majority of the office work for several occupations and industries can be carried out from our own homes, although this comes with much debate. Although workers have reported to feel lonely in a work-from-home setup it still remains a preferred work structure, as 91 percent of the employees globally prefer either a full or a hybrid work structure. Its advantages are notably more significant for women, who typically shoulder more child-rearing and household related responsibilities (Scandura and Lankau, 1997), therefore, these points highlight the importance of studying this phenomenon.

While an increase in remote work has the potential to improve several economic and social indicators such as worker well-being, gender and regional equality. Its overall impact on labour productivity is still ambiguous. Previous research papers such as Ferrara(2022), Mori(2021), Kathryn (2010) and Siha (2006) have focused primarily on qualitative analysis. Although several empirical studies have been conducted revealing that WFH is associated with increasing productivity, it is not very clear whether this is mainly due to occasional work from home or frequent work from home and whether there exists a causal relation.

Given work from home is here to stay, and will continue to increase in the long run, it is important to understand how it will affect various aspects of work-life, performance, and inequality. This study takes into account that selection into working from home might be driven by observed and unobserved characteristics and that it may be an endogenous variable. Endogeneity can occur due to sample selection, measurement error, omitted variables, or simultaneous causality (Hill, Johnson, Greco, O'Boyle, & Walter, 2021).

As of December 2023, approximately 537 million employees in China work from home, compared to only 199 million in 2020 (Thomala, 2024). This represents a nearly 170% increase. Remote work in China has evolved considerably due to technological progress and global trends. The country's strong digital communication infrastructure and flourishing tech industry have enabled many companies to efficiently handle remote work. The government has promoted policies supporting flexible work environments, particularly in major urban areas such as Beijing and Shanghai. Consequently, it is vital to understand the potential impacts of working from home. Assessing how WFH influences performance and well-being is of great significance. Using cross-sectional data from the Chinese General Social Survey (CGSS), this study finds that work-from-home is an effective factor to increase employee performance, however, long hours of remote work or frequently working from home has a negative effect on productivity.

The empirical analysis developed in this paper has important findings as it shows that work from home has a positive impact on labor productivity. The objective and contribution of this study is to add to the existing literature in the following way: In contrast to the vast majority of existing studies, this study focuses on several different models and causal quantitative methods to rigorously scrutinize the underlying relationship between these factors. The major contribution of this study, therefore, is to examine the role of remote work, and shedding light on its impact on individual performance.

Firstly, the study uses the regular OLS procedure to estimate the association between working from home, autonomy, training and productivity. Next, the study specifies a multinomial logit model to evaluate which individuals are more prone to working from home. In order to account for a possible selectivity bias arising from the fact that OLS ignores the endogeneity of selection into remote work, the second approach is to estimate average treatment effect on the treated (ATT), applying matching methods based on propensity scores. However, as this strategy might not eliminate a bias resulting from unobserved heterogeneity, the study additionally estimated the effect of remote work on productivity using instrumental variable (IV)-free Gaussian copula approach and Lewbel (2012) method.

The remainder of the thesis is arranged as follows: Chapter 2 conducts a thorough literature review of past papers and builds the hypothesis of this study. Chapter 3 discusses the data and variables;

chapter 4 presents the empirical models. A thorough overview of the regression results is present in Chapter 5, Chapter 6 provides the summary of the findings and the economic discussion. Finally, Chapter 6 concludes.

# 2. Literature Review & Hypothesis Formulation

The landscape of work has undergone significant changes, due to technological advancement and improvements in the organizational structure. The contemporary labor market has witnessed a notable increase in the adoption of remote work setups. This shift has not only redefined traditional notions of work but has also prompted a surge in scholarly interest in understanding the implications of remote work. Despite the growing body of literature, there exists a gap that requires further empirical research in work from home. This literature review aims to critically assess the vast number of existing literatures, and bridge this gap.

One recent study was conducted by Kazekami, (2020), which investigated the factors affecting labor productivity in Japan, for this analysis both difference in difference method, as well as, instrumental variables were employed. First, this study finds that telework increases labor productivity, however, this holds true only when it is done in reasonable working hours. Therefore, according to this study when telework hours are too long, telework decreases labor productivity. In contrast, teleworkers feel more stress balancing work and domestic chores, however, these emotions do not diminish the productivity of the individuals. It can be argued that low intensity teleworkers and high intensity teleworkers experience different telework outcomes. Similarly, Bentley et al. (2016), explains how low intensity teleworkers face different organizational support needs from high intensity teleworkers, and have different outcomes from remote working. As low intensity teleworkers only work remotely for a few days in a week, and thus need less technological support and suffer from less social isolation. In his paper he cites Neufeld and Fang, (2005), as well as, Belangeretal (2012), who found that the outcomes of telework is a time dependent concept, and can lead to disadvantages when workers carry out remote working for more than two or three days in a week (Gajendran and Harrison, 2007).

Therefore, as organizations increasingly adopt and adapt to remote work practices, the concept of working from home has evolved beyond being merely a workplace flexibility option. Its effects on productivity, job satisfaction, and the broader socioeconomic impact on individuals and society holds significant implications. Arguably, individuals working from home experience less interference, are not exhausted from hours of commuting, are more satisfied and thus can demonstrate a higher level of productivity. Yang and Zheng (2011) use a logistic regression model to study how flexible work programs affect workers' actualization of productivity. The findings from this study revealed that the highest level of productivity actualization is associated with individuals who enjoy a realistic flexible work schedule. Therefore, it is well understood that a flexible work schedule is an important determinant of productivity. Indeed, Fonner and Roloff, (2010), concluded that high intensity teleworkers tend to be more satisfied than traditional in-office employees as a result of less work-life conflict. Thus, due to their work arrangement and setup, remote workers are able to achieve significant benefits.

The argument that telework is a positive factor in increasing productivity appears more convincing when there is evidence that teleworkers enjoy relatively less distractions (Mann et al., 2000), feels reduced work pressure and higher autonomy, experiences work-life balance by having increased opportunity for leisure as they can have greater control over work schedules (Mann and Holdsworth, 2003) and experiences less stress caused by crowded trains and buses during rush hour (Kazekami, 2019). In support of  the contention that working from home increases productivity, a field experiment was conducted in a chinese travel agency, the employees were randomly assigned to work from home or in the office for a period of time, and results depicted that remote working led to an increase in performance, higher productivity, fewer sick-days and higher work satisfaction (Bloom et al., 2015).

On the other hand, an empirical study performed by DuBrin(1991) revealed that tasks which are repetitive, structured and measurable can be done easily from home, and such employees showed an increase in productivity working from home. Furthermore, Alipour et al. (2021)'s paper, utilizing instrumental variables, reported the causal effect of working from home (WFH) on work relations and public health. Interestingly, it was observed that during COVID-19 fewer individuals had filed applications for shorter working hours as a result of being able to work from home. In other words, during the pandemic reliance on telework reduced the probability of employees filing for short-time work applications. In the analysis above I have reviewed evidence for the beneficial role of telework in increasing individual productivity, resulting in the following prediction:

H1: Working from home increases an individual's productivity and shares a causal relationship, however, conducting telework for long hours has a negative effect on productivity.

Secondly, autonomy has been identified as a positive factor affecting productivity. Autonomy can be referred to as the degree of flexibility and independence one receives in carrying out their task and the methods and decisions implemented to complete a task or achieve a goal (Hackman and Oldham, 1975). Work flexibility or autonomy has been an important factor, demonstrating significant effects in deciding employee productivity (MacEachen et al., 2007). Looking at the evidence for why autonomy might result in higher productivity, a number of studies have found that providing employees with more independence allows them to take ownership of their task, innovate, manage and plan their work in a way that suits their skills and expertise. Early research related in this topic revealed that autonomy was associated with higher job satisfaction, and lower turnover intentions (Spector, 1986). This is because autonomy provides a sense of supervisory appraisal and reflects professional status (Schwalbe, 1985). So, it motivated individuals to perform better and retain this status.

Similarly, another study examined the relationship between autonomy and support climate on pay level satisfaction (PLS), and the findings indicated that a high autonomous climate decreases the negative effects of low PLS (Schreurs et al., 2015). Demonstrating that autonomy not only allows individuals to feel more motivated which increases productivity but also allows individuals to be more satisfied with their earnings. The assertion that autonomy is positively associated with productivity, gains greater support when considered within the framework of Self-Determination Theory (SDT) perspective. The theory proposes that all human beings have three basic psychological needs - competence, autonomy and relatedness. Moreover, if an individual is satisfied through these basic needs, it will promote optimal motivational traits which increases effective engagement (Ryan and Deci, 2000). Thus, it is reasonable to expect that autonomous climate will have a positive influence in productivity as individuals who enjoy higher autonomy are more motivated. I, therefore predict the next hypothesis in this study:

H2: Employees who receive greater autonomy in their workplace demonstrate higher productivity.

Finally, this study also analyzes the empirical effect of job training on productivity. Earlier studies and contemporary research alike have dedicated substantial attention on the effects on job training and employee productivity. Early investigations attempted to find a foundational link between job training programs and enhanced job performance, which allowed us to understand the effects of skill development on individual and organizational effectiveness. In fact, firms with high levels of investment in employee training have observed an increase in productivity and job performance as opposed to firms who invested less in training (MacDuffie & Kochan, 1995). Although, there exists causal considerable debate on the relationship between these two factors, no studies have explored this association in a work-from-home framework. Thus, this study will analyze how job training affects productivity of individuals who conduct remote work as well as traditional office work.

Additionally, several research utilizing micro data has found that off-the-job training has a more significant effect on productivity than on-the-job training. Schonewille (2001), uses UK labour force survey data to measure the impact of training on improving labour productivity, this study uses on-the-job training and off-the job training as independent factors affecting productivity. Although the results depicted a positive and a statistically significant association between training and labour productivity, it was inconclusive whether it is mainly due to on-the-job or off-the-job training. Therefore, this finding is in line with the hypothesis that job training and productivity have a positive association. In support of this assertion, research has found job training to have an important role in productivity. Research conducted in Ireland, suggests that although general training has a statistically positive effect on productivity growth, no such effect is observable for specific training (Barrett & O'Connell, 1999). So far empirical research shows that training is an important factor determining productivity and job performance, however it does not indicate that only on-the-job training is beneficial to increase productivity. Thus, this study bridges the gap by analyzing the empirical effect of training on productivity. The final hypothesis of this study can be summarized as:

H3: Individuals with training demonstrate higher productivity, as opposed to employees who received no training.

The objective of this study is to add to the existing literature in the following way. In contrast to the vast majority of existing studies, this paper focuses on investigating the causal effect of work from home and productivity. In addition to that the study also determines which individuals are more likely to opt for remote work.

# 3. Data

### 3.1 Data Source and Sample Size

The study will employ secondary data as its framework. The micro-data has been sourced from the Chinese General Social Survey (CGSS) (variable definition, appendix Table A.2). Which is a survey run by the academic institutions: Renmin University of China and the Hong Kong University of Science and Technology. The primary sample size of the research was 10,968 Chinese citizens. After removing the missing values and focusing on only the working age population, the sample size stood at 4,434 individuals. For this analysis only the 2015 survey wave was used, and other waves could not be used due to the unavailability of the primary independent variable - *work from home.*

### 3.2 Descriptive Statistics

Several demographic variables were considered in this study given their association with the primary independent variable and the outcome variable, productivity. Such as age, which is related to career and family demands, as well as organization position and income earned (Raghuram & Weisenfeld, 2004). Bailey & Kurland (2002), explains how the time spent in particular roles or job experience of an individual also plays a role in determining work arrangements. Therefore, this variable also plays a key factor in determining the income of an individual. This study emphasized both regular and remote workers, approximately 4400 employees aged 64 years and younger were considered.

Table 1, below, shows the descriptive statistics, the mean and standard deviation of the continuous variables. There were almost an equal proportion of male and female respondents in our sample. In addition, 91% of the employees were regular office-based employees, and the remaining 9% worked from home. On the other hand, the majority of individuals in the sample (52%) held a high school degree, with only 23% attaining college education and a mere 4% undergoing training to enhance their job skills. Moreover, approximately 18% of the individuals reported that they have complete independence and autonomy to make their own decisions in their workplace, and 13% were members of the union. For this analysis, the dataset was restricted to the working-age population, specifically individuals aged 15 to 64. The mean sample age was 42 years old, with an average of 11 years of working experience, and 83% of respondents were married.

Stratifying the data sample between female workers and male workers, a noteworthy finding emerged: there was a statistically significant difference in the average productivity levels between the two groups. Male workers, earning slightly higher incomes and having relatively more experience and but less training, demonstrated greater productivity compared to female workers. The results of these demographic differences are provided in the Appendix, Table A.1.

**Table 1**

*Summary Statistics*

| Discrete Variables | Frequency | Percent |
|---|---|---|
| Does not WFH | 4028 | 90.8 |
| Works from home | 406 | 9.2 |
| Received college education | 990 | 22.3 |
| Receives low/no autonomy | 3647 | 82.3 |
| Receives full autonomy | 787 | 17.7 |
| Received training | 180 | 4.1 |
| No children | 2531 | 57.1 |
| Respondent location - City | 2679 | 60.4 |
| Respondent location - Rural | 1755 | 39.6 |
| Union member | 593 | 13.4 |
| WFH - Always | 36 | 0.8 |
| WFH - Often | 63 | 1.4 |
| WFH - Sometimes | 107 | 2.4 |
| WFH - Rarely | 200 | 4.5 |

| Discrete Variables | Frequency | Percent | | |
|---|---|---|---|---|
| Male | 2497 | 56.3 | | |
| Female | 1937 | 43.7 | | |
| Married individuals | 3671 | 82.8 | | |
| Continuous Variables | Mean | Minimum | Maximum | SD |
| Income | 41136 | 400 | 5000000 | 1.00E+05 |
| Weekly work hours | 49 | 1 | 219 | 2.00E+01 |
| Productivity | 23 | 0.116 | 2170 | 7.00E+01 |
| Age | 42 | 18 | 64 | 1.00E+01 |

# 4. Model and Methodology

The estimation strategy proceeds in three steps. First, the study uses the regular OLS procedure to estimate the association between working from home, autonomy, training and productivity. Next, the study specifies a multinomial logit model to evaluate which individuals are more prone to work from home. In order to account for a possible selection bias arising from the fact that OLS ignores the endogeneity of selection into remote work, the second approach is to estimate an average treatment effect on the treated applying matching methods based on propensity scores. However, as this strategy might not eliminate a bias resulting from unobserved heterogeneity, the study thirdly estimated the effect of remote work on productivity using instrumental variable (IV)-free Gaussian copula approach and Lewbel method.

## 4.1 Baseline (OLS) Estimation

The econometric model in this study has the general form

$$\ln Productivity_i = \alpha_0 + \gamma WFH_i + \alpha_2 AUT_i + + \alpha_3 TRN_i + \alpha_4 UNN_i + \alpha_5 MAR_i + \alpha_6 EDU_{i+} \alpha_7 GEN_i + \alpha_8 AGE_i + \alpha_9 AGE^2{}_i + \alpha_{10} CHD_i + \alpha_9 REG_i + \epsilon_i \quad (1)$$

Where $\ln Productivity_i$ is the natural logarithm of productivity, $WFH_i$ represents the individuals who work from home, frequently, occasionally and rarely, $AUT_i$ is a dummy variable of autonomy/complete independence, $TRN_i$ is the dummy variable of training, which separates individuals who have received training to the individuals who have received no training. $UNN_i$ represents the individuals who are members of the union, the remaining control variables are *MAR* representing marital status, *EDU* represent years of education, *GEN* is gender, *AGE*, $AGE^2$ is age squared, *CHD* is a dummy representing whether they have a child, and *REG* which represents whether the individual resides in a rural or urban geographic area. Finally, $\epsilon_i$ is a stochastic error term with the assumptions of normal distribution, zero mean, and finite variance.

In this model, the dependent variable is defined as the natural log of productivity. The productivity variable was created by using two variables from the survey, which are the weekly working hours and labor income. The number of working hours per week was then multiplied by 48 (4 weeks per month multiplied by 12 months in a year), to get the total number of working hours in a year, assuming a consistent work schedule throughout the year. The annual labor income was then divided by the annual working hour to get the income earned per hour of work (Kazekami, 2020). Which is a measure of earnings efficiency and productivity analysis. The use of the natural log of annual productivity in the models allows to interpret the regression coefficients as a percentage change in productivity for a one-unit change in the independent variable. Table A.2 In the appendix, continues the definition and setup of all the other variables chosen in this study.
Since the available data allows to distinguish between the different frequency of work-home-from, a modified version of Eq. 1 is also used in this analysis:

$$\ln Productivity_i = \alpha_0 + \gamma_2 WFH.FREQ_i + \gamma_3 WFH.OCC_i + \gamma_4 WFH.RARE_i + \alpha_2 AUT_i + \alpha_3 TRN_i + \alpha_4 UNN_i + \alpha_5 MAR_i + \alpha_6 EDU_{i+} \alpha_7 GEN_i + \alpha_8 AGE_i + \alpha_9 AGE^2{}_i + \alpha_{10} CHD_i + \alpha_9 REG_i + \nu_i \quad (2)$$

Where $WFH.FREQ_i$ ,$WFH.OCC_i$,$WFH.RARE_i$ are all dummy variables, accounting for different levels of working from home. Three measures of work from home have been computed following Alipour's (2021) model. First the employees who “always” work from home, defined as WFH.freq, second the share of employees who often and sometimes were grouped into the work from home occasional dummy variable (WFH.occ). Third, the share of employees working at home rarely or have the option of working from home were set as WFH.rare. In the absence of selection problems with respect to remote working, the work-from-home coefficients $\gamma$, $\gamma_2$, $\gamma_3$ and $\gamma_4$, in model (1) and (2) can consistently be estimated using the conventional OLS procedure.

In order to see how productivity varies between individuals who received training or has autonomy and also conducts remote work interaction terms are introduced into the model, shown in equation 3 below:

$$\begin{aligned}\ln Productivity_i = {} & \alpha_0 + \gamma_2 WFH.FREQ_i + \gamma_3 WFH.OCC_i + \gamma_4 WFH.RARE_i + \alpha_1 WFH_{Freq} * AUT_i \\ & + \alpha_2 WFH_{OCC} * AUT_i + \alpha_2 AUT_i + \alpha_4 WFH_{Freq} * TRN_i \\ & + \alpha_5 WFH_{OCC} * TRN_i + \alpha_6 TRN_i + \alpha_7 UNN_i + \alpha_8 MAR_i + \alpha_9 EDU_{i+} \alpha_{10} GEN_i + \alpha_{11} AGE_i \\ & + \alpha_{12} AGE^2{}_i + \alpha_{13} CHD_i + \alpha_{14} REG_i + \nu_i\end{aligned} \tag{3}$$

Where $WFH_{Freq} * AUT_i +\ , WFH_{OCC} * AUT_i, WFH_{Freq} * TRN, WFH_{OCC} * TRN_i$ are all interaction terms.

The dataset also allows to segregate between blue-collar and white-collar workers, which will allow us to see the differences in productivity among varying occupation. Equation 4, includes the variables $BLUE_i, WHITE_i\ and\ WHITE * WFH_i$, to capture this effect. Here, BLUE and WHITE represents the dummy variables of blue-collar worker and white-collar workers, respectively. In order to proxy for blue collar workers, I created a dummy for all those who have lower education levels (no education, primary/junior high school) to be 1, 0 otherwise. Similarly, in order to account for white-collar workers, individuals who have completed general high school and above were categorized as 1. Finally, the interaction term $WHITE * WFH_i$ will reveal the productivity differences of white-collar workers who work from home.

$$\begin{aligned}\ln Productivity_i = {} & \alpha_0 + \gamma WFH_i + \alpha_2 AUT_i + + \alpha_3 TRN_i + \alpha_4 UNN_i + \alpha_5 MAR_i + \alpha_6 EDU_{i+} \alpha_7 GEN_i \\ & + \alpha_8 AGE_i + \alpha_9 AGE^2{}_i + \alpha_{10} CHD_i + \alpha_{11} REG_i + \alpha_{12} BLUE_i + \alpha_{13} WHITE_i \\ & + \alpha_{14} WHITE * WFH_i + \epsilon_i\end{aligned} \tag{4}$$

## 4.2 Propensity Score Analysis

The decision process of whether an individual works from home can be argued to be endogenous, as it can be driven by individual and firm characteristics. Failure to consider the selection process of both employees and employers often leads to biased and inconsistent parameter estimates. Propensity score analysis offers a range of methods to mitigate the impact of confounding variables and estimate treatment effects. The PSM approach estimates the Average treatment effect (ATE) and the Average treatment effects on the treated (ATT).

Propensity score analysis attempts to balance a set of variables across treatment and control groups by matching similar participants. These variables that are balanced are referred to as covariates. The purpose of this step is to create exchangeable groups, which allows us find the impact of non-randomized interventions (Godard-Sebillotte et al., 2019), the intervention or treatment in this case is remote working. For this model, the variables used to balance were the same control variables that were used in the baseline OLS regression, the outcome measured is productivity, the grouping variable (treatment group) here will be remote employees versus office/on-site employees. Therefore, the study matches the variables starting again with the classical OLS explanatory and control variables.

For this analysis the propensity scores represent the probability of participants working from home, determined by their baseline characteristics. The model is a logistic regression which models the likelihood or probability of belonging to the intervention group versus the control group. Therefore, to estimate the propensity score a binary choice model for work from home is estimated using the logit maximum likelihood method. Predictors or covariates included in the propensity score model (logistic regression) will be the same variable as the baseline OLS regression.

The logit model estimates the probability of working from home relative to working in an office. The estimated probabilities or propensity scores, are then used to match individuals in the analytic sample. For example, if an individual who works from office has a 70% propensity score that individual is ‘‘matched’’ with an individual who works from home and has the same propensity score. All ‘‘matched’’ individuals who do not work from home are included in a control group and all ‘‘matched’’ who work from home are included in the treatment group. Matching by propensity scores assume that any difference between individuals who work from home and those who work from office can be accounted for by observable pre-treatment characteristics (Titus, M. A., 2007).

The ultimate goal is to attempt to balance covariates related to working from home. The propensity score is considered correctly specified when the balance between the two groups is achieved. The balance can only be evaluated on observed characteristics, if there exist unobserved confounders which are not taken into account then the two groups are not exchangeable and the model will not mimic a randomized control trial. To determine the balance between the two groups standardized differences are used. Standardized differences refer to a statistical measure used to compare the means or proportions of two groups in a study. It calculates the differences in means in units of the pooled standard deviation. It is recommended to consider the difference between the two groups as negligible when the threshold is less than or equal to 0.1 (Austin, 2011). Additionally, standardized mean differences of 0.1 and variance ratios approaching 1 but less than 2 could be considered balanced (Zhang et al., 2019). The results of the standardized differences and variance ratio of the covariates is presented in Figure 1, Empirical Estimation Results section.

Finally, the last step involves estimating the effect of productivity of WFH (0,1) using matching techniques based on propensity score estimation and calculating the estimates of the average treatment effect (ATE) and average treatment effect on the treated (ATT). Specifically, for this study full matching is implemented. The fundamental assumption behind employing propensity scores is based on the absence of unobserved confounding variables, as the propensity score can only establish balance between groups based on observed variables. Therefore, in order to increase robustness of the findings the study also applies instrument free methods to establish causality.

## 4.3 Gaussian Copula

To further address the potential endogeneity concerns in this analysis, I have employed the Gaussian Copula method. This choice is motivated by several key factors that make it particularly suitable for this study. Employing the Gaussian Copula method enhances the robustness of my findings, providing a more accurate depiction of the causal relationship between work-from-home practices and employee productivity. Moreover, the method is well-suited for cross-sectional data, like the Chinese General Social Survey (CGSS) used in this study. It effectively handles the data's structure and helps in deriving unbiased estimates of the effect of remote work on productivity. Park and Gupta (2012) introduced the method of Gaussian Copula to solve the issue of endogeneity arising from omitted variable bias and simultaneous causality, which leads to inconsistent estimates. Unlike the traditional Instrumental Variable method, this approach is an instrument free method, and does not require observed instruments to correct for endogeneity but relies on the distributional assumptions to identify the impact of the independent endogenous variable (Muehler et al., 2007).

The ability of Gaussian Copula Correction Approach to recover the true parameters relies on the model meeting four important assumptions. Firstly, the endogenous regressor needs to be continuous or have a certain number of discrete outcomes, therefore, it will support the endogenous dummy variable, work from home, in this analysis. The second assumption requires that the endogenous regressor cannot have a binomial distribution (needs to be non-normal), due to parameter identification problems. The third criterion is that the error term is assumed to be normally distributed and finally the fourth condition emphasizes the distributional properties, and thus requires that the distribution between the endogenous regressor and the error term can be explained by a Gaussian Copula. Therefore, according to Muehler (2007), if the model satisfies these conditions, the Gaussian Copula correction approach provides estimates that are almost similar to the estimates obtained from IV regression with medium strength instruments.

Following the works of Park and Gupta (2012), the general model is,

$$y = \alpha + \beta x + \beta_c xc + \varepsilon^* \qquad (4)$$

Here, ε* is assumed to be normally distributed white noise and the additional regressor defined as xc = $\Phi^{-1}\left(F_x(x)\right)$, where $\Phi^{-1}$ denotes the inverse of the standard normal CDF (Cumulative Distribution Function) and $F_x$ the CDF of x. However, since $F_x$ is unknown, the observed or estimated empirical CDF of x, $\widehat{F_x}$, is used. Thus, we rewrite the equation in its extended form:

$$y = \alpha + \beta x + \beta_c \widehat{xc} + \varepsilon^* \qquad (5)$$

Where, $\widehat{xc} = \Phi^{-1}\left(\widehat{F}_x(x)\right)$

Least square regression can then be applied on this extended regression equation to obtain the estimates of the parameters. Therefore, the procedure involves two stages, first the empirical CDF of the endogenous regressor is computed, second the estimated values from the first stage is used to construct the likelihood function. As a result, the standard errors and confidence intervals are derived from the sampling distributions generated through bootstrapping.

### 4.4 Lewbel Instrumental Variable

Another instrument-free approach was applied in this study, which is the identification through heteroskedasticity estimator (Lewbel, 2012). Lewbel proposes the use of heteroscedasticity in the endogenous and mismeasured variable to identify and construct an internal instrument to correct for endogeneity. Therefore, for robust modeling of endogeneity, Equations (2) was re-estimated using Lewbel's (2012) heteroscedasticity identified endogenous regressor estimator, which does not rely on any exclusion restrictions.

The two-stage estimation in this approach has the general form

$$Y_1 = Z\beta + Y_2\gamma + \epsilon_1 \tag{6}$$

$$Y_2 = Z\alpha + \epsilon_2 \tag{7}$$

In this study, the vector $Z$ consists of the exogenous variables such as education level, respondent age and gender. On the other hand, $Y_2$ is the work from home dummy variables and $Y_1$ is the logarithm of productivity.

Identification can be achieved without any exclusion restriction, if there exists a vector of exogenous variable $Z$ and the error is heteroscedastic. Lewbel (2012) provides identification and a simple two-stage least squares estimator to obtain unbiased coefficients. The key assumptions which need to be met for applying Lewbel (2012) are:

1. $Cov(Z, \epsilon_1\epsilon_2) = 0$
2. $Cov(Z, \epsilon_2^2) \neq 0$

Where the first condition is the standard exogenous assumption, that the vector consisting of the exogenous variables will be uncorrelated with the error terms. The second condition requires the error to be heteroscedastic.

For this model, the three dummy variables of work from home were used as demonstrated in Equation (2). The variables chosen to construct the instruments were, education level, age and gender. As it can argued that these variables are exogenous to productivity and are suitable for vector $Z$. The two-stage estimation proposed by Lewbel was implemented in this analysis. In the first stage the endogenous variable work from home, is regressed on all the control variables and the vector $Z$. From the first stage, the vector of residuals $\widehat{\epsilon_2}$, are obtained. These estimated residuals are then used to create instruments.

As shown by Lewbel (2007), identification requires that the error terms in the first stage regressions are heteroskedastic. The results of Breusch-Pagan test of heteroskedasticity show that the null of homoscedastic errors is clearly rejected in each case with a P-value equal to 0.00. In the second stage the set of instruments is then used to estimate the causal effects of work from home on productivity

# 5. Empirical Estimation Results

5.1 Ordinary Least Squares (OLS) Results

The results displayed in Table 2 from the Ordinary Least Squares (OLS) analysis indicate that all primary independent variables exhibit statistically significant coefficients. To begin with, the results depict a positive correlation with work from home and productivity, and it is significant at 1%. Secondly, it is observed that individuals who are completely independent at making their own decisions in their workplace are 23% more productive than individuals who have less autonomy, the result is significant at 10%. Similarly, trained individuals also show a positive relation to productivity, which reveals that individuals who have received training to improve their job skills are 8.9% more productive (significant at 1%, $p<0.01$). Employees who are union members also demonstrated higher productivity (significant at 5%, $p<0.05$). On the other hand, it is observed that females and individuals who have children display lower productivity. Moreover, individuals living in rural regions also show a negative association with productivity. As expected, it is observed that education has a positive effect on the dependent variable. In other words, a one-year increase in education increases productivity by 13.5%.

When remodeling the work from home variable into three dummies it is observed to have more insightful results which aligns with the previous research work. The results in the second column of in in Table 2 show that long hours of remote working or individuals who frequently work from home have lower productivity. Individuals who frequently work from home are 24.9% less productive than individuals who do not frequently work from home. On the other hand, individuals who occasionally work from home are 20.5% more productive than individuals who do not work from home. This result is supported by the findings of Kazekami(2020). Therefore, work from home increases productivity but only when it is done moderately, if individuals work from home on a regular basis, it is likely to decrease their productivity. This may be explained by individuals working longer hours due to not having a fixed end time, leading them to experience challenges in balancing work and personal life, additionally employees working from home may have poor ergonomic facilities leading to physical strain and lower productivity (Ferrara et al., 2022). The coefficients of autonomy, training and union dummy, in this regression display almost similar results as the first regression, increasing the reliability of the findings. The control variables in both the models also depict similar results and significance.

Incorporating the interaction terms and occupation proxies reveals that individuals who have received training and works from home show a positive association with productivity. This effect is observed in both the models (results in column 3 and 4, in Table 2). Therefore, individuals who have received training will demonstrate higher productive irrespective of whether they frequently or occasionally work from home. On the other hand, using education as a moderating variable show that a white-collar worker who works from home will have a negative effect on productivity compared to the baseline, the coefficient is statistically significant at the 5% level. White-collar workers who work from home experience a significant decrease in productivity, relative to either white-collar workers who do not work from home, or blue-collar workers. In practical terms, it suggests that working from home may not be as effective for white-collar workers in maintaining productivity levels as it might be for other types of workers.

Therefore, while working from home may be beneficial in general, its effect on productivity is lower for white-collar workers compared to others. This could be due to reduced collaboration and communication, which are often essential in white-collar jobs (Bloom et al., 2015). Additionally, the blurred work-life boundaries in remote settings might contribute to reduced efficiency for these workers (Gajendran & Harrison, 2007).

**Table 2**

*OLS Regression Results*

| | Dependent Variable: ln (Productivity) | | | | |
|---|---|---|---|---|---|
| | (1) Single Dummy WFH Model | (2) Three dummies WFH Model | (3) Single Dummy WFH Model with interaction terms | (4) Three dummies WFH Model with interaction terms | (5) Single Dummy WFH Model with proxy terms |
| WFH | 0.097***<br>(0.011) | ------- | 0.094***<br>(0.017) | ------- | 0.167***<br>(0.074) |
| WFH_freq | ------- | -0.249**<br>(0.123) | ------- | -0.218***<br>(0.073) | ------- |
| WFH_occ | ------- | 0.205***<br>(0.043) | ------- | 0.144***<br>(0.002) | ------- |
| WFH_rare | ------- | 0.067<br>(0.052) | ------- | 0.097*<br>(0.056) | ------- |
| Trained | 0.089***<br>(0.031) | 0.088***<br>(0.028) | 0.043**<br>(0.017) | 0.001<br>(0.012) | 0.207***<br>(0.012) |

| | Dependent Variable: ln (Productivity) | | | | |
|---|---|---|---|---|---|
| | (1) Single Dummy WFH Model | (2) Three dummies WFH Model | (3) Single Dummy WFH Model with interaction terms | (4) Three dummies WFH Model with interaction terms | (5) Single Dummy WFH Model with proxy terms |
| Full Autonomy | 0.231*<br>(0.137) | 0.235*<br>(0.139) | 0.234<br>(0.150) | 0.236<br>(0.148) | 0.199<br>(0.147) |
| Union member | 0.171**<br>(0.086) | 0.171**<br>(0.084) | 0.172**<br>(0.086) | 0.177**<br>(0.083) | 0.297***<br>(0.091) |
| Female | -0.336***<br>(0.106) | -0.334***<br>(0.106) | -0.336***<br>(0.107) | -0.335***<br>(0.108) | -0.348***<br>(0.131) |
| Education | 0.135***<br>(0.007) | 0.135***<br>(0.007) | 0.135***<br>(0.007) | 0.135***<br>(0.007) | ------- |
| Married | 0.110***<br>(0.018) | 0.110***<br>(0.017) | 0.110***<br>(0.018) | 0.105***<br>(0.019) | 0.105***<br>(0.019) |
| Age | 0.061***<br>(0.018) | 0.061***<br>(0.017) | 0.061***<br>(0.018) | 0.061***<br>(0.018) | 0.061***<br>(0.023) |
| Age Square | -0.001***<br>(0.000) | -0.001***<br>(0.000) | -0.001***<br>(0.0001) | -0.001***<br>(0.0001) | -0.001***<br>(0.0002) |
| Has Children | -0.061***<br>(0.014) | -0.059***<br>(0.012) | -0.061***<br>(0.015) | -0.057***<br>(0.014) | -0.068***<br>(0.024) |
| Location | -0.419***<br>(0.039) | -0.418***<br>(0.040) | -0.419***<br>(0.039) | -0.418***<br>(0.040) | -0.492***<br>(0.0004) |
| WFH*Trained | ------- | ------- | 0.073***<br>(0.016) | ------- | ------- |
| WFH_freq*Trained | ------- | ------- | ------- | 0.443***<br>(0.016) | ------- |
| WFH_Occ*Trained | ------- | ------- | ------- | 0.278***<br>(0.073) | ------- |
| WFH*Full Autonomy | ------- | ------- | -0.030<br>(0.131) | ------- | ------- |
| WFH_freq*Full Autonomy | ------- | ------- | ------- | -0.251<br>(0.182) | ------- |
| WFH_occ*Full Autonomy | ------- | ------- | ------- | 0.026<br>(0.159) | ------- |
| Blue collar | ------- | ------- | ------- | ------- | 0.446***<br>(0.009) |
| White collar | ------- | ------- | ------- | ------- | 1.114***<br>(0.128) |
| White collar*WFH | ------- | ------- | ------- | ------- | -0.214**<br>(0.009) |
| No. of Observations | 4422 | 4420 | 4420 | 4416 | 4421 |
| $R^2$ | 0.3146 | 0.3153 | 0.3143 | 0.3152 | 0.2831 |
| F-statistic: | 186<br>[0.000] | 158<br>[0.000] | 157.3<br>[0.000] | 121<br>[0.000] | 146.8<br>[0.000] |

*Note:* Heteroscedastic robust SEs in parentheses; except for F-statistic, p-values in square brackets.
*, ** and *** denote statistical significance at the 10%, 5% and 1% levels of significance

## 5.2 Multinomial Logistic Regression Results

The results of the multinomial logistic regression in terms of the relative risk ratio (RRR), offer valuable insights into the characteristics of individuals who are more inclined to engage in remote work. The relative risk ratio demonstrates how the probability of an outcome falling in a specific group compares to the probability of the outcome falling in the reference group changes with the independent variable. If the coefficient has a value greater than one, it indicates the chances of the outcome falling in the specific group relative to the chance of the outcome falling in the reference group increases as the independent variable increases. Similarly, an RRR of less than 1, portrays a negative relation or in other words it means that the chances of the outcome falling in the reference group decreases as the independent variable changes one unit, holding all other variables constant. In this analysis the reference group is "work from home - never".

From the results in Table 3(a)(in terms of the relative risk ratio), it can be concluded that holding all other variables constant, employees who are completely independent to make their own decisions are more likely to always work from home. If an employee were to increase his or her status of autonomy by one unit (the dummy variable having a value of 1, so the individual has more independence in their workplace), the relative risk for preferring to work from home frequently as opposed to "work from home - never" would increase by a factor of 3.608 given the other variables in the model are held constant. In other words, given a one unit increase in autonomy, the relative risk of being in the WFH-always group would be 3.608 times more likely when the other variables in the model are held constant, the result is statistically significant at 1%.

Secondly, the results show that the variable trained also demonstrates a positive relation and is significant at 1%. So, an individual who has received training is 41.38 times more likely to work from home, compared to someone who received no training to improve their job skills. On the other hand, union members are less likely to work from home, possibly because the most unionized industries are the industries that cannot conduct work from home due to its structure. According to the Bureau of Labor Statistics (2024), the highest unionization rates were in education, library occupations and protective service occupations. Moreover, females are more likely to work from home more frequently than males, adding to this, married individuals are more likely to work from home occasionally but not always. Those who have children are more likely to work from home frequently as opposed to working in an office, this may be because female married individuals with children usually have more household and childcare responsibilities than their male partner (Scommegna, 2020), making WFH a more preferable option.

Finally, holding all other variables constant, as the number of years of education increases, individuals are less likely to choose to work from home. The results reveal that educated individuals are 0.84 times less likely to prefer to work from home on a regular basis which is inconsistent with previous literature.

The estimates of the marginal effects in Table 3 (b) aid to measure the expected change in probability of a particular category with respect to a unit change in an independent variable. The interpretation of the results of marginal effects and RRR follow the same direction. Individuals who have more autonomy in their workplace has a tendency to always work from home by 1%. Additionally, trained individuals are 6.1% more likely to sometimes work from home as opposed to working from office. Finally, a one increase in education years decreases the probability of always working from home by 0.1%. This may reflect that educated individuals prefer to occasionally conduct remote working as opposed to “always” working from home, as they realize that working from home on a regular basis can lead to certain drawbacks such as burnout or difficulty with collaboration. However, it is to be noted that according to results, the overall predictive power of the estimated multinomial logit model is 14 percent. It means that this model and its coefficients do not have sufficient predicting power of the employee’s behaviour.

**Table 3**
*Multinomial Logistic Regression Results*

| | Dependent Variable: WFH | | | |
|---|---|---|---|---|
| | Always | Often | Sometimes | Rarely |
| | (1) | (2) | (3) | (4) |
| **(a) Relative Risk Ratio** | | | | |
| Autonomy | 3.608*** | 2.321*** | 0.988 | 0.970 |
| | (0.040) | (0.207) | (0.248) | (0.204) |
| Trained | 41.383*** | 35.635*** | 32.958*** | 50.134*** |
| | (0.031) | (0.118) | (0.218) | (0.190) |
| Union Member | 0.187*** | 0.767*** | 0.546*** | 0.771 |
| | (0.003) | (0.048) | (0.131) | (0.222) |
| Female | 1.543*** | 0.667 | 0.790 | 0.774 |
| | (0.025) | (0.251) | (0.205) | (0.161) |
| Education | 0.843** | 0.895** | 1.010 | 0.959 |
| | (0.070) | (0.047) | (0.036) | (0.029) |
| Married | 0.626*** | 1.950*** | 1.946*** | 1.018 |
| | (0.012) | (0.091) | (0.088) | (0.240) |
| Age | 1.202*** | 0.961 | 1.091*** | 0.920*** |
| | (0.030) | (0.027) | (0.024) | (0.018) |
| Has Children | 1.246*** | 0.909 | 0.667** | 1.197 |
| | (0.023) | (0.259) | (0.197) | (0.190) |
| Akaike Inf. Crit. | 3,252.852 | 3,252.852 | 3,252.852 | 3,252.852 |
| **(b) Marginal Effects** | | | | |
| Autonomy | 0.010*** | 0.012*** | 0.001 | -0.001 |
| | (0.003) | (0.004) | (0.006) | (0.008) |
| Trained | 0.023*** | 0.038*** | 0.061*** | 0.126*** |
| | (0.004) | (0.005) | (0.007) | (0.009) |
| Union Member | -0.013 | -0.002 | -0.011 | -0.004 |
| | (0.010) | (0.006) | (0.008) | (0.009) |
| Female | 0.004 | -0.005 | -0.004 | -0.008 |
| | (0.003) | (0.004) | (0.005) | (0.006) |
| Education | -0.001** | -0.001* | 0.001 | 0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Married | -0.004 | 0.008 | 0.014* | -0.003 |
| | (0.004) | (0.006) | (0.007) | (0.009) |
| Age | 0.001 | 0.000 | 0.002 | -0.003 |
| | (0.001) | (0.001) | (0.002) | (0.002) |
| Has Children | 0.002 | -0.001 | -0.010* | 0.008 |
| | (0.003) | (0.004) | (0.006) | (0.007) |
| Number of observations | 4,434 | | | |
| Wald chi square | 484.66 [0.0000] | | | |
| Pseudo $R^2$ | 0.1402 | | | |
| Log pseudolikelihood | -1587.4253 | | | |

*Note:* Heteroscedastic robust SEs in parentheses; except for Wald chi square-statistic, p-values in square brackets.

*, ** and *** denote statistical significance at the 10%, 5% and 1% levels of significance

### 5.3 Propensity Score Analysis Results

Therefore, next this study uses the propensity score matching method to estimate the impact of working from home on productivity. The final sample consists of 4,434 observations, where the total number of observations in the control group is 4028 and the treatment group consisted of 406 individuals. The mean attributes of this sample before the propensity analysis are shown in Table – 5, below.

**Table 5**

*Mean attributes of treated and control group*

| | Mean (S.E) | |
|---|---|---|
| | Control Group (WFH = 0) | Treatment Group (WFH = 1) |
| Log Productivity | 2.366 | 2.647 |
| | (1.192) | (1.172) |
| Union Member | 0.132 | 0.153 |
| | (0.338) | (0.360) |
| Trained | 0.013 | 0.315 |
| | (0.113) | (0.465 |
| Full Autonomy | 0.172 | 0.229 |
| | (0.378) | (0.421) |
| Female | 0.436 | 0.441 |
| | (0.496) | (0.497) |
| Married | 0.826 | 0.842 |
| | (0.379) | (0.365) |
| Education | 5.581 | 6.113 |
| | (3.202) | (3.281) |
| Age | 42.501 | 40.714 |
| | (11.144) | (10.733) |
| Has Children | 0.424 | 0.483 |
| | (0.494) | (0.500) |
| No. of Observation | 4028 | 406 |

*Note:* Heteroscedastic SEs in parentheses. WFH = 0, represents the individuals who work from office, WFH = 1, includes all individuals who work from home.

**Figure 1** lists the comparison of the standardized mean differences between treatments for each of the covariates in the observed data and after inverse probability weighting by the estimated PSs. A general rule of thumb to identify good balancing is to see whether the standardized mean differences are less than or equal to 0.1 for each covariate. Additionally, standardized mean differences of 0.1 and variance ratios approaching 1 but less than 2 could be considered balanced (Zhang et al., 2019). In this case, I achieved a good balance for each of the covariates after weights.

**Figure 1**

*Standardized differences and variance ratio of the covariates*

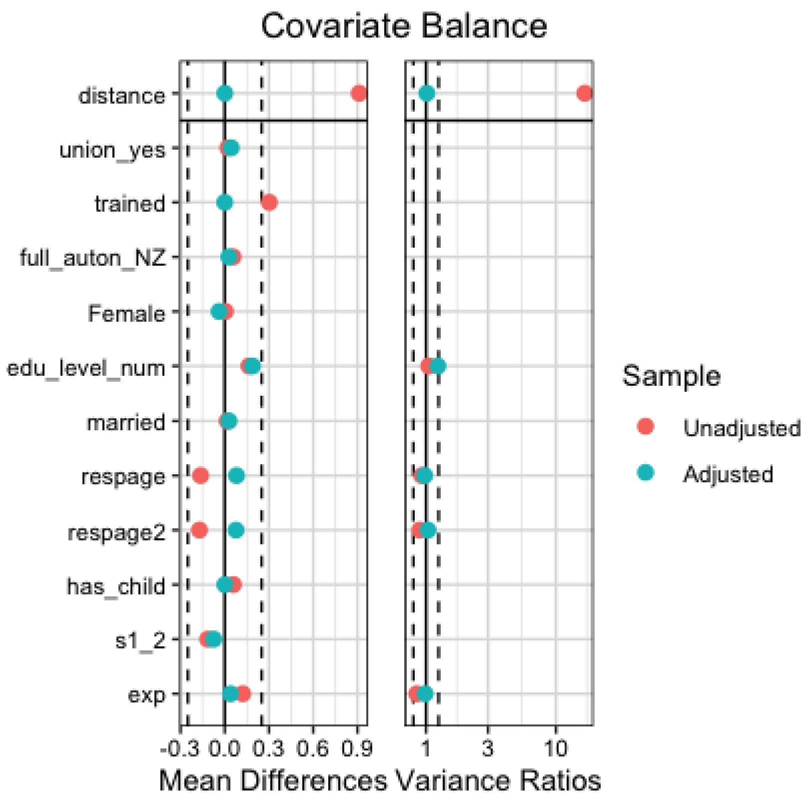

**Table 6**

*Propensity Score Model Estimates*

| | Propensity score application method: Full Matching | |
|---|---|---|
| | ATE | ATT |
| WFH | 0.146** | 0.152** |
| | (0.050) | (0.067) |
| No. of observations | 4421 | |
| $R^2$ | 0.3357 | |
| F-statistic: | 188 | |
| | [0.000] | |

*Note:* The propensity score estimation model includes the same covariates as the baseline regression model: union member, trained, full autonomy, gender, education level, marital status, age, age square, whether the respondent has a child and location.
ATE = Average Treatment Effect; ATT = Average Treatment Effect on the Treated
Heteroscedastic robust SEs in parentheses; except for F-statistic, p-values in square brackets.
*, ** and *** denote statistical significance at the 10%, 5% and 1% levels of significance

The ATE of working from home on productivity, as measured by the logarithm of productivity (log of income earned per hour work), is 0.146 (results shown in Table 6, below) and is statistically significant at 5%. Similarly, the ATT estimates show a 15.2 percentage point increase. This suggests that individuals working from home, after accounting for observed covariates, exhibit an increase in productivity compared to their counterparts who do not work from home. On average, remote workers earn 15.2 percentage point higher income per hour of work than office workers, holding all other variables constant. Thus, the effect is increased in this analysis compared to ordinary least square regression.

### 5.4 Gaussian Copula Correction Results

Table 7 presents the results of the Gaussian Copula correction applied to the regression model examining the relationship between work-from-home (WFH) and productivity. Since the endogenous variable is a discrete variable, the estimation is carried out using augmented OLS estimation which is based on Gaussian Copulas. Additionally, the reported standard errors are the bootstrapped standard errors, due to the inference being done in two steps. The percentile confidence intervals are also reported as the distribution of the bootstrapped parameter is skewed.

The results of this analysis suggest that there is a statistically significant positive causal relationship between WFH and productivity. The estimated coefficient of 0.16 indicates that, on average, individuals who work from home experience a 0.16 unit increase in productivity compared to those who do not, holding other factors constant. However, it's important to note that the actual effect could range from a decrease of 0.051 to an increase of 0.233 units, based on the confidence intervals. However, it is to be noted that the limitation of the Gaussian Copula correction approach is that it cannot handle the endogeneity that may arise from self-selection.

**Table 7**

*Gaussian Copula Correction Estimates:*

| | Point Estimate | Lower Boots CI (95%) | Upper Boots CI (95%) |
|---|---|---|---|
| WFH | 0.160**<br>(0.073) | -0.051 | 0.233 |
| Trained | 0.091<br>(0.074) | -0.058 | 0.227 |
| Full Autonomy | 0.231***<br>(0.045) | 0.146 | 0.317 |
| Union Member | 0.170***<br>(0.041) | 0.093 | 0.251 |
| Female | -0.337***<br>(0.030) | -0.391 | -0.279 |
| Education | 0.136***<br>(0.006) | 0.125 | 0.147 |
| Married | 0.110**<br>(0.046) | 0.023 | 0.203 |
| Age | 0.061***<br>(0.011) | 0.039 | 0.082 |
| Age Square | -0.001***<br>(0.000) | -0.001 | -0.001 |
| Has Children | -0.063*<br>(0.034) | -0.124 | 0.005 |
| Location | -0.419***<br>(0.036) | -0.487 | -0.347 |

*Note:* The table also reports the bootstrapped standard errors and the confidence interval. The confidence intervals are also the bootstrapped confidence intervals, due to the non-normality of the bootstrapped parameters (Gui, 2020).

*, ** and *** denote statistical significance at the 10%, 5% and 1% levels of significance

## 5.5 IV Lewbel Analysis Results

Table 8 (below), provides the findings from the two versions of the work from home model estimated using the Lewbel (2012) method. The Lewbel (2012) estimates of the "Three dummies WFH variable" model carry virtually identical signs to the OLS regression estimates. However, there are some exceptions such as, the sign and significance level of the WFH dummy in the "Single Dummy WFH" model is negative and insignificant. The internal instrumental variable approach employed here ensures robustness and further enhances the reliability of the findings. Notably, the coefficient for the work from home variable(occasion) is 0.19, again indicating that working from home occasionally has a positive impact on the productivity of employees, the result is statistically significant at 5%. On the other hand, the negative coefficient of frequently working from home on productivity aligns with the findings of previous literatures.

The complete autonomy variable has a positive effect and is significant at 1% and the estimated coefficient is 0.24. The baseline estimates of the effects of working from home and productivity reported in Table 2(column 2) are larger than the Lewbel IV estimates, but the difference is small. According to the Lewbel (2012) estimates, working from home and having complete independence in one's job role are both statistically significant (at 1 percent level) on an employee's productivity. The positive effects of working from home on productivity are larger than the effects of traditional office work. The explanatory powers of the Lewbel (2012) estimates range from 30.6% to 31.5%. The Wald (F) statistics and respective *p*-values attest to the joint significance of the coefficient estimates in both the models. The evidence from this alternative identification approach thus provides strong support to the conclusions reached earlier based on the Gaussian Copula correction estimates that rely on using augmented OLS estimation which is centered on Gaussian copulas.

**Table 8**

*Lewbel (2012) Estimates*

| | Dependent Variable: ln (Productivity) | |
|---|---|---|
| | (1) Single Dummy WFH Model (WFH) | (2) Three dummies WFH Model (WFH_freq, WFH_occ, WFH_rare) |
| WFH | -0.332<br>(0.933) | - |
| WFH_freq | - | -0.625*<br>(0.350) |
| WFH_occ | - | 0.196**<br>(0.081) |
| WFH_rare | - | 0.057<br>(0.078) |
| Trained | 0.370<br>(0.614) | 0.110<br>(0.087) |
| Full Autonomy | 0.241***<br>(0.045) | 0.240***<br>(0.040) |
| Union Member | 0.160**<br>(0.054) | 0.169***<br>(0.048) |
| Female | -0.342***<br>(0.033) | -0.333***<br>(0.030) |
| Education | 0.134***<br>(0.007) | 0.135***<br>(0.006) |
| Married | 0.117*<br>(0.050) | 0.105*<br>(0.047) |
| Age | 0.060***<br>(0.011) | 0.061***<br>(0.011) |
| Age Square | -0.001***<br>(0.000) | -0.001***<br>(0.000) |
| Has Children | -0.062<br>(0.037) | -0.059<br>(0.036) |
| Location | -0.430***<br>(0.044) | -0.418***<br>(0.035) |
| No. of observations | 4420 | 4420 |
| $R^2$ | 0.306 | 0.3145 |
| Wald (F) test: | 183.4*** | 158*** |

*Note:* Heteroscedastic robust SEs in parentheses; except for Wald test, p-values in square brackets.
*, ** and *** denote statistical significance at the 10%, 5% and 1% levels of significance.
WFH_freq represent individuals who frequently work home, WFH_occ and WFH_rare includes the group who occasionally, and rarely work from home, respectively.

To ensure the robustness of the findings of the Lewbel (2012) estimates, several diagnostic tests have been conducted. These additional tests, reported in Table 9, consistently support the results, affirming the reliability of the IV Lewbel estimates. In order for the IV Lewbel estimate to be a better estimator, the model needs to satisfy two conditions (standard exogenous assumption and the error needs to be heteroscedastic). The diagnostic tests indicate that both the models suffer from heteroskedasticity at 1% level. This satisfies the first condition of using Lewbel (2012) estimator. Secondly, the models also satisfy the exogeneity assumption; the null of exogeneity of the set of Lewbel IVs cannot be rejected with a P-value equal to 0.48 for the Sargan overidentifying restrictions statistics, meaning that the extra instruments are valid (are uncorrelated with the error term). Finally, the weak instrument test rejects the null hypothesis, demonstrating that at least one instrument created in this analysis is strong, so the instruments based on heteroskedasticity perform well. Therefore, this model provides consistent and unbiased estimates.

**Table 9**

*Diagnostic Tests conducted for the Lewbel (2012) estimates*

| | (1) Single Dummy WFH Model (WFH) | | (2) Three dummies WFH Model (WFH_freq, WFH_occ, WFH_rare) | |
|---|---|---|---|---|
| | Statistic | *p*-value | Statistic | *p*-value |
| Breusch − Pagan $(\chi^2)^a$ | 132.24 | <2e-16 *** | 141.59 | <2e-16 *** |
| Weak instruments $(F)^b$ | 5.689 | 0.001*** | 434.240 | <2e-16 *** |
| Wu − Hausman$^c$ | 0.215 | 0.643 | 1.505 | 0.220 |
| Sargan Test $(J)^d$ | 1.845 | 0.397 | 1.449 | 0.485 |

*Note:* [a]Breusch-Pagan test for heteroskedasticity in OLS estimates. Test conducted using OLS regression without robust SEs. $H_o$: residuals are homoscedastic.
[b]Weak instruments − for Lewbel (2012) estimates in table 8. $H_o$: Instruments are only weakly correlated with the endogenous regressor – that is, weakly-identified model.
[c]Wu − Hausman for instrument overidentification restriction - for Lewbel (2012) estimates in table 8. $H_o$:Cov(x,e)=0 The variable is not endogenous
[d]Sargan Test for instrument overidentification restriction - for Lewbel (2012) estimates in table 8. $H_o$: Cov(z,e)=0 The covariance between the instrument and the error term is zero, that is, the instruments are all valid.
*** denote rejection of relevant null hypothesis at the 1% levels of significance.

5.6 Robustness Check

Further test for omitted variable bias and stability of the coefficients of the work from home dummies – WFH, WFH_freq, WFH_occ and WFH_rare , is conducted using Oster (2019) method. A common procedure to test for robustness to omitted variable bias is to evaluate coefficient and R-squared movements before and after including control variables. Therefore, Oster (2019) requires analyzing both the restricted and full versions of the model. The restricted version of the model focuses solely on the dependent variable (productivity) in relation to the explanatory variable of interest (e.g., WFH dummies). The full version of the estimated model typically represents the full specification used in the baseline regressions. It often involves a sensitivity analysis that estimates how much the coefficient would change if there were omitted variables affecting the key explanatory variable. A higher sensitivity parameter often suggests that the estimated effect is less sensitive to omitted variables and thus more robust.

The Oster test helps to evaluate how stable the results are in the face of potential omitted variable bias. The estimated coefficient of proportionality of the WFH dummies from the Oster (2019) test is greater than 1, in all the tests (Table 10). Therefore, the results suggest that the effect of work from home on productivity remains stable or robust and does not suffer from any coefficient instability. This robustness enhances the credibility of my findings and suggests that the observed effect of WFH is not unduly influenced by omitted variables, further reinforcing the reliability of my results.

**Table 10**

*Oster (2019) Test for Coefficient Stability*

| | Dependent Variable: ln(Productivity) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Restricted | Full model | Restricted | Full model | Restricted | Full model | Restricted | Full model |
| Treatment | (1) | (2) | (1) | (2) | (1) | (2) | (1) | (2) |
| WFH | 0.281*** | 0.127** | | | | | | |
| WFH_freq | | | -0.365 | -0.238 | | | | |
| WFH_occ | | | | | 0.415*** | 0.239*** | | |
| WFH_rare | | | | | | | 0.256*** | 0.099 |
| Controls | NO | YES | NO | YES | NO | YES | NO | YES |
| Constant | YES | YES | YES | YES | YES | YES | YES | YES |
| No. of observations | 4,434 | 4,434 | 4,434 | 4,434 | 4,434 | 4,434 | 4,434 | 4,434 |
| $R^2$(WITHIN) | 0.005 | 0.295 | 0.001 | 0.295 | 0.004 | 0.295 | 0.002 | 0.295 |
| $R_{Max}$ | | 0.384 | | 0.384 | | 0.384 | | 0.384 |
| $\beta_{Restricted} - \beta_{Full}$ Work From Home Measure | | 0.154 | | -0.127 | | 0.176 | | 0.157 |
| $R_{Max} - R_{Full}$ | | 0.089 | | 0.089 | | 0.089 | | 0.089 |
| $R_{Full} - R_{Restricted}$ | | 0.290 | | 0.294 | | 0.291 | | 0.293 |
| Bias-adjusted treatment effect ($\beta^*$) ($R_{Max} = R_{Full} + (R_{Full} - R_{Restricted})$) Work From Home Measure | | 0.080 | | -0.200 | | 0.186 | | 0.051 |
| Coefficient of proportionality $(\delta)^a$ ($R_{Max} = \min(1.3R_{Full}, 1)$) Work From Home Measure | | 2.703 | | 6.25 | | 4.464 | | 2.079 |

*Note:* $R_{Max}$ denotes the $R^2$ from regression estimate of the treatment effect. $R_{Full}$ denotes the $R^2$ from the regression estimate of the treatment effect using a full set of observed control variables. $R_{Restricted}$ denotes $R^2$ from regression estimate of the treatment effects using the outcome and treatment variables only.

$^a\delta > 1$ implies robustness of the treatment effect.

*, ** and *** denote statistical significance at the 10%, 5% and 1% levels of significance

# 6. Summary of Findings and Conclusion

Using data from the Chinese General Social Survey (CGSS), this study investigates the effects of remote work, autonomy, and training on employee performance measured by their productivity. The empirical analyses confirmed all hypotheses, each showing significant statistical results. This suggests that the sample data effectively represents the population. The results support the hypothesis that working from home increases an individual's productivity and shares a causal relationship, however, conducting telework for long hours has a negative effect on productivity. Second, the empirical findings indicate that employees who receive greater autonomy in their workplace demonstrate higher productivity, which is in congruence with the results of MacEachen et al. (2007). Another finding of the study was how job training demonstrated significant effects in determining employee productivity, which is expected as training helps employees to attain new skills and motivates them to attain organizational goals and helps with personal development, this finding is supported by Kochan’s (1995) study, who found that organizations that invest more in training demonstrated higher productivity levels.

The primary focus of this study was to analyze the causal impact of work from home on productivity, and several regressions have been run to investigate this, all providing supporting results. Therefore, these results are in accordance with previous works of Bloom (2015) and Kazekami (2020), working from home occasionally every week allows individuals to be more productive. On the other hand, full-time working from home has been shown to negatively impact employees' performance in this study, as it is difficult to accomplish all tasks remotely, creativity is limited in isolation, and it can be difficult to stay motivated, as seeing our colleagues can drive a sense of inspiration and healthy competition.

Therefore, an important insight of this study is that employees should work from home occasionally every week, as it will reduce the exhaustion from commuting, employees will take fewer sick leaves, will be able to better manage their work and personal life and be more satisfied. However, it is to be noted that the results will vary among individuals. A study conducted by Mehdi & Morissette, (2021), found that whether teleworkers accomplish more work per hour or not, varies across industries and workers characteristics. The multivariate analysis, using ordered logit models, revealed that public administrators, health care workers and social assistants had reported doing more work per hour, than teleworkers in educational services and production industries. This is a limitation of this study, due to the unavailability of the variable in the dataset it could not capture the differences in varying occupations.

Additionally, key variables such as collaboration, communication, and work-life balance were missing. Including these variables would have provided more comprehensive and insightful results. We must also take into account the results related to autonomy and training only account for association properties and our interpretations are limited to such, thus we cannot infer causality. As a result, these two variables may be subject to endogeneity bias. Future research can look into the causal effect of these variables by using longitudinal data.

This study is an informative insight for policymakers at various dimensions, such as those involved in labor and employment departments, officials who specialize in workforce development and human resource management. Therefore, this study has two significant implications. Firstly, the study extends the literature on remote work and the causal relationship contributes to the theoretical understanding of work from home dynamics. It sheds light on remote frequency and its effect on productivity, highlighting the significance of considering the structure of remote work arrangements. The study points out that employers can work on a strategy that will maximize employee output in a remote or hybrid work-setup, while avoiding potential negative effects associated with prolonged remote work. The findings also underscore the importance of investing in job training and empowering employees with greater autonomy even in a traditional office setup providing relevant training opportunities. Providing relevant training opportunities and higher independence not only improves individual productivity but also overall organizational growth.

# A: Appendix

**Table A.1:**

*Demographic differences of male and female employees, mean and standard deviation*

| | Male | Female | P-Test |
|---|---|---|---|
| Observation(n) | 2497 | 1937 | |
| Income (mean (SD)) | 48863.9 | 31174.82 | <0.001 |
| | (130502.8) | (43196.1) | |
| Weekly work hours (mean (SD)) | 50.29 | 47.17 | <0.001 |
| | (19.9) | (20.14) | |
| Productivity (mean (SD)) | 26.54 | 18.24 | <0.001 |
| | (82.17) | (31.38) | |
| Work from home (%) | 227 | 179 | 0.905 |
| | (9.1) | (9.2) | |
| Received college education (%) | 528 | 462 | 0.035 |
| | (21.2) | (23.9) | |
| Age (mean (SD)) | 42.73 | 41.83 | 0.008 |
| | (11.5) | (10.59) | |
| Receives full autonomy (%) | 448 | 339 | 0.733 |
| | (17.9) | (17.5) | |
| Has training (%) | 81 | 99 | 0.002 |
| | (3.2) | (5.1) | |
| Experience (mean (SD)) | 11.36 | 9.42 | <0.001 |
| | (11.92) | (10.11) | |
| Has children (%) | 1044 | 859 | 0.096 |
| | (41.8) | (44.3) | |
| Rural (%) | 1041 | 714 | 0.001 |
| | (41.7) | (36.9) | |
| Union member (%) | 359 | 234 | 0.029 |
| | (14.4) | (12.1) | |
| Always works from home | 15 | 21 | |
| | (0.6) | (1.1) | |
| Often works from home | 39 | 24 | |
| | (1.6) | (1.2) | |
| Sometimes works from home | 59 | 48 | |
| | (2.4) | (2.5) | |
| Rarely works from home | 114 | 86 | |
| | (4.6) | (4.4) | |
| Married (%) | 2039 | 1632 | 0.026 |
| | (81.7) | (84.3) | |

**Table A.2**

*Variable definition*

| Variable | Type | Definition |
|---|---|---|
| Work From Home (WFH) | Dummy | 1 if respondent works from home in any frequency, 0 otherwise (never works from home) |
| Work From Home - frequently (WFH_Freq) | Dummy | 1 if respondent always works from home, 0 otherwise |
| Work From Home - occasionally (WFH_Occ) | Dummy | 1 if respondent often or sometimes works from home, 0 otherwise |
| Work From Home - rarely (WFH_Rare) | Dummy | 1 if respondent rarely works from home, 0 otherwise |
| Full Autonomy | Dummy | 1 if respondent is completely independent to make their own decisions in their work, 0 otherwise |
| Trained | Dummy | 1 if respondent have received training to improve their job skills, 0 otherwise |
| Annual labor income | Continuous | Total professional/labor income previous year |
| Annual working hours | Continuous | Number of working hours per week multiplied by 48 |
| Productivity | Continuous | Ratio of annual labor income by the annual working hour |
| Gender | Dummy | 1 if respondent is a female, 0 if male |
| Age | Continuous | Number of years |
| Age Square | Continuous | Number of years squared |
| Education | Continuous | Years/highest level of education |
| Martial Status | Dummy | 1 if individual is married, 0 otherwise |
| Presence of Child | Dummy | 1 if individual has a child, 0 if the individual has no children |
| Region | Dummy | 1 if respondent lives in the city, 0 otherwise |
| Union member | Dummy | 1 if respondent is a member of the union, 0 otherwise |

# References:


Alipour, J.-V., Fadinger, H., & Schymik, J. (2021). *My home is my castle: The benefits of working from home during a pandemic crisis. Evidence from Germany* (Working Paper 329). ifo Working Paper. https://www.econstor.eu/handle/10419/222850

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Barrett, A. and O'Connell, P.J. (1999), Does Training Generally Work? The Returns to In-company Training, IZA discussion paper No. 51, Forschungsinstitut zur Zukunft der Arbeit, Bonn.

Bentley, T.A.; Teo, S.T.T.; McLeod, L.; Tan, F.; Bosua, R.; Gloet, M. The Role of Organisational Support in Teleworker Wellbeing: A Socio-Technical Systems Approach. Appl. Ergon. 2016, 52, 207–215.

Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does Working from Home Work? Evidence from a Chinese Experiment *. *The Quarterly Journal of Economics*, *130*(1), 165–218. https://doi.org/10.1093/qje/qju032

Baum, C. F., & Lewbel, A. (2019). Advice on using heteroskedasticity-based identification. *The Stata Journal*, *19*(4), 757–767. https://doi.org/10.1177/1536867X19893614

Conen, W.S. (2012) 'Ageing and employers' perceptions of labour costs and productivity: a survey among European employers', *International Journal of Manpower*, Vol. 33, No. 6, pp.629–647.

Deci, E. L. (1975). Intrinsic motivation. New York, NY: Plenum.

DuBrin, A. J. (1991). Comparison of the job satisfaction and productivity of telecommuters versus in-house employees: A research note on work in progress. Psychological Reports, 68, 1223-1234.

Ferrara, B., Pansini, M., De Vincenzi, C., Buonomo, I., & Benevene, P. (2022). Investigating the Role of Remote Working on Employees' Performance and Well-Being: An Evidence-Based Systematic Review. International Journal of Environmental Research and Public Health, 19(19), Article 19. https://doi.org/10.3390/ijerph191912373

Fonner, K. L., & Roloff, M. E. (2010). Why Teleworkers are More Satisfied with Their Jobs than are Office-Based Workers: When Less Contact is Beneficial. Journal of Applied Communication Research, 38(4), 336–361. https://doi.org/10.1080/00909882.2010.513998

Gajendran, R.S.; Harrison, D.A. The Good, the Bad, and the Unknown about Telecommuting: Meta-Analysis of Psychological Mediators and Individual Consequences. J. Appl. Psychol. 2007, 92, 1524.

Godard-Sebillotte, C., Karunananthan, S., & Vedel, I. (2019). Difference-in-differences analysis and the propensity score to estimate the impact of non-randomized primary care interventions. *Family Practice*, *36*(2), 247–251. https://doi.org/10.1093/fampra/cmz003

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. Journal of

Applied Psychology, 60, 159–170.

Hill, A., Johnson, S., Greco, L., O'Boyle, E., & Walter, S. (2020). Endogeneity: A Review and Agenda for the Methodology-Practice Divide Affecting Micro and Macro Research. Journal of Management, 47, 014920632096053. https://doi.org/10.1177/0149206320960533

Kazekami, S. (2020). Mechanisms to improve labor productivity by performing telework. *Telecommunications Policy*, *44*(2), 101868. https://doi.org/10.1016/j.telpol.2019.101868

MacDuffie, J.P. and Kochan, T.A. (1995) 'Does US firms invest less in human resources? Training in the world auto industry', *Industrial Relations: A Journal of Economy and Society*, Vol. 34, No. 2, pp.147–168.

MacEachen, E., Polzer, J., & Clarke, J. (2008). "You are free to set your own hours": Governing worker productivity and health through flexibility and resilience. Social Science & Medicine, 66(5), 1019–1033. doi:10.1016/j.socscimed.2007.11.013

Mann, S., Varey, R., Button, W., 2000. An exploration of the emotional impact of tele-working via computer-mediated communication. J. Manag. Psychol. 15, 668e690.

Mann, S., Holdsworth, L., 2003. The psychological impacts of teleworking: stress, emotions and health. New Technol. Work Employ. 18, 196e211.

Mehdi, T., & Morissette, R. (2021). *Working from home: Productivity and preferences*.

Muehler, G., Beckmann, M., & Schauenberg, B. (2007). The returns to continuous training in Germany: New evidence from propensity score matching estimators. *Review of Managerial Science*, *1*(3), 209–235. https://doi.org/10.1007/s11846-007-0014-6

Neufeld DJ, Fang Y. Individual, social and situational determinants of telecommuter productivity. Information & Management. 2005 Oct 1; 42(7):1037–49.

Park, S., & Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science, 31*(4), 567–586. https://doi.org/10.1287/mksc.1120.0718

Raghuram, S., & Wiesenfeld, B. (2004). Work-nonwork conflict and job stress among virtual

workers. Human Resource Management, 43, 259-277.

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. Journal of Personality and Social Psychology, 45, 736–750.

Scandura, T.A., Lankau, M.J., 1997. Relationships of gender, family responsibility and flexible work hours to organizational commitment and job satisfaction.

Journal of Organizational Behavior 18, 377–391.

Schreurs, B., Guenter, H., van Emmerik, I. J. H., Notelaers, G., & Schumacher, D. (2015). Pay level satisfaction and employee outcomes: The moderating effect of autonomy and support climates. *The International Journal of Human*

*Resource Management*, *26*(12), 1523–1546. https://doi.org/10.1080/09585192.2014.940992

Schonewille, M. (2001). Does training generally work?: Explaining labour productivity effects from schooling and training. International Journal of Manpower, 22, 158–173. https://doi.org/10.1108/01437720110386467

Schwalbe, M. L. (1985). Autonomy in work and self-esteem. Sociological Quarterly, 26, 519–532. Seibert, S. E., Wang, G., & Courtright, S. H. (2011). Antecedents and consequences of psychological and team empowerment in organizations: A meta-analytic review. Journal of Applied Psychology, 96, 981–1003.

Sherif, A. (2024), Topic: Work from home & remote work. Statista. Retrieved, from https://www.statista.com/topics/6565/work-from-home-and-remote-work/

Spector, P. E. (1986). Perceived control by employees: A meta-analysis of studies concerning autonomy and participation at work. Human Relations, 39, 1005–1016.

Titus, M. A. (2007). Detecting selection bias, using propensity score matching, and estimating treatment effects: An application to the private returns to a master's degree. *Research in Higher Education*, *48*(4), 487–521. https://doi.org/10.1007/s11162-006-9034-3

Yang, S., & Zheng, L. (2011). The paradox of de-coupling: A study of flexible work program and workers' productivity. Social Science Research, 40(1), 299–311. https://doi.org/10.1016/j.ssresearch.2010.04.005