Wage Penalties for Workplace Amenity: Evidence from remote work in the United States

Jacob Sano*

July 5, 2025

Abstract

Following its popularization during the COVID-19 pandemic, work from home (WFH) has become a highly desired workplace amenity. In this paper, I present an empirical estimation of the wage penalty that workers in the US face for the ability to WFH using a nationally representative dataset. Estimation results show that a worker moving from infrequent WFH to frequent or full WFH is associated with an earnings reduction of approximately 2%. This penalty is smaller than existing estimates of workers' valuation of WFH, suggesting that workers may not yet be paying the full value they place on the amenity. In further analysis, I find that there is heterogeneity in the wage penalty across industries: industries with a higher share of WFH-compatible jobs–such as "information" and "finance and insurance"–tend to exhibit lower, or even nonexistent, penalties. Additionally, I present evidence that women raising children face a significantly steeper penalty for WFH than do other groups. This suggests possible limitations in using WFH as a tool for reducing the "child penalty" and narrowing the gender wage gap.

*Graduate School of Economics, Waseda University.
Nishiwaseda 1-6-1, Shinjuku-ku, Tokyo, 169-8050, Japan. Email: jacobfs@akane.waseda.jp
I thank Munechika Katayama for research guidance. I further extend my gratitude to participants at the 3rd Tokyo-Taipei-Seoul Macroeconomics Network Workshop, participants at the WINPEC Macroeconomics Research Seminar, Shun-ichiro Bessho, and Ali Turfah for helpful comments. This research was supported by JST SPRING Grant No. JPMJSP2128 and JSPS KAKENHI Grant No. 25KJ2158.

Data availability:

All data used in this paper are publicly available. SWAA data can be accessed from https://wfhresearch.com/data/. Users of SWAA data must register their information and intended purpose of use prior to download. SIPP data can be downloaded from https://www.census.gov/programs-surveys/sipp/data/datasets.html. J2J data can be downloaded from https://j2jexplorer.ces.census.gov/. BLS CES data can be downloaded from https://data.bls.gov/series-report. See Appendix F for more information related to data access.

Disclosure statement:

Funding organizations: I gratefully acknowledge the support of JST SPRING Grant No. JP-MJSP2128 (Waseda University; Tokyo, Japan) and JSPS KAKENHI Grant No. 25KJ2158 (Japan Society for the Promotion of Science; Tokyo, Japan).

Support: I have no other sources of support to disclose.

Positions: I hold no position in any organization whose policy positions, goals, or financial interests relate to this work.

Relatives/partners: No relative or partner holds positions in any organization whose policy positions, goals, or financial interests relate to this work.

Review of the work: No other party had the right to review the paper prior to its circulation.

Institutional review board (IRB) approval: This research does not involve the collection of data on human subjects and is not subject to IRB review.

1 Introduction

For workers, amenities are an important part of total compensation. Amenity-based benefits such as the comfort of a private office, the peace of mind of a safe work environment, or the freedom of being able to work from one's own home can significantly alter the disutility that a worker will face in a job. Therefore, workers will carefully consider amenities when deciding whether to enter into (or terminate) an employment contract with a firm. Capable firms, informed on worker valuation of amenities, can use amenities to better attract and retain talented workers.

Among identical jobs, the amenities a job provides introduce across-job heterogeneity affecting worker utility. In a competitive labor market, this suggests the existence of a compensating wage differential that equalizes total compensation for identical jobs with different levels of amenities. As provided amenities increase, wage is expected to decrease, and vice versa in the presence of workplace disamenities. This paper investigates how compensating wage differentials translate to wage penalties for one particular amenity—the ability to work from home (hereinafter abbreviated as WFH).

WFH has been an amenity of particular interest in recent years due to its explosive growth during the recent COVID-19 pandemic, and to the often-divisive public discourse that surrounds it. Even after the end of the pandemic, WFH has continued to be a hot-button topic. One of the first actions of the US Trump administration after taking office was to end WFH for federal employees. WFH later became a battleground issue in the Australian national election when the leader of the Liberal Party suggested that it would pursue similar actions to Trump given the opportunity.[1] With both strong proponents and strong detractors, WFH promises to be in the public eye for the foreseeable future.

There is also evidence to suggest that WFH is impacting macro trends. Bagga et al. (2025) suggest that worker tastes for WFH are partially responsible for the unique employment trends observed in the US post-pandemic. Additionally, though industries in which a high proportion of jobs are suited to WFH tended to have higher wages prior to the pandemic, it appears that post-pandemic wage growth between industries with high potential for WFH and industries with low potential for WFH have converged. While the causes underlying industry wage growth are assuredly complex, the wage penalty for the ability to WFH may be playing a role in this trend.[2]

Of course, workplace amenity, WFH included, is a consequential topic for firms as well. An understanding of the value workers assign to particular amenities is important for firms to offer the appropriate wage necessary to attract or retain the level of worker they desire. Unfortunately, estimating worker valuation for amenities is often difficult, as pecuniary compensation and workplace amenities tend to concentrate around a small number of highly skilled workers—a phenomenon described in Rosen's "superstar effect" (Rosen, 1981). Maestas et al. (2023)

[1]The Liberal Party later abandoned this stance in the face of public outcry, but would nonetheless go on to suffer heavy losses in the election.

[2]See Appendix A for a more detailed explanation of these trends and how they were measured.

demonstrate how estimating worker valuation of amenities using cross-sectional data often yields a coefficient with the opposite sign of what is expected, as high-wage workers tend to also receive the highest level of amenities.[3] Due to this difficulty, empirical estimates of compensating wage differentials for amenities generally require rich panel data, but no such data are currently available with regard to WFH.

This paper, however, estimates worker valuation of WFH using repeated cross-sectional data from the Survey of Working Attitudes and Arrangements (Barrero et al., 2021). It overcomes the lack of longitudinal information by leveraging the extreme, near-instantaneous increase in the incidence of WFH following the COVID-19 pandemic, and by using pre-pandemic compensation as a proxy for a worker's ability to garner a high salary. The inclusion of this proxy as a control allows for an estimation of the relationship between WFH frequency and a worker's pecuniary compensation that is not confounded by the "superstar effect."

As a main result, I find that workers face a wage penalty in exchange for the ability to WFH. Moving from low WFH frequency (< 50% of working days WFH) to high WFH frequency (≥ 50% of working days WFH) is associated with an approximate 1.79% reduction in salary. Moving from low WFH frequency to full-time WFH is associated with an approximate 2.34% reduction in salary. Though not an enormous penalty—especially when viewed through the eyes of a worker who may not be able to participate in the labor force without a WFH option—it is certainly not negligible and suggests that firms that efficiently offer WFH have an advantage in attracting talented workers. Additionally, I observe that the wage penalty for WFH is not uniform across industries. Industries that have a higher share of jobs that can be done remotely, such as "information" and "finance and insurance," have a lower wage penalty for WFH, even when accounting for productivity differences between WFH and in-office work. This may suggest that some industries view WFH as more of a right, or inevitability, than a privilege. Finally, I present evidence that women raising children face a significantly steeper penalty for WFH than do other groups. WFH is undoubtedly a powerful tool in allowing greater labor force participation among mothers, but this result suggests that there are currently limitations to how much WFH can reduce the "child penalty" and the related gender wage gap.

The valuation of workplace amenities, particularly the option to WFH, has recently seen increased interest in economics literature. Barrero et al. (2021) address this issue by directly asking workers about their valuation of WFH in the Survey of Working Attitudes and Arrangements. The study finds that workers view the option to WFH 2 to 3 days per week as, on average, equivalent to 7.2% of earnings, highlighting a significant pecuniary valuation of WFH. Maestas et al. (2023) discuss the methodological challenges associated with estimating the pecuniary valuation of amenities. Through analysis of data from the American Working Condition Survey, they provide an estimate of the pecuniary value workers assign to WFH at 4.2% of salary. Mas and Pallais (2017), analyzing a sample of applicants to a call center position, find that those

[3] This effect may be at work in studies such as White (2018) and Pabilonia and Vernon (2024) which demonstrate that what appear to be wage premia for WFH can be observed in cross-sectional data.

in their sample were, on average, willing to sacrifice 8% of salary for the ability to WFH.[4] De Fraja et al. (2022) focus on workers in the UK and, using results from a novel survey, find that respondents report an average valuation of 2 days of WFH per week at approximately 8% of wages. To investigate the impact of the emergence of WFH on inequality, they set up a general equilibrium model which suggests that workers who are able to WFH will experience reduced wage growth. They test this empirically and find that workers in occupations that can be done under WFH experienced 2-7% lower wage growth in the post-pandemic period. Although their research has a slightly different focus and examines a different population, their empirical work is likely the most comparable existing research to the analysis presented in this paper.

The relationship between workplace amenities and income inequality has also been examined in earlier foundational research. Hamermesh (1999a, 1999b), for example, investigates how workplace amenities, such as shift timing and workplace safety, vary with income levels. He finds that higher-wage workers tend to enjoy greater amenities, underscoring the positive correlation between income and workplace conditions. This phenomenon suggests that traditional measures of income inequality may underestimate the true extent of inequality when workplace amenities are considered.

This paper contributes to the literature by approaching valuation of WFH from an empirical angle, employing recent survey data that includes a large sample of workers across the United States from all industries. So far, existing research has investigated valuation mainly either directly through self-reporting in survey responses, or through experiments involving choosing between hypothetical job offers. These approaches are effective in obtaining a pure valuation of the amenity, but this valuation is not necessarily equivalent to the wage penalty a worker will ultimately face when their valuation is subject to actual market forces. This empirical approach allows for measuring the "outcome" wage penalty that results from the decisions of both workers and firms.

The rest of the paper is organized as follows: Section 2 describes the main data source and data preparation. Section 3 presents main findings and heterogeneous effects across respondent characteristics. Section 4 checks result validity using various imputation methods. Section 5 concludes.

2 Data

This paper uses data from the Survey of Working Attitudes and Arrangements (SWAA). The SWAA is a repeated cross-sectional online survey that began in May 2020 and is still actively collecting data at the present time of writing. The survey mainly collects information relevant

[4]Utilizing choice between job offers (real or hypothetical) is a popular method for estimating valuation of workplace amenities. He et al. (2021) and Nagler et al. (2024) use this method to estimate WFH valuation for workers in China and Germany, respectively. Wiswall and Zafar (2017) use this method to estimate valuation of flexible working arrangements for US workers, but do not focus on WFH specifically.

to the respondents' "working arrangements" and, in particular, WFH incidence and respondent attitude toward WFH. The survey also collects demographic and income information along with a broad array of other topical questions. The survey takes a number of measures to promote the reliability of the data, including removing "speeders," implementing attention checks, and shuffling response choice display order. Respondents are restricted to working-age persons (20-64 years of age). Originally, the survey was restricted to employed persons who had an income of 10,000 USD or more in the previous calendar year, but later versions of the survey removed this restriction. For consistency, this paper will use only data that meets the original restrictions.[5]

A new version of the survey instrument is published monthly. New versions of the instrument add and delete some topical questions, but the main structure of the survey has been generally consistent from its outset, collecting data from applicants on topics such as income, demographics, job type and industry, WFH frequency, desire to WFH, and employer WFH policies. Variables from the dataset used in the main analysis are briefly described in Table 1.

In reviewing the data, I found no unexpected features or obvious outliers. Overall, the data largely reflects the trends one would expect to see in the larger population. One exception to this is the income distributions for certain regions. Mean income for respondents located in military location codes is extremely high relative to all other locations. In the opposite manner, observations from Puerto Rico have a very low mean income. I remove observations from these locations from analysis. This restriction only removes 15 observations from the dataset and does not significantly affect the results of the analysis.

I focus the analysis on non-farm economic activity by removing respondents who are employed in the "agriculture" industry. WFH likely has a different nuance in this industry than in others, as "working from home" on a farm likely means that the individual is working and living on a farm that they or their family owns. In such a case, it would not make sense for there to be a penalty for WFH, and the observation would not be relevant to this analysis. Again, this restriction does not meaningfully alter the results of the analysis.[6]

The SWAA provides weights for each observation that match the Current Population Survey (CPS) in terms of age, sex, education, and earnings. Except for where otherwise stated, analyses throughout this paper use these weights to promote results that are more representative of the US working population. This paper uses SWAA data through December 2024. For more information on the survey, refer to Barrero et al. (2021).

[5]There are some other data sources that collect information on WFH incidence, such as the Current Population Survey and the Census Household Pulse Survey. These datasets were not found suitable for this analysis due to limited length and frequency of data collection and lack of information on critical controls. At first glance, there appear to be large discrepancies between the WFH incidence reported across these datasets. However, Buckman et al. (2025) Buckman et al. show that the differences in WFH incidence across datasets are due mostly to misalignment of target populations and survey design. They find that when harmonizing the datasets, the WFH incidence reported in the SWAA aligns very closely with that reported in the Census Household Pulse Survey. The Current Population Survey seems to be an outlier in terms of reported WFH incidence, but Buckman et al. present strong evidence that this is due to idiosyncrasies in the question design.

[6]Robustness checks for the main analysis where these data restrictions are lifted are presented in Appendix C.

Table 1: Description of variables

Variable	**Description**	**Type**
Date	Year-month date in which the response was submitted.	Date
Income	Annual income of the respondent in the previous calendar year. The midpoint of the bin is taken when used in regression.	Binned
Income 2019	Annual income of the respondent in the calendar year 2019. The midpoint of the bin is taken when used in regression.	Binned
Gender	Respondent's gender (female, male, other/prefer not to respond).	Categorical
Occupation	Respondent's occupation or job type.	Categorical
Industry	Industry in which the respondent is employed.	Categorical
Age	Respondent's age.	Binned
Location	Location of the respondent's place of residence (all US states and military location codes).	Categorical
Ethnicity	Ethnicity of the respondent.	Categorical
Education	Highest level of education attained by the respondent (through PhD).	Categorical
WFH frequency	Percent of work the respondent did in the previous week in a WFH environment. Conceptually continuous in [0,100], but binned by the survey instrument.	Binned

For analysis purposes, I create an additional categorical variable using the values of *WFHfrequency* that denotes the *level* of frequency at which a respondent performs WFH. This variable, *WFHlevel*, is defined as "none" when *WFHfrequency* is 0% (meaning that the respondent works 0% of their work days under WFH), "low" when *WFHfrequency* $> 0\%$ and $< 50\%$, "high" when *WFHfrequency* $\geq 50\%$ and $\leq 97\%$, and "full" when *WFHfrequency* $> 97\%$. This allows for an intuitive interpretation of the wage differential between those who work from home frequently and those who do not, makes irrelevant some idiosyncrasies in how respondents reported their WFH frequency in the SWAA, and precludes some nonlinearity issues that may arise from using *WFHfrequency* as a continuous variable. *WFHfrequency* can, if desired, be used in analysis as a continuous variable without changing the sign of the results. Results of a robustness check employing this method are presented in Appendix C.

I additionally employ a variable available in the SWAA dataset that records the respondent's self-reported productivity change between WFH and in-office work. If respondents are significantly less productive under WFH, there is a possibility that the wage penalty for WFH is not due to the amenity itself, but rather due to the fact that WFH workers are producing less. To account for this, I exclude from analysis respondents who report that they are more than 10% less productive under WFH. Self-reported productivity can be problematic if respondents are unable to accurately assess their own productivity. However, even with some level of measurement error, respondents who experience problematically low productivity under WFH are likely to be excluded from analysis under this threshold. This variable could be used as an explicit control in the regression, but if there is significant measurement error which is correlated with *WFHfrequency*, the coefficient of interest may be biased. Therefore, I opt to restrict the sample over adding the control to regression specifications.[7]

Summary statistics for selected variables are available in Table 2. As visible in the table, *WFHfrequency* is bimodal at the extreme left and right of the distribution. Approximately three-quarters of the observations fall into the "none" or "full" levels. Due to the large number of observations and the symmetry of the distribution, this does not pose a problem for analysis. Sample distributions for other variables are available in Appendix B.

Some robustness checks in this paper additionally use data from the Survey of Income and Program Participation (SIPP) to impute income data. More information on this dataset will be presented in Section 4.

3 Analysis

This section presents methods and results for the main analysis of the wage penalty for WFH and for analysis of heterogeneous effects across respondent characteristics.

[7] There is a risk that failing to control for increases in productivity under WFH may lead to underestimation of the wage penalty for WFH. This is discussed further in Section 4. Results of a robustness check that explicitly uses self-reported productivity as a control are presented in Appendix C.

Table 2: Summary of Selected Variables

Continuous	Count	Mean	Median	Std. Dev.	Min	Max
Income	221,646	101.60	65.00	150.49	15.00	1,000.00
Income 2019	128,247	102.16	65.00	154.53	2.50	1,000.00
WFH Frequency	170,408	47.08	40.00	45.16	0.00	100.00
WFH Effectiveness	147,547	9.73	0.00	18.02	-40.00	40.00
Categorical	Count	-	-	-	-	-
WFH Level	170,408	-	-	-	-	-
none	71,200	-	-	-	-	-
low	17,443	-	-	-	-	-
high	19,255	-	-	-	-	-
full	62,510	-	-	-	-	-

Note: Income is measured in thousand USD. WFH Effectiveness reports as a percentage how much more effective the respondent is under WFH compared to working at the office. A negative value means the worker is more effective at the office than under WFH. Italics denote a level of the categorical variable *WFHlevel*.

3.1 Main Analysis

I first estimate the wage penalty that US workers face for the amenity of WFH. For this estimation, a worker's most recent wage is regressed on the worker's current level of WFH, the worker's pre-pandemic wage (as a proxy for the worker's ability to garner a high wage), and controls for demographics, occupation, industry, and education.

The SWAA has data available from May 2020, but the relationship between wage and the ability to WFH was abnormal in the early stages of the pandemic as the economy faced severe consumption and employment shocks, and lockdown policies forced companies to choose between implementing WFH or shuttering operations entirely. To obtain a more representative result, I remove from analysis data influenced strongly by the pandemic.

Figure 1 shows the seasonally-adjusted US employment rate from 2010 to 2024. As expected, there is an enormous spike in unemployment starting in early 2020 corresponding to the COVID-19 pandemic. Visually inspecting the data, we can see that the spike appears to have completely subsided in December 2021, following which the trend begins to run flat.

In addition, early in the pandemic, firms were still operating under the assumption that WFH was a temporary measure and were therefore unlikely to take measures to adjust wages in accordance with the amenity. However, as firms gained experience with WFH, they began to revise their plans to include long-term WFH implementation, as seen in Figure 2. Figure 2 shows the average actual WFH frequency for SWAA respondents (by percent of full work days done under WFH) and the average percent of days that firms planned to allow their workers to WFH looking one year into the future.[8] In 2020 there was a large discrepancy between the

[8]Percent of full work days done under WFH for a respondent is calculated as:
$\left[\left(\frac{1}{40}\right)\text{ (average weekly workhours)}\right] \times \textit{WFHfrequency}$.

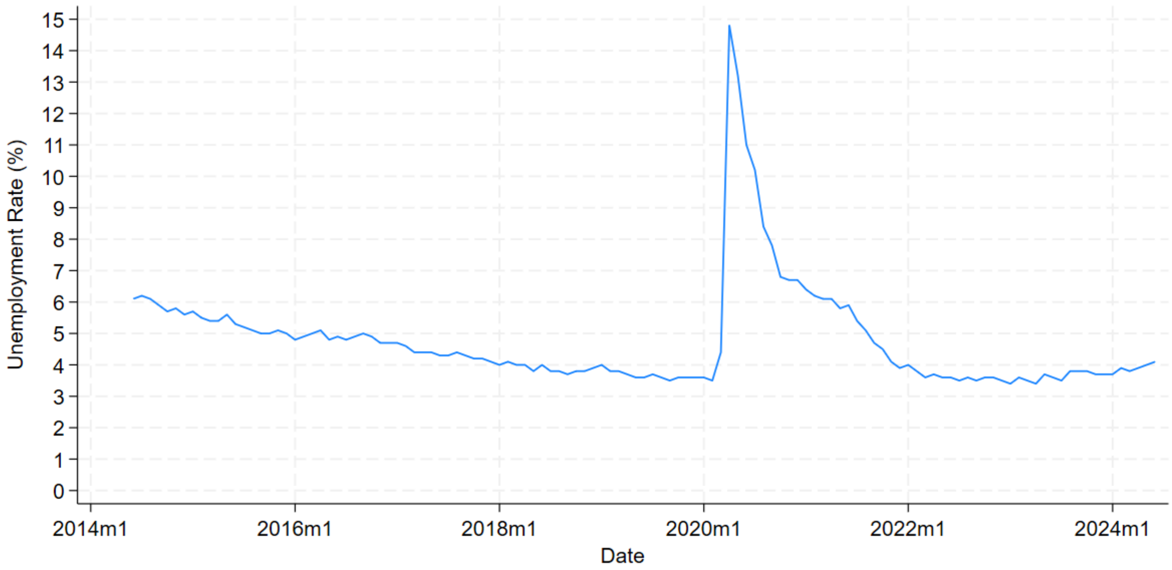

Figure 1: US Unemployment Rate (2014-2024)

Source: US Bureau of Labor Statistics

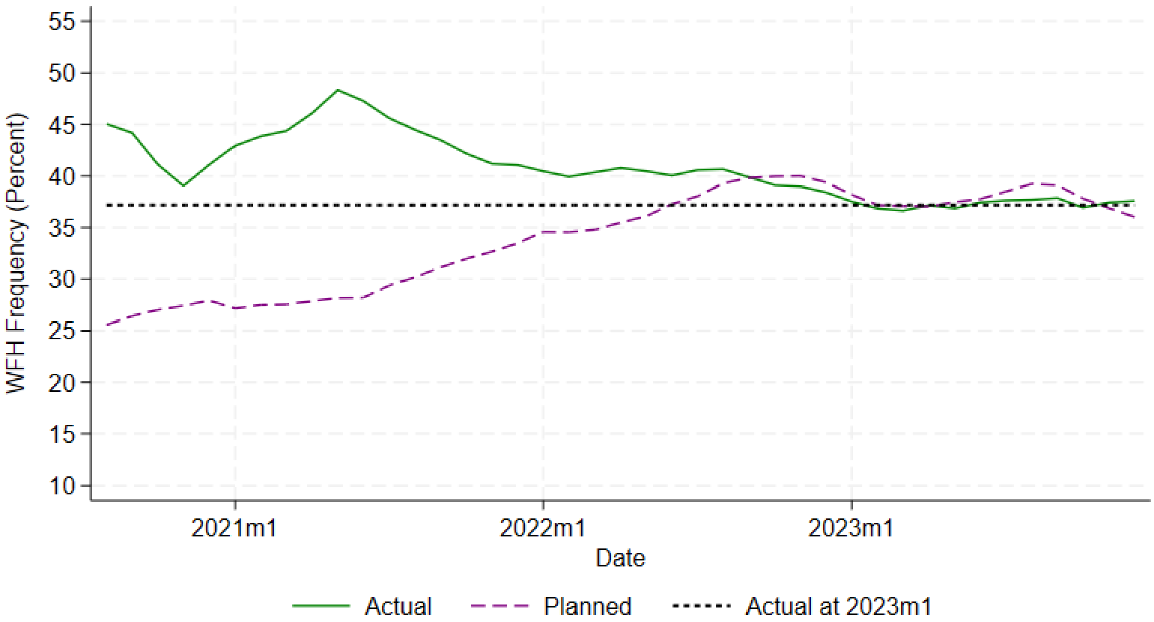

Figure 2: Average WFH Frequency (Actual vs. Planned)

Note: Data smoothed at a 6-month moving average. 'Actual' represents the average full days done WFH per week. 'Planned' represents firms' average expectation for how many full days per week their employees will be allowed to WFH one year in the future. Average WFH in January 2023 is added for reference.

measures, but firm expectations for WFH frequency increased rapidly, and by early 2022, firms had generally accurate expectations for the future of WFH.

With this information, I decide to exclude from analysis all data from before January 2022.

This should remove the majority of the pandemic-related noise from the data, while still allowing for a large enough sample size to obtain meaningful estimates. This cutoff also allows for a full three-year period of data (January 2022 - December 2024) to be used in the analysis. Having full years of data helps to remove issues that may arise from seasonality in the data, and allows for clean inter-temporal imputation (discussed in Section 4.1).[9]

For illustrative purposes, before running the main regression I estimate a specification that faces the problem of the "superstar effect" that biases upward the estimation of the coefficient of interest. This is a normal OLS regression with the following form:

$$\ln(income_i) = \alpha + \sum_{j=1}^{3} \left[\beta^{(j)} \cdot level_i^{(j)}\right] + \boldsymbol{\psi}'\mathbf{X}_i + u_i, \quad j \in \{none, high, full\}, \tag{1}$$

where j represents the levels of *WFHlevel* (with the "low" level serving as the base group) and $level^{(j)}$ is a binary variable indicating whether the observation belongs to the j-th level. $\mathbf{X}$ and $\boldsymbol{\psi}$ are vectors of control variables and corresponding coefficients, where controls include gender, occupation, industry, age, location, ethnicity, and education.[10] The results of this regression are displayed in column (1) of Table 3.

These results would tepidly suggest that those who WFH frequently (but not fully) may make slightly more than those who WFH infrequently, and that those who WFH fully make slightly less than those who WFH infrequently. Neither of these estimates are statistically significant, and this is certainly not a result that would suggest the presence of a wage penalty. We gain more meaningful results, however, upon inclusion of the *income2019* control. Results of a regression of the following form are reported in column (2) of Table 3.[11]

$$ln(income_i) = \alpha + \sum_{j=1}^{3} \left[\beta^{(j)} \cdot level_i^{(j)}\right] + \gamma \cdot ln(income2019_i) + \boldsymbol{\psi}'\mathbf{X}_i + u_i \tag{2}$$

Including the respondent's 2019 income in the regression controls for the worker's ability to "garner a high wage". Critically, it controls for this ability irrespective of the worker's choice to WFH, as WFH was only legitimized as a popular amenity at the onset of the COVID-19 pandemic in 2020. Without the control, results are confounded by the fact that those who earn more are often more able to gain the amenity of the ability to WFH. After controlling for the individual's "earning potential," the results clearly suggest the existence of a wage penalty for the ability to WFH. An increase in WFH frequency from low to high is associated with an approximate 1.79% reduction in income. An increase in WFH frequency from low to full is

[9] January 2022-December 2024 is the preferred data period, but the results are robust to changing the cutoff to July 2022 or January 2023. See Appendix C for details.

[10] The control variables are handled as categorical variables, so $\mathbf{X}$ acts a vector capturing all of the dummy variables necessary to define the categories of each control.

[11] Regressions employ heteroskedasticity robust standard errors, but not clustered standard errors. Data are not obtained using stratified sampling, and groups within the sample that we may expect to display meaningful heterogeneity in the distribution of errors (e.g., industries and occupations) are controlled for as covariates.

Table 3: Income on WFH level

	(1) Income (no control)	(2) Income (with control)
Full	−0.0171	−0.0234***
	(0.0107)	(0.00733)
High	0.0123	−0.0179*
	(0.0132)	(0.00922)
Low	–	–
None	−0.0795***	−0.0371***
	(0.0124)	(0.00884)
ln(Income_2019)		0.6760***
		(0.00826)
Observations	60,102	60,102
Adjusted R^2	0.342	0.731

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

associated with an approximate 2.34% reduction in income. Results are robust to a number of alternate specifications and data restrictions. Refer to Appendix C for details.

These results are smaller in magnitude than the findings of previous studies such as Barrero et al. (2021) (7.2% valuation on the ability to WFH 2-3 days per week), Maestas et al. (2023) (4.2% valuation on the ability to WFH), and Mas and Pallais (2017) (8% valuation on the ability to WFH). It is difficult to compare these results directly, as the studies employ different methods (for example, Barrero et al. (2021) ask respondents directly how much they would be willing to sacrifice for the ability to WFH) and each paper has slightly different definitions of the ability to WFH. It is interesting, though, that this paper, which empirically analyzes WFH incidence and wage in a broad sample of workers, reports a smaller magnitude than the other studies, which measure a more "pure" valuation. This may suggest that workers, on average, pay less of a wage penalty for the amenity of WFH than their full pecuniary valuation of the amenity.

Readers will note that in Table 3, there is a significant negative coefficient on the "none" level of *WFHlevel*, indicating that workers who do not WFH at all make less than those who WFH at a low frequency, even when accounting for "earning potential" via 2019 income. This appears to occur because the controls in the regression cannot perfectly predict which respondents are employed in jobs that are not possible to perform under WFH. Such jobs are expected to be lower-paying than jobs that can be done under WFH, and we therefore see a severe downward bias on the "low" coefficient. The dataset does not provide sufficient information to reliably separate those who do not WFH because the nature of their job renders it impossible from those who do not WFH because their firm simply does not provide the amenity, so it is not feasible to remove this bias. Therefore, analysis in this paper will focus on the relative wage penalty for those who perform a nonzero fraction of their job under WFH.

3.2 Heterogeneous Effects

The results above suggest that, on average, workers face a modest wage penalty for the amenity of WFH. However, it is unlikely that this wage penalty is the same for all workers. In this section, I investigate the heterogeneity of the wage penalty across industries and other selected sub-groups.

Industry is an interesting possible source of heterogeneity in the wage penalty for WFH as firms within the same industry are likely to be competing for the same pool of workers. Additionally, industries can take on unique cultures and norms that may affect how the amenity of WFH is perceived. To investigate the wage penalty across industries, I run a variant of the main regression for each industry listed in the SWAA dataset.

A difficulty in this exercise is that splitting the dataset by industry results in a small number of observations for each regression. To help counteract this loss in power, I reconstruct the *WFHlevel* variable to combine the "high" and "full" levels into a single "high" level. (Observations are given the category "high" if they have *WFHfrequency* $\geq 50\%$.) The "low" and "none" levels are unchanged from their previous definitions. In addition to allowing for more robust and meaningful estimation with a smaller sample size, combining the "high" and "full" levels allows for a more intuitive comparison across industries. The coefficient estimate by industry for the "high" level of *WFHlevel* is visualized in Figure 3.

Figure 3 shows that while, overall, there is a positive wage penalty for WFH, there are some industries, such as "information" and "finance and insurance," for which the wage penalty appears nonexistent. This could potentially indicate that norms in these industries have grown to accept WFH as more of a standard practice due to high incidence.

To explore this further, I use results from Dingel and Neiman (2020), which provides an index of the share of jobs in an industry that can be done from home. In Figure 3, the industries are ordered from top to bottom in ascending order of the Dingel and Neiman index.[12] When plotted, a clear positive relationship emerges, suggesting that the wage penalty for WFH is smaller in industries where a larger share of jobs can be done from home.

There are some outliers in center of the distribution in the "Mining," "Arts & Entertainment," and "Other" industries, but there are reasons to believe that these do not disrupt the overall trend. "Mining" has very few observations in the dataset (see Appendix B), leading to a very large standard error in the coefficient estimate. For "Other" and "Arts & Entertainment," it is difficult

[12]The industries in the SWAA dataset align closely with the industries in Dingel and Neiman's study, but are not identical. In cases where there may be differences in industry definitions, I use my best judgment for matching. Refer to Dingel and Neiman (2020) for more information on the Dingel and Neiman index and industry scores.

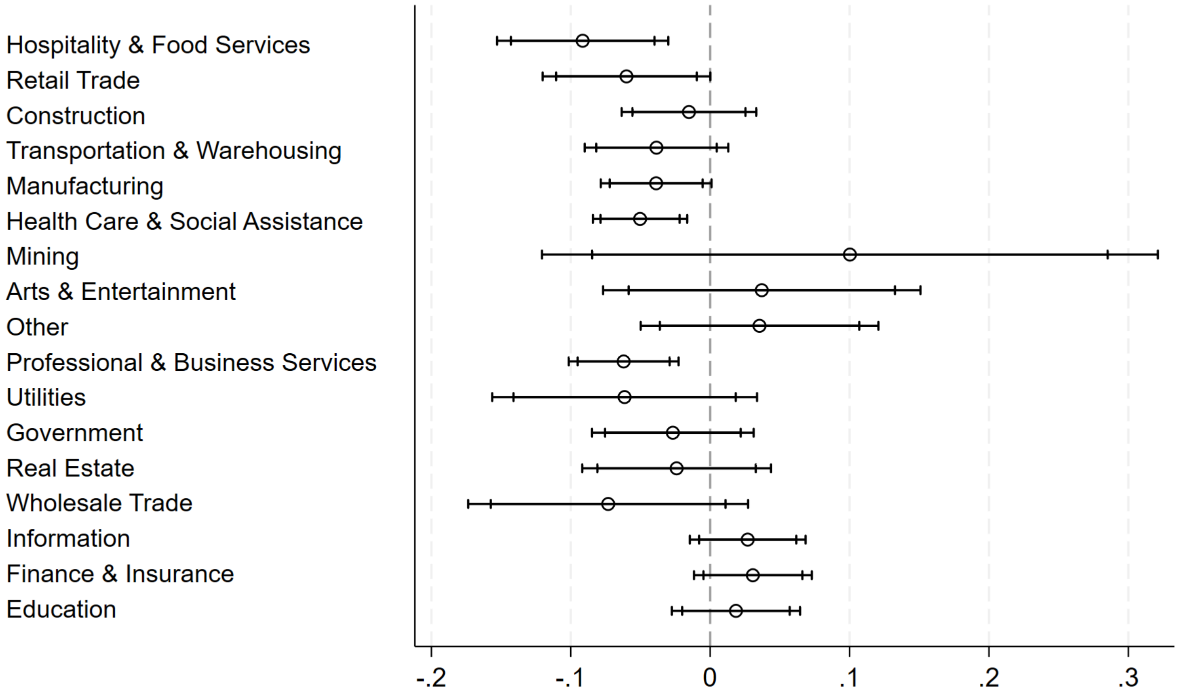

Figure 3: "High" Coefficient Estimate by Industry

Note: Spikes represent 90% and 95% confidence intervals.

to be confident in where they should truly lie in terms of share of jobs that can be done under WFH. The "Other" industry is nebulous in construction, and the "Arts & Entertainment" industry may be increasingly composed of people creating videos and other digital content for social media, which is generally done from home.

To test the significance of the trend, I run a regression on data from all industries that implements an interaction term between the Dingel and Neiman index and each level of *WFHlevel*. The regression is identical to the main regression, but with the addition of the interaction term. As the Dingel and Neiman index is continuous bounded on the interval $[0, 1]$, the interaction term represents the difference in the wage penalty for WFH between an industry with a 0% share of jobs that can be done from home and an industry with a 100% share of jobs that can be done from home. The results of this regression are presented in Panel A of Table 4. The positive and significant coefficients on the interaction terms support the hypothesis that the wage penalty for WFH is smaller in industries where a larger share of jobs can be done from home. It appears likely that for certain industries in the US, WFH is being regarded as more of a standard practice, or "inevitability," than as a privilege.[13]

Another point of interest with regard to WFH is the potential of WFH to allow people to

[13]Since the Dingel and Neiman index is a measure of the share of jobs that *can* be done from home, it does not directly point to the prevalence of WFH in an industry. However, Bartik et al. (2020) demonstrate that the Dingel and Neiman index is a very strong predictor of industry level patterns of WFH incidence, so here I do not make a distinction between the potential for WFH and its realized incidence.

Table 4: D&N Index and Gender/Family Interactions

Dependent Variable: Income	Coefficient	SE
Panel A: D&N Index Interaction		
Full	−0.0711***	(0.0137)
Full × D&N Index	0.1220***	(0.0297)
High	−0.0563***	(0.0176)
High × D&N Index	0.0979***	(0.0376)
Observations	60,102	
Adjusted R^2	0.732	
Panel B: Gender and Parental Status Interactions		
Full	−0.0526**	(0.0232)
Full × Female	0.0231	(0.0322)
Full × Has children	0.0726***	(0.0278)
Full × Female × Has children	−0.0618	(0.0386)
High	−0.0757**	(0.0308)
High × Female	0.0934**	(0.0407)
High × Has children	0.0564	(0.0358)
High × Female × Has children	−0.0930*	(0.0488)
Female	−0.0941***	(0.0302)
Female × Has children	0.0330	(0.0346)
Has children	−0.0142	(0.0253)
Observations	42,479	
Adjusted R^2	0.733	

Standard errors in parentheses.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

enter or remain in the workforce who otherwise would not be able to work. For example, those who have child care responsibilities may be unable to work in an office environment, as they must be available to tend to the needs of their children, but may be able to WFH. In such a case, WFH could be a viable tool to help combat the "child penalty" that damages the earning potential of those with child care responsibilities and, as child care responsibilities have been traditionally been borne mostly by women, contributes to the gender income gap.[14]

However, this assumes that the wage penalty for WFH is homogenous across groups. If there are discriminatory practices in the way wages are set under WFH that result in women with child care responsibilities taking a larger penalty to WFH, the effect of WFH on the child penalty and gender wage gap becomes ambiguous. To test for the presence of these kinds of differential penalties, I run a regression with full interactions between gender, *WFHlevel*, and whether the respondent is raising children. The regression is again identical to the main regression, with only the addition of the interaction terms. The dummy variable used for gender takes value 1 for female respondents, and the dummy variable for having children takes value 1 when the respondent is living with one or more children under the age of 18. Results are reported in Panel B of Table 4.

The results of this regression unfortunately show that women raising children face a larger penalty for WFH than do other groups. The coefficient on the double interaction term is negative for both the "high" and "full" levels, with the double interaction term for the "full" level being significant at a 10% level. The coefficient on the "high" level double interaction term has a p-value of 0.109. It is not clear whether this is responding to a higher valuation for WFH for those in this group, or if the increased penalty is the result of a separate mechanism, but these results suggest that, under current conditions, WFH alone will not be sufficient to fully close the wage gap facing women with child care responsibilities.

Another interesting point here is that the coefficient on the interaction between *WFHlevel* and whether the respondent is raising children is positive and significant for the "full" level. This is initially counterintuitive, but it may be the result of reverse causality. It is possible that those who are in jobs, or have access to jobs, that allow them to WFH fully without penalty are more likely to have children. At this point, there is insufficient evidence to draw any conclusions regarding this relationship, but it could be a topic of interest for future research.

4 Discussion

In this section I present analysis necessary to verify the validity of regression results in the presence of timing differences in key variables, and briefly discuss further considerations regarding the interpretation of results in this paper.

[14]The "child penalty" is a common topic in economics literature and generally refers to the divergence in wage between men and women that is driven by unequal allocations of child care responsibilities. Papers that touch on this topic include Kleven et al. (2019), Bertrand et al. (2010), and Angelov et al. (2016).

4.1 Data Timing Incongruency

The results presented in the previous section face a potential issue that arises from the details of the definition of *income* and *WFHfrequency* in the SWAA. In the survey instrument, respondents are asked to provide their *current* frequency of WFH, and their income in the *previous year*. This creates an incongruency in the timing that the two variables represent. The regressions in the previous section therefore implicitly assume that one of the following is true: (1) Income does not change from the previous year to the time at which the respondent replies to the survey. Or, (2) WFH frequency and other possibly time-varying regressors do not change from the previous year to the time at which the respondent replies to the survey.

While there is likely to be enough stability in the variables to allow the results of the previous regressions to be valid numerically, the above assumptions are clearly not true. To address this issue and verify the legitimacy of the results in the previous section, I carry out two additional investigations. First, I impute the *WFHlevel* for each respondent in the previous year to match the timing of the income data and rerun the main regression using the imputed data. Second, I use data from the Survey of Income and Program Participation (SIPP) to impute income for respondents in the year of the survey response to match the timing of the WFH frequency data and rerun the main regression using the imputed data. The results are compared with the original regression to check for consistency. The remainder of this section describes the imputation methods and the results of the imputed regressions.

4.1.1 *WFHlevel* Imputation

First, I match the timing of the WFH frequency data to the income data by imputing the *WFHlevel* for each respondent in the previous year. Since there are no other datasets with the requisite data on WFH incidence, this is accomplished within the SWAA dataset by matching a respondent in year t with the respondent that exhibits the most similar characteristics in year *t-1*. For this, I employ a random forest based matching procedure.

One difficulty in matching with the information provided in the SWAA dataset is that the variables useful for matching are all categorical, and some, such as location, have a large number of levels. This leads to a dimensionality problem that will significantly hurt the performance of distance-based matching algorithms. To address this, I use a random forest based approach, as random forests have been shown to be robust to high-dimensional data (Capitaine et al., 2021).

For each year pair, t and *t-1*, $t = \{2022, 2023, 2024\}$, I run the following process: Using the observations in year t and *t-1*, a random forest model consisting of 150 trees is fitted using age, gender, occupation, industry, ethnicity, and location. These variables are treated as categorical.[15] To each observation, an outcome variable (taking value A or B) is assigned at random. This arbitrary outcome allows for a classification approach where the partitioning structure of the forest is based solely on the structure of the independent variables.

[15] Recall that age is binned in the SWAA dataset.

After running each observation through the random forest, the terminal node assignments for each observation across all trees in the forest are obtained. For each observation in year *t*, the fraction of trees in which it shares the same terminal node with each observation in *t-1* is calculated. The observation in *t-1* that exhibits the highest proportion of shared terminal nodes is selected as the match for the observation in *t*.

WFHfrequency for each observation in *t* is subsequently imputed by transferring the corresponding *WFHfrequency* value from the matched observation in *t-1*. This process is repeated for each year pair, and the imputed *WFHfrequency* is used to create the imputed *WFHlevel* variable. The imputed *WFHlevel* variable is then used to rerun the main regression. The results of this regression are presented in column (2) of Table 5. The results are of the same sign, and of slightly larger magnitude than the results of the main regression.

A limitation of this method is that imputing *WFHlevel* to be used in rerunning the regression results in a generated regressor. Because there is uncertainty in the process of generating the imputed variable, the standard errors of the regression are not valid for inference. This is a common problem with imputed data, and can be overcome by calculating bootstrapped standard errors that include the variable generating process in the bootstrap. However, due to the computational demands of calculating proximities between observations in the random forest, bootstrapping is not feasible for this analysis.

In order to obtain valid standard errors, I use a second method to impute *WFHlevel* for the previous year. This method creates a proximity score using a probit model, and is a variant of the frequently used propensity score matching method. Details of this method are presented in Appendix E. The results of the regression using this method are presented in column (3) of Table 5. This method suffers more from the dimensionality of the data, and, therefore, likely performs worse in matching relative to the random forest method. However, it is quicker to compute and allows for bootstrapping. The standard errors reported have been bootstrapped using 500 iterations. The results of this regression are of the same signs and of similar magnitude to the results of the main regression. The coefficient on the "full" level is significant, but the coefficient on the "high" level is not significant at any reasonable alpha.

4.1.2 *Income* Imputation

Next, I match the timing of the income data to the WFH frequency data by imputing the *income* of each respondent in the year the respondent completed the survey. Since there are major longitudinal datasets that contain income information, I use an auxiliary dataset to impute the income data. The Survey of Income and Program Participation (SIPP) is a longitudinal survey conducted by the US Census Bureau that tracks individuals within a household over several years, collecting data on income, employment, and other topics. This provides a rich source of data that provides insight into how SWAA respondents' income is likely to change over the course of a year.

Table 5: Regression Results with Imputed Values

	(1) Main Regression	(2) Imputed WFH (RF)	(3) Imputed WFH (score)	(4) Imputed Income
Full	-0.0234^{***}	$-0.0285^{\dagger}$	-0.0301^{**}	-0.0189^{**}
	(0.00733)	(0.00793)	(0.0127)	(0.00886)
High	-0.0179^{*}	$-0.0329^{\dagger}$	−0.0177	−0.0154
	0.00922	0.0104	0.0145	0.0112
Low	–	–	–	–
None	-0.0371^{***}	$-0.0566^{\dagger}$	-0.0294^{**}	-0.0318^{***}
	(0.00884)	(0.00828)	(0.0136)	(0.0108)
Base Observations‡	60,102			
Adjusted R^2	0.731	0.713	0.732	–

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
† Standard errors are not valid for inference.
‡ Additional observation sets are generated in imputation.

A number of steps are taken to clean the SIPP data to allow for matching to the SWAA data. First, relevant variables are regrouped to match the granularity of the SWAA data. In general, variables in the SIPP dataset have more detailed levels than those in the SWAA, so data matching mostly takes the form of aggregating the SIPP data to the level of the SWAA. Education, ethnicity, industry, and occupation are manually regrouped to match the SWAA data.[16] Age is binned to match the SWAA data. Location is matched by state. To match the SWAA population of interest, observations for individuals less than 20 years old or 65 years or older, and those who report an initial income that (when scaled up to yearly income) does not meet the SWAA income threshold are removed. SIPP income data are rescaled to the SWAA income level (counted in thousand USD) by dividing by 1,000.

There exists some seasonality in the SIPP monthly income data which is likely due to the timing of bonus payments and other irregular income sources.[17] To adjust for seasonality while keeping the yearlong trend of income growing over time, the monthly income is regressed on month, where month is a continuous measure of time, and month dummies. The residuals from this regression represent the non-seasonal variation around the time trend of the monthly income data. This non-seasonal variance plus the time trend gives the seasonally adjusted monthly income data. These seasonally adjusted data are used in the imputation process.

To facilitate the observation of change in an individual's income across years, only indi-

[16] SIPP data are based on four digit industry and occupation codes. SWAA data have custom lists. Regrouping the SIPP industries and occupations to match the SWAA data is mostly straightforward. Where there is ambiguity, I use my best judgment.

[17] Details on the income seasonality present in the SIPP data can be found in Appendix D

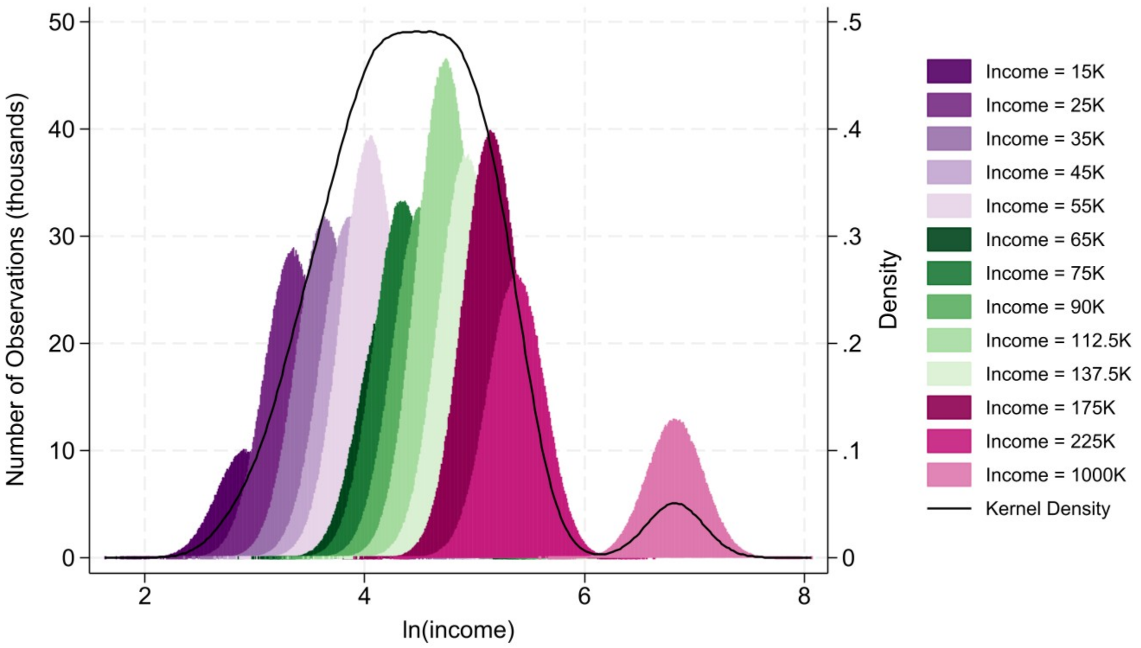

Figure 4: Predictive Distribution for Imputed Income

viduals who have multiple full years of data are kept. Furthermore, observations in any year for which the individual does not have 12 complete months of data are discarded. The SIPP data are then collapsed so that each observation represents an individual-year pair. For any observation for which there is an observation in the previous year for the same individual, a variable is created which represents the individual's income in the previous year.

Once the data are prepared, a regression is run on the SIPP data of the following specification:

$$ln(income_i) = \alpha + \beta \cdot ln(prevyrincome_i) + \gamma \cdot ln(prevyrincome_i)^2 + \delta \cdot ln(prevyrincome_i)^3 + \boldsymbol{\psi}'\mathbf{X}_i + u_i, \quad (3)$$

where $\mathbf{X}$ is a vector of dummies for the categorical variables gender, ethnicity, education, industry, occupation, age (binned), region, and year, and $\boldsymbol{\psi}$ is a vector of corresponding coefficients. *prevyrincome* is the respondent's income in the previous year. The results of this regression become the basis for the imputation of income in the SWAA data.

To calculate inference-valid standard errors in the downstream regression, multiple imputation is employed. With the posterior predictive distribution defined by estimating the above regression, 200 draws are made from the results for the imputed value of *income* for each observation. Figure 4 shows the distribution of the drawn values broken down by income level, overlaid with a density plot of the overall distribution. Visual inspection suggests that the generating distribution is mixed normal, with a largely normal overall shape driven by well-behaved distributions centered around each midpoint of the binned income variable.

The main regression is run with the imputed draws for *income* as the dependent variable. To obtain the coefficient point estimates and standard errors, the output is collapsed according to Rubin's rules (Rubin and Schenker, 1986). The results of this regression are reported in column (4) in Table 5.

The results of this regression again closely match those obtained in the main regression, suggesting that the results in Table 3 are valid despite the timing incongruency between the income and WFH frequency data. Since these results were generated through the multiple imputation process, the standard errors are valid for inference. The "full" coefficient is significant at a 5% level, and the "high" coefficient has a p-value of 0.167, which is reasonable given the additional uncertainty created in the imputation process. Overall, imputing the data in both directions and rerunning the main regression yields results that are generally consistent with those obtained by running the main regression on the original data. This suggests that there is indeed sufficient stability in the measures of *income* and *WFHfrequency* to make the results of the main regression valid.

4.2 Further Considerations

There are a few important considerations in interpreting the results of this analysis. First, I do not introduce a model as the basis of analysis. That several controls must be included in the regression to obtain a meaningful estimate of the wage penalty for WFH suggests that wage and WFH frequency are jointly endogenous to some other factors, but there is no model to make explicit the nature of this endogeneity or to suggest a functional form for the regression. The choice to forgo a model-based approach in favor of more purely empirical methods was made due to data availability. WFH frequency and wage as a function of WFH amenity depend critically on both worker choices and firm choices. While the SWAA dataset provides a rich source of data on worker choices, there is very little information currently available on how firms make choices regarding WFH.[18] Given the infeasibility of accounting for the firm side of the relationship, I believe that the empirical approach presented here is most appropriate.

A related area that warrants further exploration is the mechanism through which WFH is associated with wage penalties. Wage penalties for an amenity can be conceived of most simply as a compensating wage differential, but it is possible that the penalty, or part of the penalty, is manifesting through different channels. Such a channel could be as simple as a pervasive cultural bias, or unwillingness to pay people who WFH frequently as much as those who WFH rarely. Another possible channel could be related to across-industry worker migration. For example, if many people decide simultaneously that they want to WFH, and decide to look for a WFH position in a certain WFH-friendly industry, the influx of job seekers may lead to a decrease in the wages for the job seekers once they match in the new industry. Assuming these

[18]The benchmark model of amenity determination described in Mas (2025) relies on the firm's choice in minimizing expenditure while getting workers to their target utility. However, it is currently unclear what costs (if any) firms incur by providing the WFH amenity. Without a better understanding of this, it is difficult to apply standard models to the WFH amenity.

job seekers have a higher propensity to WFH, this would lead to a negative correlation between WFH frequency and wage (conditional on the relevant controls). This sort of phenomenon can be conceptualized as a form of "wage penalty," but is distinct from the classical idea of a compensating wage differential.

It is worth noting that this kind of migration does not seem to be present in the data. Figure 5 presents the job transition matrix between industries created using 2022 and 2023 data from the US Census Bureau's Job-to-Job Flows (J2J) dataset. The industries in the matrix are ordered ascending by the Dingel and Neiman index,[19] so the elements to the left of (below) the diagonal represent job transitions from industries with a high share of jobs that can be done from home to industries with a low share of jobs that can be done from home, and elements to the right (above) the diagonal represent job transitions from industries with a low share of jobs that can be done from home to industries with a high share of jobs that can be done from home. The values on the bottom-left of the matrix tending to be slightly larger than those on the top-right suggests that there are some frictions in moving from lower WFH industries to higher WFH industries. This does not come as a particular surprise, as the Dingel and Neiman index appears to correlate positively with the skill level of the industries.

There are also reasons to believe that the estimates of the wage penalty for WFH in this paper represent lower bounds of the actual magnitudes. One reason is that the most important control in the regression is the respondent's income directly prior to the COVID-19 pandemic. This works well as a control for the respondent's "ability to garner a high wage" under non-WFH conditions, but it is likely that for some respondents this ability has changed slightly between 2019 and the time of response. In a way related to the previous discussion of the "superstar effect," those who enjoyed an increase in potential wage are, due to diminishing returns to income, more likely to sacrifice some of the potential income and choose non-pecuniary amenities. This will tend to bias upward (toward zero) the estimate of the relationship between WFH and income.

Additionally, proximity to the COVID-19 pandemic may be reducing the magnitude of the coefficient estimates relative to a "long run" measure of the wage penalty. Restricting the data period used in analysis removes the period during which the labor market was being directly affected by the pandemic, but there are other indirect ways in which the pandemic may be affecting the estimates. For example, the process through which WFH leads to reduced income may require an extended period of time to be fully realized. If a worker who remained in their job switched from in-office work to WFH because of the pandemic, it is very unlikely that the company would immediately reduce the worker's wages as a result. Instead, the wage penalty would more likely manifest as a reduction in the worker's wage growth relative to colleagues working in the office. It is also possible that there are some companies in the sample which were allowing workers to WFH, but simultaneously planning to switch back to in-office work in short order. Companies with such expectations may have believed it unnecessary to consider reducing

[19] Industries available in Dingel and Neiman (2020) and those in the J2J dataset align closely but are not identical. In cases where the industries do not match perfectly, I use my best judgment in assigning an index to the J2J industry.

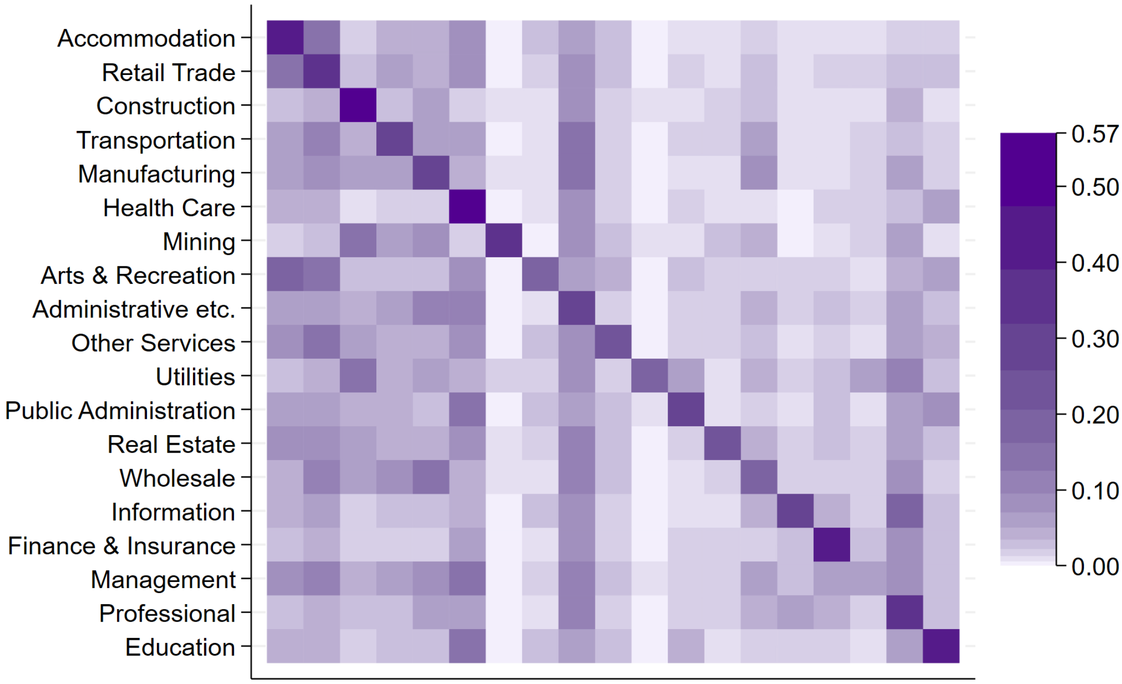

Figure 5: Job Transition Matrix

Note: Industries assigned to rows are the origin industry, ordered ascending by proportion of jobs within the industry that can be done WFH. The columns represent the destination industry and are the same industries ordered identically to the rows, so the diagonal elements are job changes within the same industry, elements below the diagonal are job transitions to industries with a smaller proportion of jobs that can be done WFH, and elements above the diagonal are job transitions to industries with a higher proportion of jobs that can be done WFH. Elements represent the proportion of workers transitioning from the origin industry who transition to the destination industry. Cutoffs for the color scale are defined by $\text{cut}_i = \min \times \left(\frac{\max}{\min}\right)^{\left(\frac{i}{15}\right)^{0.4}}$ for $i = 0, 1, \ldots, 15$, where *min* is the minimum value in the matrix and *max* is the maximum value in the matrix.

wages for workers under WFH as the arrangement was perceived as temporary. Both scenarios point discussed here point to the possibility of an underestimation of the wage penalty's true magnitude. However, I do not explore the existence or extent of such effects in this paper.

5 Conclusion

This paper uses data from the SWAA, a large repeated cross-sectional survey of US workers, to investigate the existence and magnitude of a wage penalty for the amenity of working from home. The results suggest that workers face a wage penalty for the ability to WFH, with the penalty being approximately 1.79% for workers to move from infrequent WFH to frequent WFH, and 2.34% for workers to move from infrequent WFH to full WFH.

The empirical approach used in this study offers a unique contribution to the literature as previous studies have tended to measure valuation of WFH through either self-reporting in survey responses, or through experimental methods. Such methods offer a picture of "pure" valuation of WFH, but this valuation may not be the same as the wage penalty that workers ultimately face for the amenity in the labor market. Empirical estimates in this paper reflect the actual wage penalty derived from worker valuation of the amenity meeting market forces. That the estimates in this paper are smaller in magnitude than those in previous literature may suggest that workers are not currently paying their full valuation for the ability to WFH.

Furthermore, the magnitude of the wage penalty for WFH appears to be heterogeneous across industries. When accounting for the proportion of jobs in an industry that can be done from home, the wage penalty for WFH is much smaller in industries in which it is relatively easy to WFH. In fact, in some industries where it is common to be able to WFH, the wage penalty for WFH is essentially nonexistent.

To assess how the penalty varies by gender and child raising status, an investigation is included which uses interaction terms between each WFH level indicator, gender, and whether the respondent is raising children. The estimation results demonstrate that women raising children face a larger penalty for WFH than other groups, suggesting that under current conditions, WFH alone cannot fully close the wage gap facing working mothers.

The analysis is subject to a possible issue in the timing of the data, as income is reported for the previous year while frequency of WFH is reported for the current year. To address this issue, the paper imputes the WFH level for the previous year using random forest based matching, and imputes the income for the current year using data from the Survey of Income and Program Participation. The results of these imputations suggest that the regression results are numerically valid despite the timing incongruency.

Overall, the findings of this paper suggest that workers have a positive pecuniary valuation of the amenity of WFH and that, on average, they pay a wage penalty to obtain the amenity. This is an important finding for firms competing for labor, as it suggests that firms that can offer WFH at relatively lower costs will have a competitive advantage in the labor market. As we move further into the post-pandemic world, it will be interesting to see how the wage penalty for WFH evolves over time as norms and attitudes towards WFH change. As more data become available, it will be possible to investigate this question further.

References

Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, 34(3):545–579.

Bagga, S., Mann, L. F., Sahin, A., and Violante, G. L. (2025). Job amenity shocks and labor reallocation. Working Paper 33787, National Bureau of Economic Research.

Barrero, J. M., Bloom, N., and Davis, S. J. (2021). Why working from home will stick. Working Paper 28731, National Bureau of Economic Research.

Bartik, A. W., Cullen, Z. B., Glaeser, E. L., Luca, M., and Stanton, C. T. (2020). What jobs are being done at home during the covid-19 crisis? evidence from firm-level surveys. Working Paper 27422, National Bureau of Economic Research.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–55.

Buckman, S. R., Barrero, J. M., Bloom, N., and Davis, S. J. (2025). Measuring work from home. Working Paper 32508, National Bureau of Economic Research.

Capitaine, L., Genuer, R., and Thiébaut, R. (2021). Random forests for high-dimensional longitudinal data. *Statistical Methods in Medical Research*, 30(1):166–184. PMID: 32772626.

De Fraja, G., Matheson, J., Mizen, P., Rockey, J., Taneja, S., and Thwaites, G. (2022). Remote work and compensation inequality. Available from SSRN: https://ssrn.com/abstract=4962603 or http://dx.doi.org/10.2139/ssrn.4962603.

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

Hamermesh, D. S. (1999a). Changing inequality in markets for workplace amenities*. *The Quarterly Journal of Economics*, 114(4):1085–1123.

Hamermesh, D. S. (1999b). The timing of work over time. *The Economic Journal*, 109(452):37–66.

He, H., Neumark, D., and Weng, Q. (2021). Do workers value flexible jobs? a field experiment. *Journal of Labor Economics*, 39(3):709–719.

Jann, B. (2019). heatplot: Stata module to create heat plots and hexagon plots. Available from http://ideas.repec.org/c/boc/bocode/s458598.html.

Kleven, H., Landais, C., and Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, 11(4):181–209.

Leuven, E. and Sianesi, B. (2003). psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Available from http://ideas.repec.org/c/boc/bocode/s432001.html.

Maestas, N., Mullen, K. J., Powell, D., von Wachter, T., and Wenger, J. B. (2023). The value of working conditions in the united states and the implications for the structure of wages. *American Economic Review*, 113(7):2007–47.

Mas, A. (2025). Non-wage amenities. Working Paper 33643, National Bureau of Economic Research.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–59.

Nagler, M., Rincke, J., and Winkler, E. (2024). Working from home, commuting, and gender. *Journal of Population Economics*, 37.

Pabilonia, S. W. and Vernon, V. (2024). Remote work, wages, and hours worked in the united states. Working Paper 565, U.S. Bureau of Labor Statistics.

Rosen, S. (1981). The economics of superstars. *The American Economic Review*, 71(5):845–858.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, 81(394):366–374.

White, D. (2018). Agency theory and work from home. *LABOUR*, 33.

Wiswall, M. and Zafar, B. (2017). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507.

Wright, M. N. and Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in c++ and r. *Journal of Statistical Software*, 77(1):1–17.

Appendix A Industry Wage Trends

The brief investigation into cross-industry wage trends presented in the introduction uses data on average hourly wage by industry from the Bureau of Labor Statistics' Current Employment Statistics Survey. The measure of share of jobs in the industry that can be done remotely is taken from the index provided in Dingel and Neiman (2020). The Dingel and Neiman index is mapped to industries at the 2-digit NAICS level. Due to a lack of data availability, government services, education services, and agriculture are excluded from the analysis.

For the investigation, I first run a simple linear regression of the average wage in 2019 on the mapped value of the Dingel and Neiman index. The regression returns a coefficient of 27.30 with a standard error of 5.37. Since the Dingel and Neiman index is bounded on the interval $[0, 1]$, this suggests that for each 1% increase in the share of jobs that can be done remotely, the average wage in that industry increases by approximately 0.27 USD.

Next, I regress the average annual wage growth rate in an industry on the mapped value of the Dingel and Neiman index. Only data from 2020 onward (after the onset of the COVID-19 pandemic and WFH boom) are included in the regression. The regression returns a coefficient of -1.47 with a standard error of 0.68 ($p = 0.034$). These results suggest convergence of wages between industries with a high share of jobs that can be done remotely and those with a low share of jobs that can be done remotely since the start of the COVID-19 pandemic.

Appendix B Summary Statistics

Summary statistics for important variables from the SWAA are presented in this section. Variables that are handled as categorical are presented as histograms. Variables that are handled as continuous, as well as *WFHlevel*, are presented in Table 2 in Section 2.

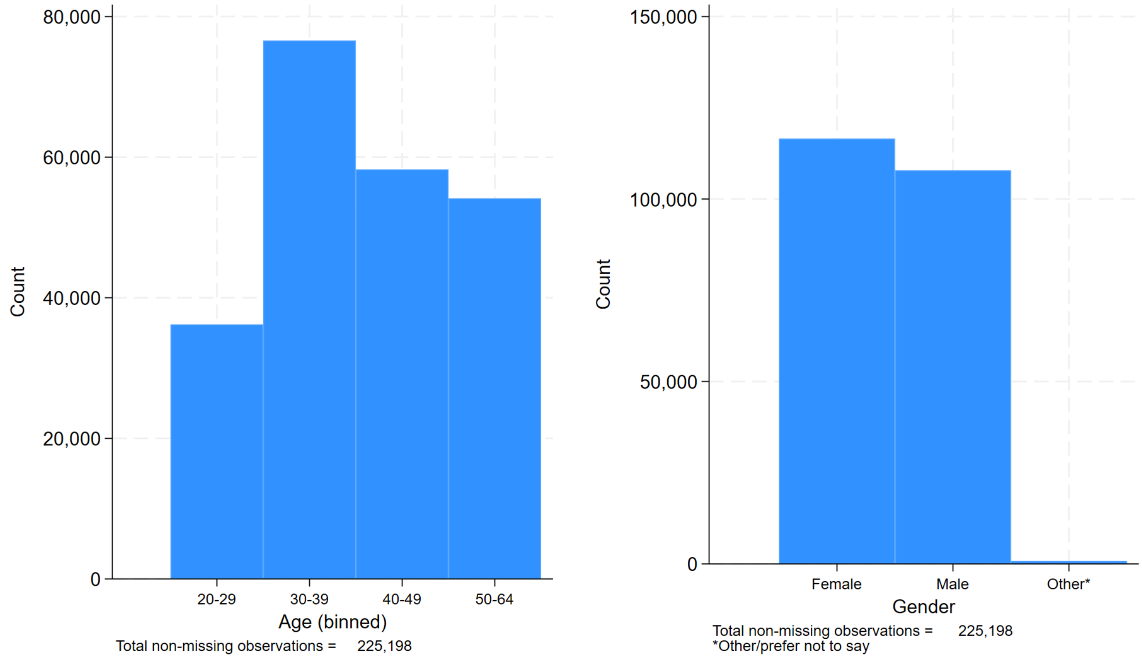

Figure 6: Age and Gender

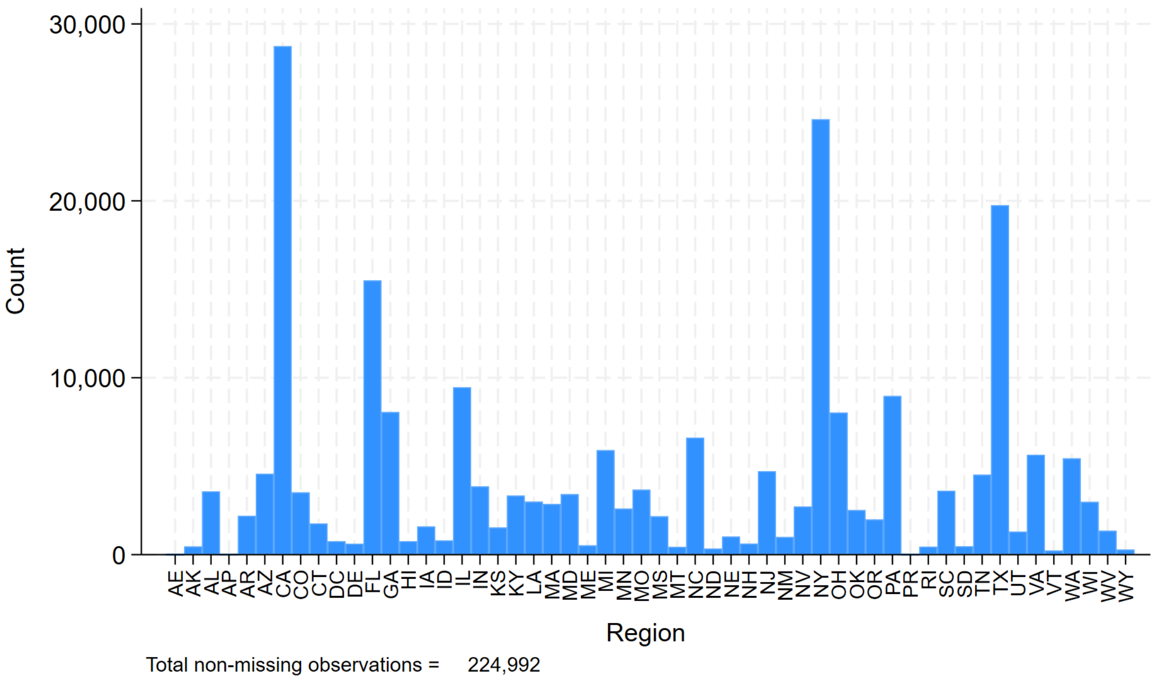

Figure 7: Region

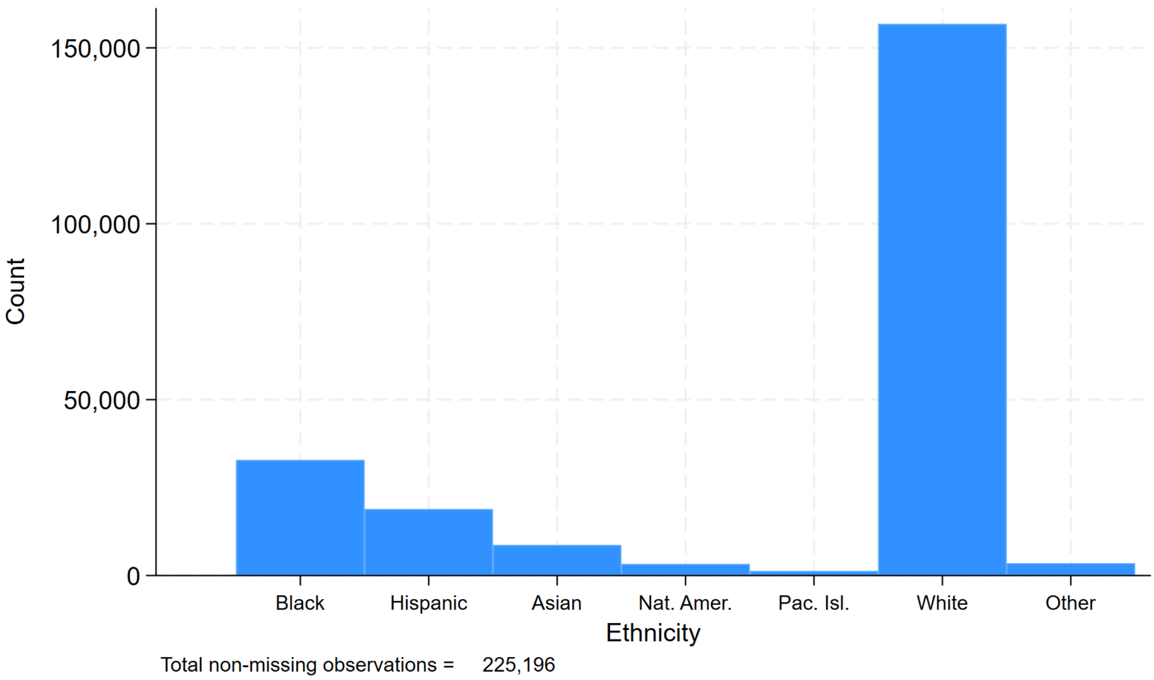

Figure 8: Ethnicity

Note:
Some labels have been shortened for presentation. Full category names are as follows:
Black = Black or African American
Hispanic = Hispanic (of any race)
Nat. Amer. = Native American or Alaska Native
Pac. Isl. = Native Hawaiian or Pacific Islander
White = White (non-Hispanic)

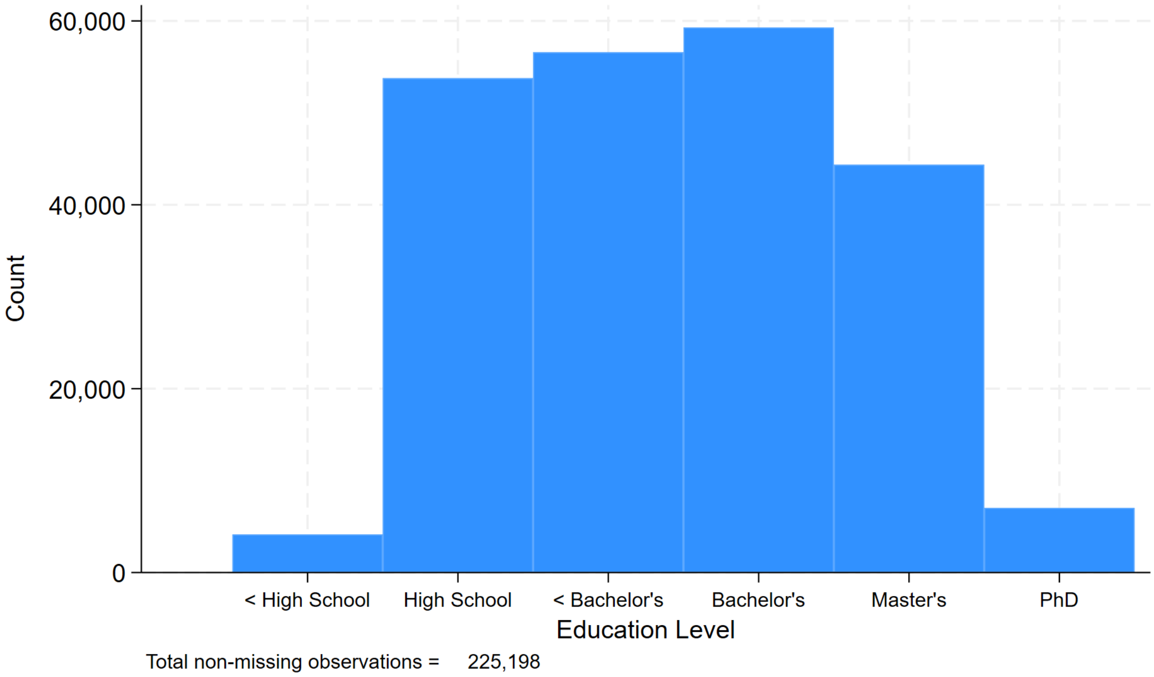

Figure 9: Education

Note:
Some labels have been shortened for presentation. Full category names are as follows:
< High School = Less than high school graduation
High School = High school graduation
< Bachelor’s = 1 to 3 years of college
Bachelor’s = 4 years of college degree
Master’s = Master’s or Professional degree
PhD = PhD

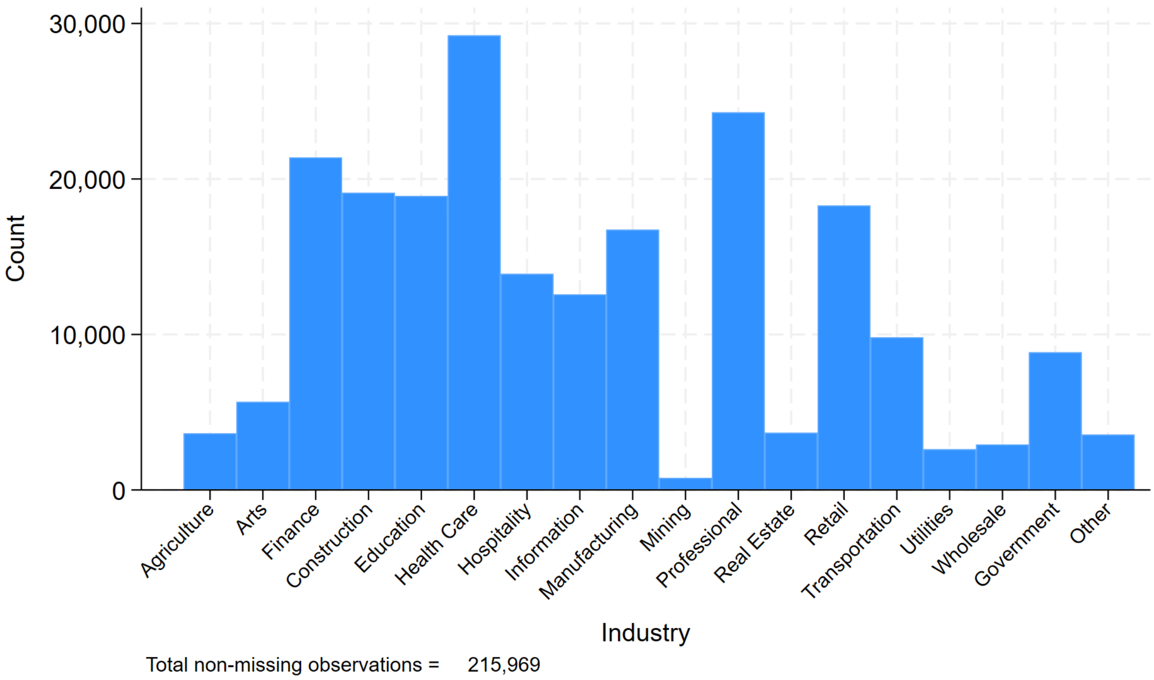

Figure 10: Industry

Note:
Some labels have been shortened for presentation. Full category names are as follows:
Arts = Arts and Entertainment
Finance = Finance and Insurance
Health Care = Health Care and Social Assistance
Hospitality = Hospitality and Food Services
Professional = Professional and Business Services
Retail = Retail Trade
Transportation = Transportation and Warehousing
Wholesale = Wholesale Trade

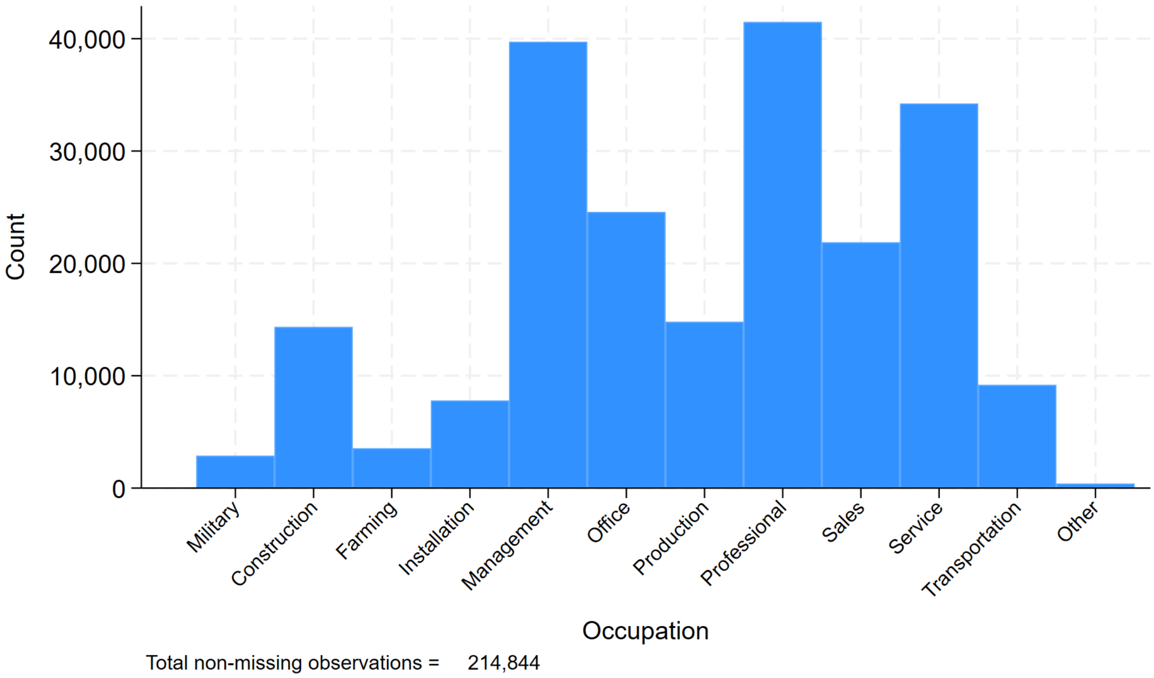

Figure 11: Occupation

Note:
Some labels have been shortened for presentation. Full category names are as follows:
Military = Armed Forces
Construction = Construction and Extraction
Farming = Farming, Fishing, and Forestry
Installation = Installation, Maintenance, and Repair
Management = Management, Business, and Financial
Office = Office and Administrative Support
Professional = Professional and Related
Sales = Sales and Related
Transportation = Transportation and Material Moving

Appendix C Robustness Checks

This appendix presents robustness checks for the main regression results presented in Table 3 in Section 3. Table 6 displays results for robustness checks dealing with alternative data restrictions and controls.

Column (1) shows results when data for all regions (including military codes and Puerto Rico) are included in the regression. Column (2) shows results when all industries (including "agriculture") are included in the regression. Column (3) shows results when all levels of WFH efficiency are included in the regression, and WFH efficiency is included as a control in the specification. Column (4) shows results when only individuals working full time (defined as 35+ hours per week) are included in the regression. Results from all of these robustness checks align closely with the original regression results.

Table 6: Robustness Checks

	(1) Income (all regions)	(2) Income (all industries)	(3) Income (efficiency control)	(4) Income (35+ hours)
Full	-0.0234^{***}	-0.0216^{***}	-0.0252^{***}	-0.0318^{***}
	(0.00733)	(0.00744)	(0.00742)	(0.00821)
High	-0.0179^{*}	-0.0163^{*}	-0.0205^{**}	-0.0228^{**}
	(0.00922)	(0.00936)	(0.00930)	(0.0113)
Low	–	–	–	–
None	-0.0371^{***}	-0.0345^{***}	-0.0319^{***}	-0.0477^{***}
	(0.00884)	(0.00896)	(0.00855)	(0.00957)
ln(Income_2019)	0.676^{***}	0.674^{***}	0.670^{***}	0.676^{***}
	(0.00826)	(0.00813)	(0.00809)	(0.0113)
Wfh_efficiency			0.000398^{***}	
			(0.000143)	
Observations	60,103	61,168	65,980	32,843
Adjusted R^2	0.731	0.729	0.725	0.737

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7 displays results for robustness checks using alternate forms of the WFH variable. Column (1) shows results when the WFH variable has an additional "medium" category. For this specification, the *WFHlevel* variable is defined as follows. "None" is defined as *WFHfrequency* of 0%, "low" is defined as *WFHfrequency* in (0%, 25%), "medium" is defined as *WFHfrequency* in [25%, 65%), "high" is defined as *WFHfrequency* in [65%, 97%], and "full" is defined as

Table 7: Robustness Checks (cont.)

	(1) Income (five levels)	(2) Income (excl. none)	(3) Income (two levels)	(4) Income (alt levels)	(5) Income (continuous)
Full	-0.0362^{***}	-0.0258^{***}	–	-0.0329^{***}	–
	(0.0119)	(0.00740)		(0.00977)	
High	-0.0233	-0.0197^{**}	-0.0244^{***}	-0.0242^{**}	–
	(0.0150)	(0.00922)	(0.00717)	(0.0105)	
Medium	-0.0277^{**}	–	–	–	–
	(0.0127)				
Low	–	–	–	–	–
None	-0.0488^{***}	–	–	-0.0454^{***}	–
	(0.0129)			(0.0110)	
WFH_frequency	–	–	–	–	-0.000294^{***}
					(0.000101)
Observations	60,249	50,019	50,019	60,249	50,019
Adjusted R^2	0.731	0.738	0.738	0.731	0.738

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

WFHfrequency in (97%, 100%]. Column (2) gives results when *WFHlevel* is defined as in the original regression, but observations in the "none" category are excluded from the regression. Column (3) gives results when the *WFHlevel* variable is split into only two levels, "high," defined as *WFHfrequency* in [50%, 100%], and "low," defined as *WFHfrequency* in (0%, 50%) (those with 0% *WFHfrequency* are not included in the regression). Column (4) gives results with the same levels of *WFHlevel* as in the original regression, but with different cutoffs. In this specification, "none" is defined as *WFHfrequency* in [0%, 15%), "low" is defined as *WFHfrequency* in [15%, 30%), "high" is defined as *WFHfrequency* in [30%, 85%], and "full" is defined as *WFHfrequency* in (85%, 100%]. Column (5) displays results when the continuous *WFHfrequency* variable is used in place of the categorical *WFHlevel* variable.

The results in both Table 6 and Table 7 are consistent with the main regression results, suggesting that the results are robust to a number of different specifications.

Results presented in Table 8 demonstrate that the main regression results are also robust to changes in the data period. Column (1) shows results when the data period starts in July 2022. Column (2) shows results when the data period starts in January 2023. Restricting the data period to these later start dates slightly increases the magnitudes of the coefficients of interest, but there is otherwise no real change in the results.

Table 8: Robustness Checks (cont.)

	(1) Income (Jul 2022-)	(2) Income (Jan 2023-)
Full	-0.0260^{***}	-0.0247^{***}
	(0.00787)	(0.00932)
High	-0.0183^{*}	-0.0212^{*}
	(0.00987)	(0.0118)
Low	–	–
None	-0.0409^{***}	-0.0393^{***}
	(0.00946)	(0.0109)
ln(Income_2019)	0.6571^{***}	0.6410^{***}
	(0.00859)	(0.00960)
Observations	55,430	44,892
Adjusted R^2	0.716	0.709

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D SIPP Income Seasonality

Figure 12 shows the income distribution by month in the SIPP dataset. January and December are the months with the highest income, potentialy reflecting the timing of bonuses or increased working hours. February is the month with the lowest income, possibly becuase it is shorter than the other months.

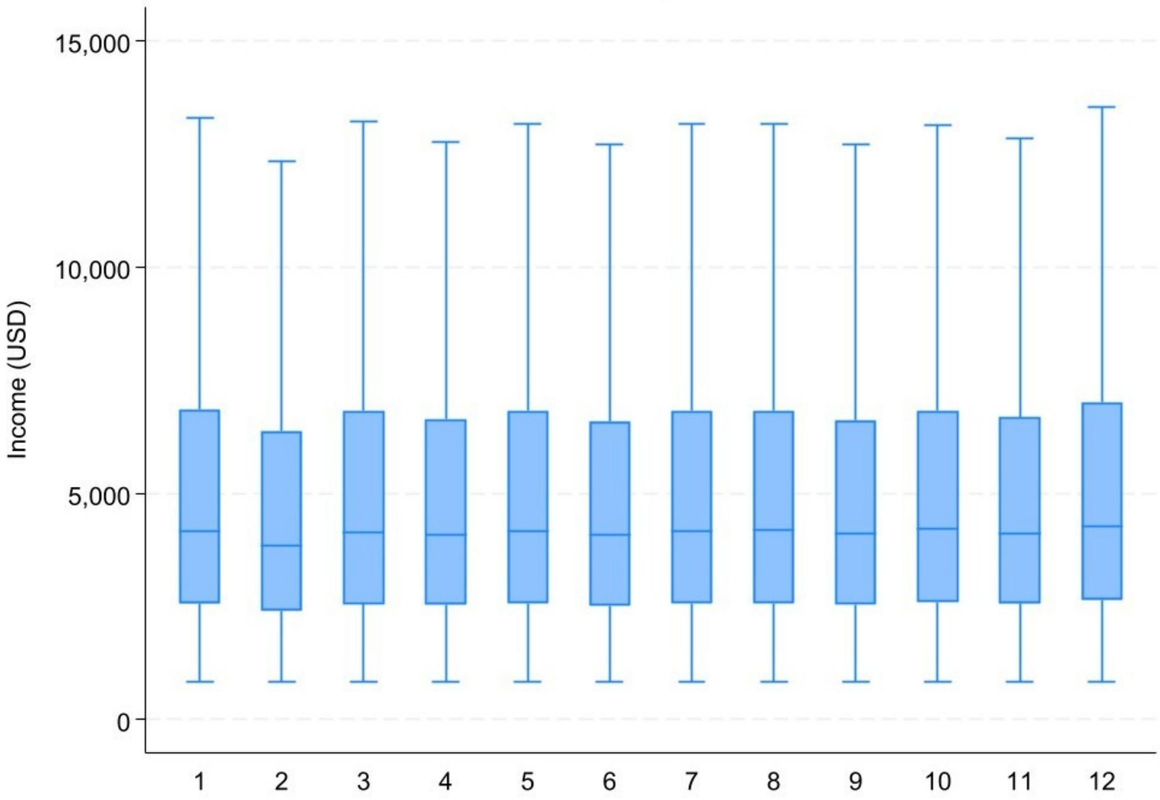

Figure 12: Income by Month in SIPP Data

Note: Boxplots show the 10th, 25th, 50th, 75th, and 90th percentiles of the income distribution for each month. Outliers are not displayed.

Appendix E Score Matching

The score matching process used in this paper is closely related to the propensity score matching process (Rosenbaum and Rubin, 1983). The main difference is in the outcome variable used to calculate the score. In classical propensity score matching, the outcome variable is a binary variable indicating whether the observation is in the treatment group or not. This yields a score that represents the propensity of an observation being selected into the treatment group, and observations are then matched across groups based on this score. This paper uses a binary variable indicating whether the observation's *WFHfrequency* is greater than 50%. Calculating a score based on this variable yields a score that represents the propensity of an observation to have a "high level of WFH," which is a much more meaningful measure to this purpose, and allows the score generation process to be based more heavily on explanatory variables that are relevant to WFH frequency. Scores for each observation are calculated using a probit regression of the following form:

$$\Phi^{-1}(P(wfhfrequency > 50 \mid X_i)) = \beta_0 + \beta_1 \cdot age_i + \beta_2 \cdot education_i + \beta_3 \cdot gender_i + \beta_4 \cdot industry_i + \beta_5 \cdot occupation_i + \beta_6 \cdot ethnicity_i + \beta_7 \cdot location_i, \quad (4)$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function, X is a vector of the covariates, and $(P(wfhfrequency > 50 \mid X_i))$ is the probability of an observation having a *WFHfrequency* of over 50%. Note that the covariates are all categorical variables, so each variable and β represent a vector of dummy variables and the corresponding coefficients.

Once the scores for each observation are calculated, each observation in year *t* is matched with the observation in year *t-1* that has the closest score. To impute *WFHfrequency* for each observation in year *t*, the *WFHfrequency* of the matched observation in year *t-1* is used. This process is repeated for each year pair, and the imputed *WFHfrequency* is used to create the imputed *WFHlevel* variable.

For matching, a caliper of 0.0001 is used. Any observation in year *t* for which there is no observation in year *t-1* with a score within the caliper is dropped from the analysis. The caliper was set after observing the distribution of scores in the dataset. Implementing the caliper results in about 1% of observations being dropped from the analysis.

Some users of matching-based imputation methods would suggest finding multiple matches for each observation in year *t* and using the average of the matched observations in year *t-1* to impute *WFHfrequency*. This is potentially beneficial as it reduces the penalty in the case the observation with the closest score is actually a bad match. However, the distribution of the variable of interest (*WFHlevel*) is bimodal at the left and right extremes. This means that using averaged values as the imputed value would result in a large number of observations being imputed to the middle of the distribution, and the resulting imputed distribution being unrepresentative of the original distribution. This method is therefore not used in this analysis.

Appendix F Data Availability and Technical Notes

Data

All data used in this paper are publicly available. Instructions for accessing each dataset are below.

SWAA

SWAA data can be downloaded from the *WFH Research* website at
https://wfhresearch.com/data/
The dataset is available for free. Users must create an account and state purpose of use. This paper uses data from the "WFH Code and Data: May 2020 to December 2024" release.

SIPP

SIPP data are available through the U.S. Census Bureau's website. This paper uses the 2021-2023 releases of the data which can be found at the following links:
https://www.census.gov/programs-surveys/sipp/data/datasets/2021-data/2021.html
https://www.census.gov/programs-surveys/sipp/data/datasets/2022-data/2022.html
https://www.census.gov/programs-surveys/sipp/data/datasets/2023-data/2023.html
For the 2021 and 2022 releases, the data used are under the "Primary Data File–STATA Data Format" download. For the 2023 release, the data used are under the "Primary Data File–Stata-formatted data" download. Other versions of the download files have not been tested for consistency with the results presented in this paper.

J2J Flows

The J2J Flows data used can be obtained through the U.S. Census Bureau's "J2J Explorer" tool located at:
https://j2jexplorer.ces.census.gov/
For this paper, data were extracted using two filters. The first filter removes Alaska, Michigan, North Carolina, and South Carolina from the list of "destination states" as data for those states are not up to date. The second filter sets the target "year/quarter" to all quarters in 2022 and 2023. The below link reproduces the filters used in this paper:
https://j2jexplorer.ces.census.gov/explore.html#1653468

Unemployment Trends

The unemployment data used in this paper are retrieved from FRED using the UNRATE series, which is sourced from the Current Population Survey. The data are available at the following link:
https://fred.stlouisfed.org/series/UNRATE

Average Wage by Industry

The industry-level wage data used in this paper are from the Current Employment Statistics Survey from the US Bureau of Labor Statistics. The data are downloadable using the BLS data tools at the following link:

https://data.bls.gov/series-report
The following series are used in this paper: CEU3000000003, CEU6054000003, CEU6055000003, CEU5500000003, CEU5000000003, CEU4142000003, CEU5553100003, CEU4422000003, CEU8000000003, CEU6056000003, CEU7071000003, CEU1021000003, CEU6562000003, CEU4300000003, CEU2000000003, CEU4200000003, CEU7072000003.

Technical Notes

Analysis in this paper is carried out mainly using Stata (version 18.5), with the random forest matching process for WFH frequency imputation handled in R (version 4.4.1). User-written packages used include *heatplot* (Jann, 2019) and *psmatch2* (Leuven and Sianesi, 2003) in Stata, and *ranger* (Wright and Ziegler, 2017) in R.