Work From Home Policy Analysis

01. Data Merging and Cleaning

Summary of Data Collection

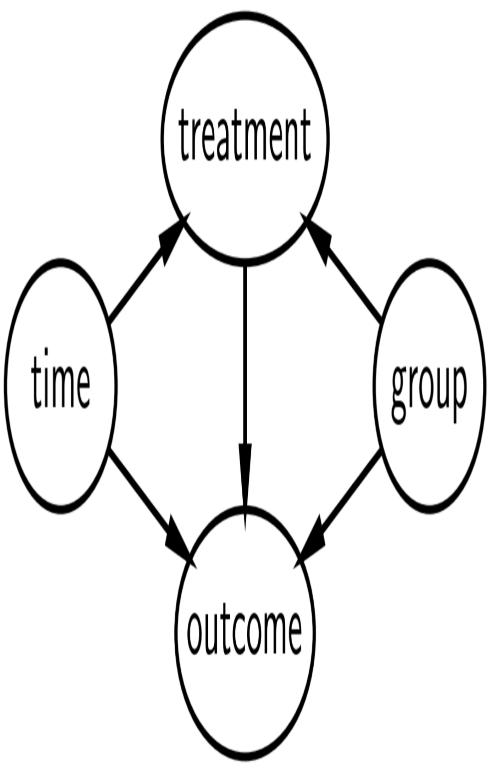
For the analysis, the goal I have is to evaluate the impact of the work from home policy(WFH) on employee retention for a Chinese travel agency. Given most of the business is done over the phone and employees making a salary through commission, the company's work from home strategy is tested with 131 employees in treatment(WFH) while the other 118 are the control(work as usual).

Merged Datasets

In order to analyze the data I merged the data for the datasets of: EmployeeStatus.dta(treatment assignment), EmployeeCharacteristics.dta(education, wages, commute), Performance_Panel.dta(monthly wage that is heavily associated to number of calls), and QuitDate.dta(quit status via month)

Cleaning Datasets

In order to simplify and correct the datasets for analysis I used the browse command to evaluate what was necessary for change and cleaning. In EmployeeCharacteristics.dta the variable rental(whether a worker was renting) had problems as it was a float and instead wanted binary identification such that I replaced the variable to rental_bin(0/1). The dataset also had problems with values given in variables age, prior_experience, and tenure which contained values of -99 that were a sign of missing data that I replaced. As for the merge of all data I also dropped non-matching data via drop if _merge != 3.



02. Empirical Strategy

Difference-in-Differences

In order to find a causal effect between the working from home policy and quitting, the empirical strategy I used was difference-in-differences(DiD). In doing so I created a variable treatmentXpost which describes the interaction between treatment and the post policy period to the outcome quit, which is 1 if an employee leaves the agency.

Logistic Regression Model

Given that the outcome was binary in quit(1 if quit, 0 if not), I believed it would be a good idea to create a logistic regression model. The logit model:

logit(P(Quit=1)) = β**0 + β1*treat_X_post_it + X_it + a_it + u_it**

This includes the constant(β**0)**, the DiD estimator(β**1*treat_X_post_it)**, and observed variables(age, gender, experience, rental, tenure,marital status, education) in X_it. I also can include unobserved variables(a_it) and the error term(u_it) as I attempt to create a precise logistic regression.

03. Tables and Figures

Table of Logit Results

Looking into the table for statistically significant variables(from ** a significance at 5% levels and *** a significance at 1% levels) we find that the treatment of the WFH policy post implementation led to a significant reduction in probability of an employee quitting with a negative coefficient.

This transformation is better observed in the table of Predicted Quit Probability given an example of Gender=Male, Age=28, Prior_Experience=2, Rental_Bin=1, Married=1, completed_HS=1:

Time Period	Probability of Quitting
Before WFH	3.84%
After WFH	1.82%

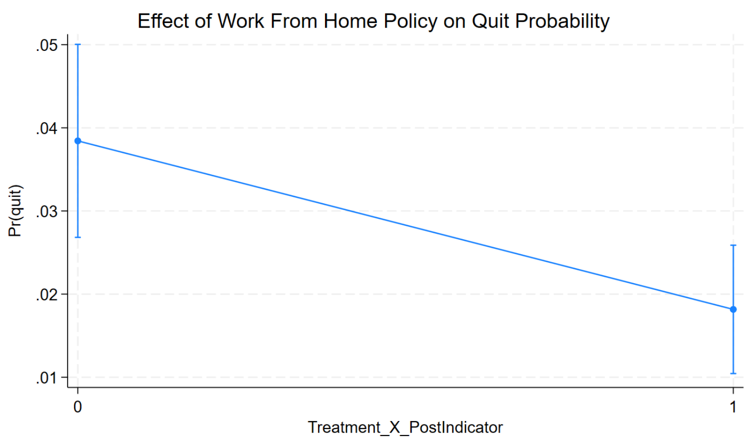
Calculating for reduction: ((3.84-1.82)/(1.82)) * 100 = 52.6% reduction in probability of this type of employee quitting.

Logistic Regression Model

In a logit model the transformation can be visualized to see the difference in work from home policy from no

Variable	Bias	Signficant	Interpretation
TreatmentXPostIndicator	Downward	Yes	Policy significantly helps reduce quit probability
Post Treatment_Period	N/A	Yes	Post_Policy Effect(No Bias)
Treatment_Group	N/A	Yes	Assginment of work-from-home workers(No Bias)
Age	Downward	No	Not significant but is downawards in bias possibly because workers are more comfortable in their careers as they get older
Tenure	Neutral	No	Not significant
Prior_Experience	Neutral	No	Not significant
Male	Upward	No	Not Significant but perhaps shows preference for job in gender comparison
Married	Upward	No	Not significant but increases quit probability prior to WFH policy possibly due to work-life conflict. Could be downward in bias by WFH.
Completed_HS	Upward	No	Not significant but higher education may open more opportunities that decentivize retention
Rental_Bin	Upward	No	Not significant but by renting and having a longer commute in-office work decentivizes retention and agency could benefit from WFH policy to negate/ shift bias
Constant	Upward	Yes	Is the baseline quit probability

implementation to implementation of policy.



04: Conclusion and Discussion

In today's evolving workplace, many companies are experimenting with work-from-home (WFH) policies to incentivize employees not to quit. For the empirical project, I analyzed data from a pilot WFH experiment that was run by a Chinese travel agency. The dataset includes detailed monthly records on 2223 observations, employee's and their backgrounds, in the company's call center that help to find what parts of a WFH policy could help the agency benefit from greater possible employee retention rates. Ultimately, my goal was to see whether moving to remote work truly makes employees less likely to quit and to understand which factors shape those decisions, especially for a company whose business model and service can be done remotely.

Focused on finding out whether or not the implementation of a work-from-home (WFH) policy has an effect on employee retention for a Chinese travel agency, I started by collecting, cleaning, and merging datasets from EmployeeStatus.dta(treatment assignment), EmployeeCharacteristics.dta(education, wages, commute), Performance_Panel.dta(a monthly wage that is heavily associated to several calls), and QuitDate.dta(quit status via month) all of which helped identify what parts of the policy are good at helping to reduce the probability of an employee quitting. To find out whether the policy was helpful and see which parts were taking away from employee retention I decided on using a difference-in-differences framework and logistic regression analysis. This strategy allowed me to find results that provide strong evidence that implementing WFH policies can significantly reduce the likelihood of employees quitting.

After cleaning the data and merging the dataset, I was able to make a logistic regression, which was useful due to its binary nature of finding out whether an employee quit such that value is 1 and 0 if not, that allowed me to visualize and test the progressive change in employee retention. Using the logistic regression model: $logit(P(Quit=1)) = \beta0 + \beta1*treat\_X\_post\_it + X\_it + a\_it + u\_it$, I was able to test the change in employee retention rates given observable inputs such as age, gender, married, and tenure, that are found in X_it of the regression, to test for change in probability of employees quitting before and after program implementation. Excited to test the regression, I tried an example of Gender=Male, Age=28, Prior_Experience=2, Rental_Bin=1, Married=1, and completed_HS=1 to see the change in probability of not quitting before and after implementation. In this example, there was a significant decrease from 3.84% before introducing the WFH policy to 1.82% in the probability of quitting after its implementation. This represents a 52.6% reduction in quit rates, a positive result that should motivate the agency to shift the business framework to one that includes more flexibility in the WFH program. Additionally, the data shows a major shift where the statistically significant interaction term of treatment_X_postindicator suggests that the WFH policy directly contributed to lower quit rates among employees in the treatment group, which helps to conclude a shift in the impact of employees married, renting, and commuting, and completed HS which were correlated to increasing probabilities of quitting before the WFH policy was implemented.

Furthermore, the analysis of additional covariates helps understand where the policy allows the agency to solve background problems that may have not been identified or solved before the program. The shift in variables such as age, tenure, and prior experience did not significantly predict quitting but had negative coefficients for age, which indicates that older employees may be more likely to stay in their positions, which could help us to assume that older employees are more comfortable in their career instead looking for another job. I also decided to include other observable and positive coefficient variables such as tenure and prior experience that weren't statistically significant for predicting quitting rates; however, were helpful in including them in the model as they would help contribute to a more precise model for estimating the change in worker retention from the policy.

Looking into the regression prior to the program the marital status affects employee turnover, with married employees possibly being more likely to quit due to greater time demands in traditional office jobs. However, work-from-home programs that allow remote work can help reduce these pressures, which may lower the quitting intentions for this group. Other factors like gender, education level, and rental status did not show a significant impact, but it is still important to understand how they influence employee retention.

The positive constant term shows the baseline probability of quitting and highlights the role of both visible and hidden factors that influence turnover. The results demonstrate that work-from-home (WFH) policies can significantly reduce employee turnover.

All together, the collected datasets that capture WFH policy effects can be analyzed and concluded to be a significant way for reducing the probability of workers quitting. The analysis helps highlight what problems stand in the way of workers and how much of an impact working from home has for incentivizing worker retention. Ultimately, the analysis I have done illustrates a positive future for an agency such as this Chinese travel agency that implements this policy.