Rachel Kim, Hyunji Park

Work from Home and Population Change: Evidence from U.S. Metropolitan Areas

INTRODUCTION

COVID-19, a highly infectious disease that started in China in late 2019, developed into a global pandemic that impacted nearly every aspect of daily life. This disease was discovered in Wuhan, China in late 2019, and was soon spread around the world. One of the most significant impacts the pandemic had was transformation in work patterns. Before the pandemic, people had worked mostly at physical offices and interacted face-to-face. After the breakout of the disease, which spread through respiratory droplets, many companies started implementing the "work-at-home" system, also known as "remote jobs," to reduce transmission risks.

Since people tend to consume more close to where they work, the reduction in office commuting decreased economic activities around workplaces—mainly traditional job centers. As the remote work system expanded across various occupations, it was no longer necessary to commute every day, weakening one of the key drivers of urban agglomeration forces: concentrated population density. Even today, numerous countries, including France and the United States, continue to utilize the remote work system.

As students who were greatly impacted by the COVID-19 pandemic, we chose a topic that we experienced and found relevant. Considering the United States has the most number of reported COVID-19 cases in the world[1]We were interested in how the population of the United States has changed in the rise of remote work. This research paper investigates the relationship between COVID-19, the remote work system, and the United States' population change in the U.S. metropolitan areas.

RESEARCH QUESTION

Our goal is to investigate the relationship between the remote work system and population change in metropolitan areas in the United States in the context of COVID-19. We utilized some key urban economics knowledge to design the model and hypothesis for our project. We looked into metropolitan areas in the United States where economic activities are concentrated and the cost of living is expensive, such as New York City, Los Angeles, and Chicago. However, gains of living in such areas offset the high costs, making many people spend high living expenses and stay there. Thus, our team predicted that with the COVID-19 pandemic, affecting the location of where people work, the areas with higher share of remote work-feasible jobs are likely to move out of their previous living areas and move to other locations that are cheaper than these cities. In addition, since the COVID-19 started in late 2019, we decided to focus on 2020 to 2022 in our research.

[1] https://data.who.int/dashboards/covid19/cases

Therefore, our team came up with the following research question: *"Did metropolitan areas with a higher share of remote work-feasible jobs experience greater population loss—or slower population growth—compared to areas with fewer such jobs?"* With this research question, this research paper strives to answer the connection between remote jobs and the population change in the United States.

LITERATURE REVIEW

The relationship between jobs that can be done at home due to COVID-19 and the United States population was covered in numerous research articles. Among these articles, the research paper conducted by Jonathan I. Dingel and Brent Neiman, "How Many Jobs Can be Done at Home?"[2], discusses the economic and demographic impact of "social distancing", which was the main term used during the COVID-19 pandemic.

This study showed that after the COVID-19 pandemic, with the assumption that all jobs had the same number of hours of work, 37% of the jobs in the United States can be performed at home. The report concluded that there is a positive correlation between the median hourly wage and the share of jobs that can be done at home as the jobs possible with remote systems are typically higher wages than jobs that cannot be done at home. Moreover, jobs such as managers, educators, computers, finance, and law are greatly able to work from home, while jobs such as farm, construction, and production workers are not able to.

Furthermore, there was a significant variation across cities in the United States regarding which jobs were able to be performed at home and which jobs were not. More than 45% of the jobs in San Francisco, San Jose, and Washington DC were able to be performed at home, and 30% or less jobs in Fort Myers, Grand Rapids, and Las Vegas were not able to change to remote jobs.

According to Bick, Blandin, and Mertens, 75.4% of the workers reported by the Real-Time Population Survey (RPS) worked at home every day, when only 8.2% of the workers worked at home in February 2020, which is before the COVID-19 pandemic[3]. With this information, our team was able to predict that there is definitely a relationship between remote work and population change.

DATA SOURCES

To explore our research question, our team found and utilized the following data. The dependent variable for our analysis is gained from the United States Census Bureau. Here, we utilized the 'annual population percentage change by metropolitan statistical area (MSA) in the United States' from 2020 through 2024. The independent variable is the share of jobs in

---

[2] https://www.nber.org/system/files/working_papers/w26948/w26948.pdf
[3] https://www.dallasfed.org/-/media/documents/research/papers/2020/wp2017.pdf

metropolitan areas that can be done remotely. We derived this measure from the widely cited Dingel and Neiman (2020) study, "How Many Jobs Can be Done at Home?" This data is static, as it only includes the year 2020. The controls are also from the United States Census Bureau, which is the initial population of year 2022.

METHODOLOGY

R Studio is the main tool for this study's data analysis because of its ability in handling large datasets, performing econometric modeling, and high-quality visualizations. R provides a wide range of packages designed specifically for statistical analysis, such as “dplyr” for data manipulation, “ggplot2” for graphing, and “lm()” for regression modeling. These codes are necessary for finding out the relationship between remote work feasibility and metropolitan cities' population change. Moreover, R is widely used in academic and professional econometrics as it ensures transparency, reproducibility, and alignment with standard research practices. R is also an open-source, which makes it a flexible and accessible choice for performing a strong, data research.

With R, the model of our research is as followed:

$$\Delta\text{Population\_it} = \beta 0 + \beta 1 * \text{WFH\_share\_i} + \beta 2 * \text{controls\_i} + \varepsilon\text{\_it}$$

“ΔPopulation_it” refers to the annual percentage change in metropolitan area population. “WFH_share_i” refers to the percentage of remote jobs in that metropolitan area, which is obtained from Dingel and Neiman data sources. “Controls_i” is the optional demographic/economic control variables, and lastly, “ε_it” is the error term, which we cannot control. With this model, programming in R Studio to analyze the data was conducted.

DATA PRESENTATION

First, we graphed the population change trends over time for major big cities that are well-known for its economic impact in the United States. From 2020 to 2024, these are the population change trend graphs for New York City, Los Angeles, and Chicago.

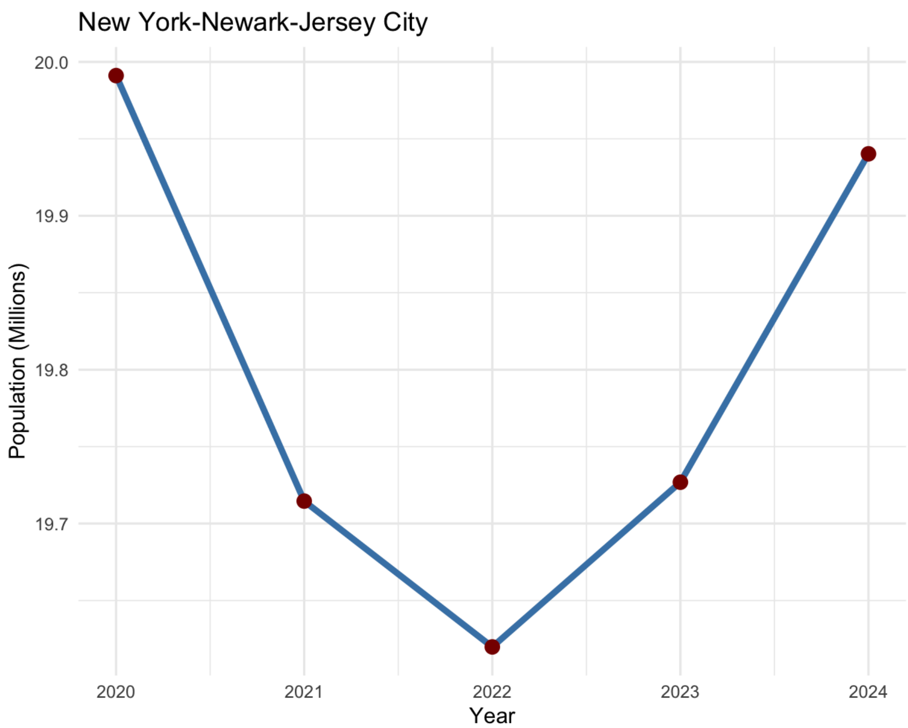

Figure 1: Population Change Trend of New York-Newark-Jersey City (2020-2024)

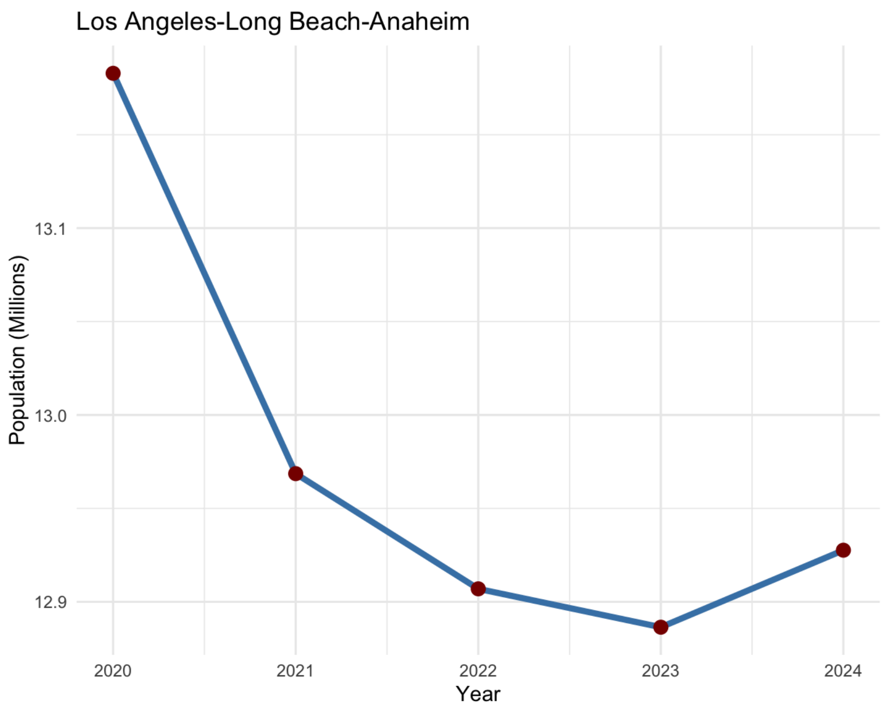

Figure 2: Population Change Trend of Los Angeles-Long Beach-Anaheim (2020-2024)

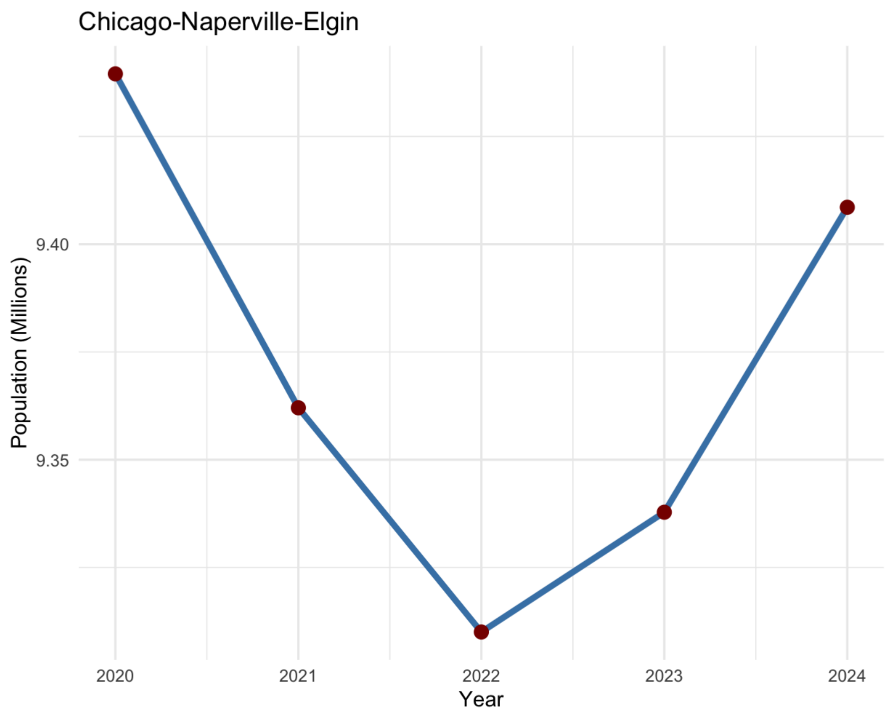

Figure 3: Population Change Trend of Chicago-Naperville-Elgin (2020-2024)

With these graphs, it is noticeable that there was a significant drop in population between 2020 to 2022 for all three cities. As this trend is common in the selected metropolitan areas, it is very likely that this trend was created as a result of the COVID-19 pandemic.

To ensure that our hypothesis is reasonable, our team created two histograms to show the percentage change for the year 2021 and 2022 in New York City, Los Angeles, and Chicago together.

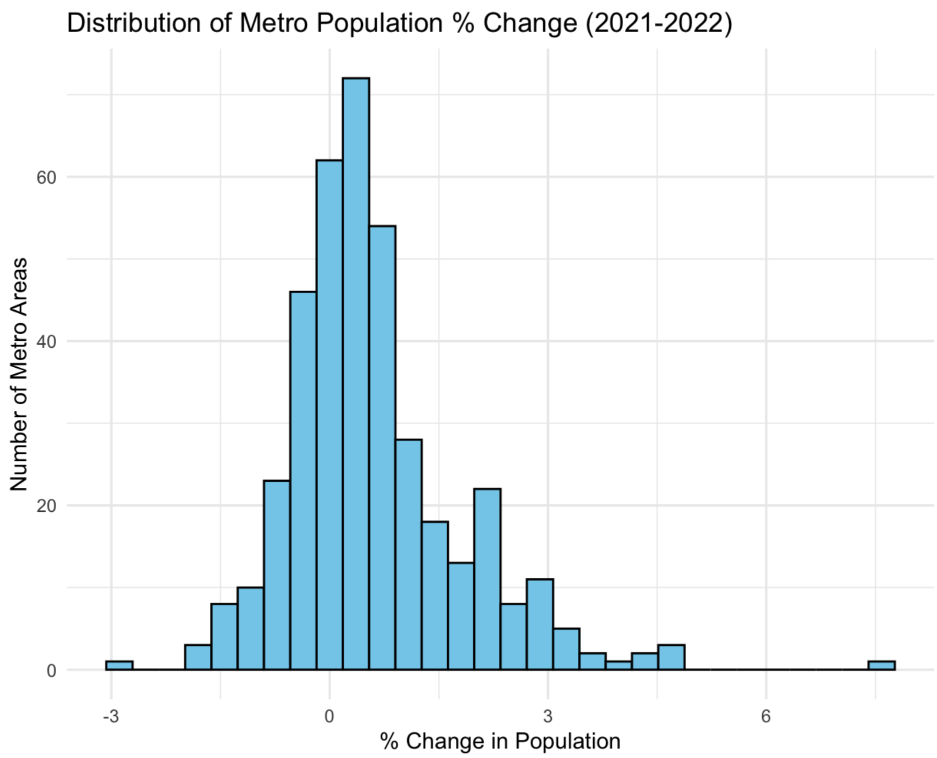

Figure 4: Distribution of Metro Population Percentage Change (2021-2022)

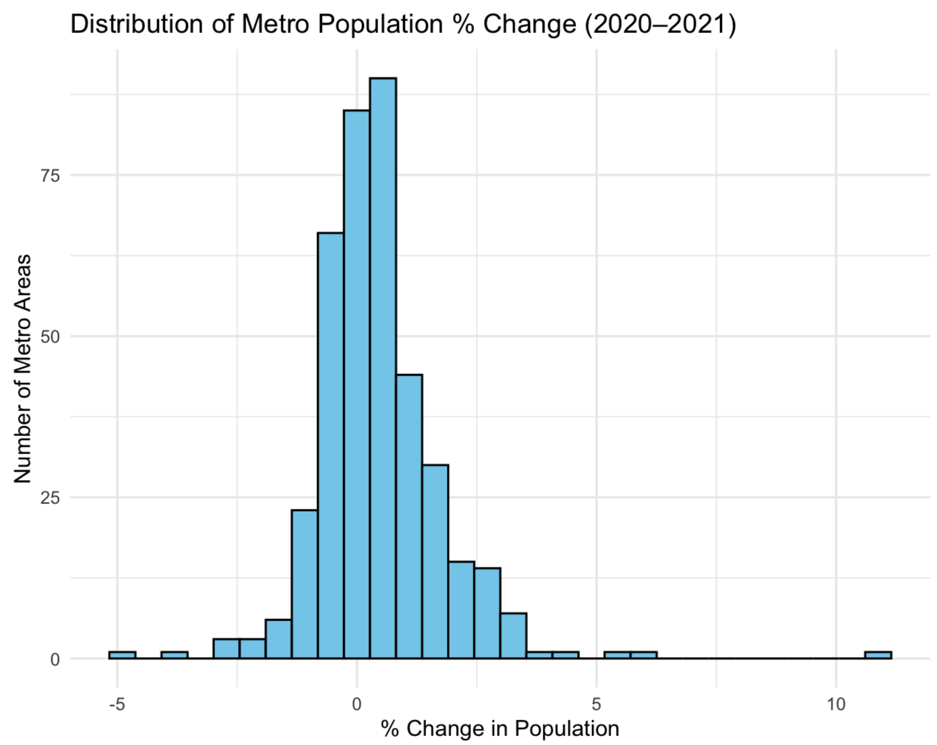

Figure 5: Distribution of Metro Population Percentage Change (2020-2021)

Here, the graphs are slightly right skewed, which means that on a national scale, there is a similar tendency of population loss in major metropolitan areas. This slight population loss proves that the COVID-19 and the change of offline jobs to remote jobs did indeed affect the United States' metropolitan areas' population.

Furthermore, for a better understanding, our team created a map of the United States that shows the population percentage change of metropolitan areas from 2020 to 2021.

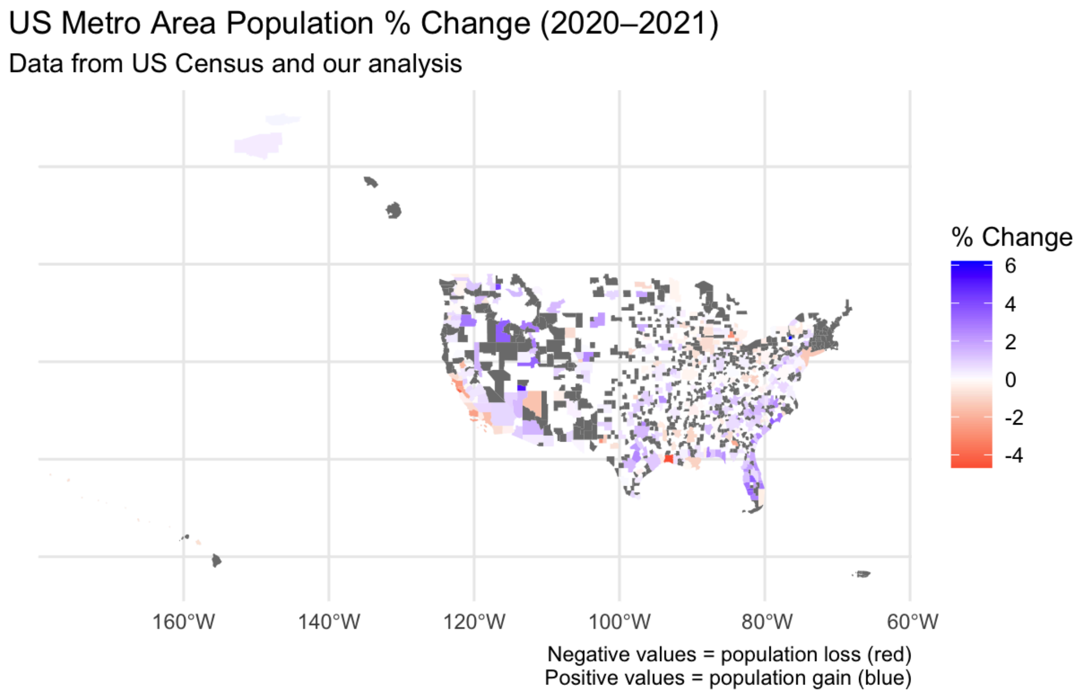

Figure 6: United States Metro Area Population Percentage Change (2020-2021)

ANALYSIS

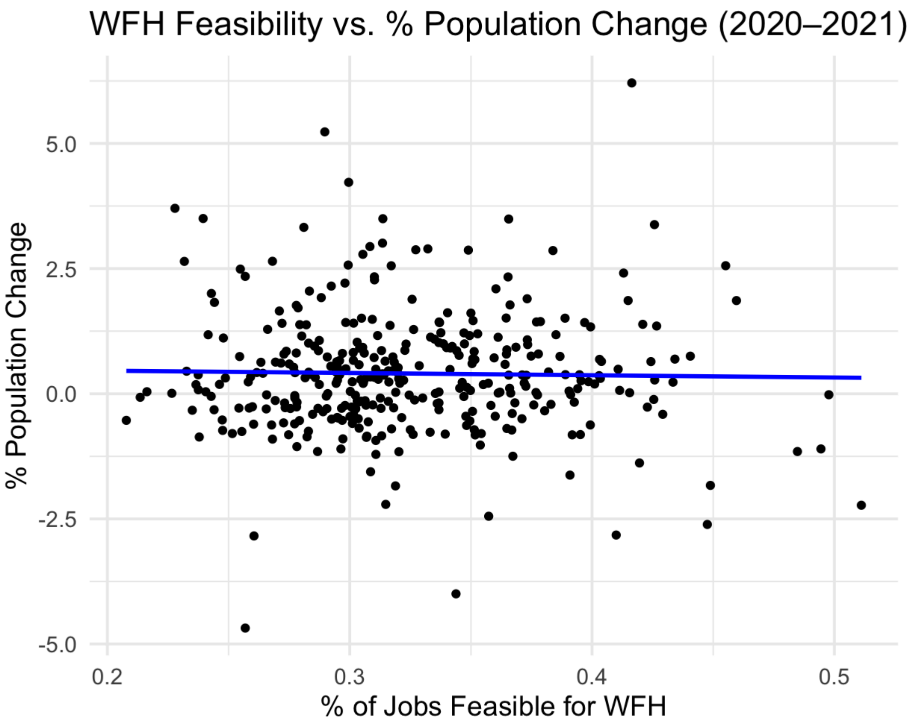

Figure 7: Work from Home Feasibility vs. Population Percentage Change (2020-2021)

The analysis shows that a higher share of work from home feasible jobs is related with slower population growth across metropolitan areas. In a simple linear regression model, it is estimated that a one-unit increase in the share of work from home feasible jobs corresponds to approximately a 0.45% point decrease in population growth between 2020 and 2021. This relationship is shown through the following regression equation:

pct_change2021 = 0.5512 − 0.4542 × teleworkable_emp

However, when the initial size of each metropolitan area, measured as the natural logarithm of the 2020 population estimate, is controlled, the effect of work from home job share becomes even more noticeable. The multi-regression model is shown through the following equation:

pct_change2021 = 0.14862 − 0.93354 × teleworkable_emp + 0.04391 × log(ESTIMATESBASE2020)

This result indicates that, after accounting for baseline metro size, the estimated negative effect of work from home feasibility nearly doubles, from −0.45 to −0.93% points. The increase in magnitude shows that larger metropolitan areas, which tend to have both higher shares of work from home feasible jobs and greater flexibility because of factors like economic diversification, infrastructure, and agglomeration economies, may have covered the true effect in the simpler model. In the absence of controls, the positive characteristics of large metros, such as attractiveness and robustness, somewhat outweigh the population-depressing effect of high remote work potential. This pattern reflects omitted variable bias in the simple regression, where failing to account for metro size restrained the full impact of work from home feasibility on population growth.

LIMITATION

There are several limitations in our data and analysis. First, Dingel and Neiman's WFH feasibility data captures a snapshot of the pre-pandemic occupational structure in 2020. Since we relied on an existing paper to measure WFH feasibility, we could not fully align the timing of our dependent and independent variables, which might cause some undesired errors in the analysis.

Second, we could not fully integrate density and other factors. Our analysis was conducted at the metropolitan statistical area (MSA) scale, which reflects economically integrated regions rather than administrative boundaries. MSAs refer to very dense areas where people live and work, but they vary in size. Thus, it was essential to take metro-level density into account in addition to the absolute population size. We could have extracted the density of each metro by having the population divided by land size, and add this new variable into our model as a control to remove confounding factors. In addition, there might be other unobserved factors that affect economic growth, such as housing supply, commute patterns, or local amenities. There might be reverse causality or simultaneity as well: population changes might influence occupational structures over time.

CONCLUSION

This study was conducted to examine whether metropolitan areas with a higher share of work from home feasible jobs experienced greater population loss or slower population growth due to the COVID-19 pandemic. By using population data from the U.S. Census Bureau and work from home feasibility estimates from Dingel and Neiman (2020), we analyzed trends across major U.S. metropolitan areas from 2020 to 2024. Initial visualizations revealed population declines in economically significant metropolitan areas such as New York, Los Angeles, and Chicago, suggesting a nationwide pattern of out-migration from high-density urban centers during the pandemic period.

Our econometric analysis with R confirmed these visual patterns as a simple regression showed that a higher share of remote work feasible jobs is associated with a 0.45% point decrease in population growth. Importantly, when we accounted for the initial population size of each metropolitan area through a multi-regression model, the negative effect of work from home feasibility became even more evident, nearly doubling to −0.93% points. This shows that larger metros, which tend to have more remote-friendly jobs and greater economic resilience, may have masked the full effect in the simpler model due to omitted variable bias.

With the controlled multi-regression model, the findings support both our research question and hypothesis: metropolitan areas with a higher share of work from home feasible jobs did, in fact, experience relatively slower population growth during the COVID-19 period. These results reflect wider changes in the spatial dynamics of work and residence, emphasizing the demographic effect of remote work in the post-pandemic era.

BIBLIOGRAPHY

Bick, Alexander, Adam Blandin, and Karel Mertens. *"Work from Home After the COVID-19 Outbreak."* National Bureau of Economic Research, Working Paper No. 28731, Apr. 2021. https://www.dallasfed.org/-/media/documents/research/papers/2020/wp2017.pdf.

Jonathan I. Dingel and Brent Neiman, *"How Many Jobs Can Be Done at Home?"* NBER Working Paper No. 26948 (National Bureau of Economic Research, April 2020), https://www.nber.org/system/files/working_papers/w26948/w26948.pdf.

World Health Organization. *"WHO Coronavirus (COVID-19) Dashboard: Cases."* WHO, https://data.who.int/dashboards/covid19/cases.

APPENDIX: CODE

```
library(dplyr)
library(tidyr)


pop <- read.csv("/Users/semi/Desktop/sp-R/final/cbsa-est2024-alldata.csv")
pop <- pop %>%
  filter(LSAD == "Metropolitan Statistical Area") %>%
  select(-c(2:3), -5, -c(17:ncol(.)))


# Create 5 new change columns
pop <- pop %>%
  mutate(
    pct_change2021 = 100 * NPOPCHG2021 / POPESTIMATE2020,
    pct_change2022 = 100 * NPOPCHG2022 / POPESTIMATE2021,
    pct_change2023 = 100 * NPOPCHG2023 / POPESTIMATE2022,
    pct_change2024 = 100 * NPOPCHG2024 / POPESTIMATE2023
  )


library(ggplot2)
library(tidyr)


# Convert wide to long for plotting
pop_long <- pop_cleaned %>%
  select(Geographic_Area, `2020`, `2021`, `2022`, `2023`, `2024`) %>%
  pivot_longer(cols = -Geographic_Area, names_to = "year", values_to = "population") %>%
  mutate(year = as.numeric(year))
```

```
# Filter for a few major metros
selected_metros <- c("New York-Newark-Jersey City, NY-NJ",
                     "Los Angeles-Long Beach-Anaheim, CA",
                     "Chicago-Naperville-Elgin, IL-IN")


# Prepare long dataset
pop_long <- pop %>%
  filter(NAME %in% selected_metros) %>%
  select(NAME, POPESTIMATE2020, POPESTIMATE2021, POPESTIMATE2022,
POPESTIMATE2023, POPESTIMATE2024) %>%
  pivot_longer(cols = starts_with("POPESTIMATE"),
               names_to = "year", values_to = "population") %>%
  mutate(year = as.numeric(gsub("POPESTIMATE", "", year)))


# Create individual plots
ny_plot <- ggplot(pop_long %>% filter(NAME == selected_metros[1]),
                  aes(x = year, y = population / 1e6)) +
  geom_line(color = "steelblue", size = 1.5) +
  geom_point(color = "darkred", size = 3) +
  labs(title = "New York-Newark-Jersey City", x = "Year", y = "Population (Millions)") +
  scale_x_continuous(breaks = 2020:2024) +
  theme_minimal()
ny_plot


la_plot <- ggplot(pop_long %>% filter(NAME == selected_metros[2]),
                  aes(x = year, y = population / 1e6)) +
  geom_line(color = "steelblue", size = 1.5) +
```

```
  geom_point(color = "darkred", size = 3) +
  labs(title = "Los Angeles-Long Beach-Anaheim", x = "Year", y = "Population (Millions)") +
  scale_x_continuous(breaks = 2020:2024) +
  theme_minimal()
la_plot


chicago_plot <- ggplot(pop_long %>% filter(NAME == selected_metros[3]),
                       aes(x = year, y = population / 1e6)) +
  geom_line(color = "steelblue", size = 1.5) +
  geom_point(color = "darkred", size = 3) +
  labs(title = "Chicago-Naperville-Elgin", x = "Year", y = "Population (Millions)") +
  scale_x_continuous(breaks = 2020:2024) +
  theme_minimal()
chicago_plot



# Plot 2-1: Histogram of % change for 2021
ggplot(pop, aes(x = pct_change2021)) +
  geom_histogram(bins = 30, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Metro Population % Change (2020–2021)",
       x = "% Change in Population",
       y = "Number of Metro Areas") +
  theme_minimal()


# Plot 2-2: for 2022
ggplot(pop, aes(x = pct_change2022)) +
  geom_histogram(bins = 30, fill = "skyblue", color = "black") +
```

```
  labs(title = "Distribution of Metro Population % Change (2021-2022)",
       x = "% Change in Population",
       y = "Number of Metro Areas") +
  theme_minimal()



# Plot 3: Map
library(sf)
library(dplyr)
library(ggplot2)

# Step 1: Read metro shapefile
metro_shapes <- st_read("/Users/semi/Desktop/sp-R/final/tl_2023_us_cbsa.shp")

# First, make sure both CBSA columns are character
metro_shapes$GEOID <- as.character(metro_shapes$GEOID)
final_data$CBSA <- as.character(final_data$CBSA)

# Step 2: Check your population data
metro_shapes <- metro_shapes %>%
  left_join(final_data, by = c("GEOID" = "CBSA"))  # GEOID = CBSA code in shapefile

# Step 3: Plot map for a specific year, e.g., pct_change2021
map <- ggplot(metro_shapes) +
  geom_sf(aes(fill = pct_change2021), color = NA) +
  scale_fill_gradient2(low = "red", mid = "white", high = "blue", midpoint = 0,
                       name = "% Change") +
```

```
  labs(title = "US Metro Area Population % Change (2020–2021)",
       subtitle = "Data from US Census and your analysis",
       caption = "Negative values = population loss (red)\nPositive values = population gain
(blue)") +
  theme_minimal()
map


###################


# Read the data
wfh <- read.csv("/Users/semi/Desktop/sp-R/final/MSA_workfromhome.csv")


# Merge
wfh$AREA <- as.character(wfh$AREA)
final_data <- left_join(pop, wfh, by = c("CBSA" = "AREA"))


# Drop rows with NA
final_data <- final_data %>%
  drop_na()


# Regression: WFH - % pop. change
reg <- lm(pct_change2021 ~ teleworkable_emp, data = final_data)
reg


# Scatterplot
library(ggplot2)
ggplot(final_data, aes(x = teleworkable_emp, y = pct_change2021)) +
```

```
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "WFH Feasibility vs. % Population Change (2020–2021)",
       x = "% of Jobs Feasible for WFH",
       y = "% Population Change") +
  theme_minimal(base_size = 14)


model_control <- lm(pct_change2021 ~ teleworkable_emp + log(ESTIMATESBASE2020),
                    data = final_data)
model_control
reg # compare the results
```