COVID-19 Pandemic and Unemployment: Evidence from Mobile Phone Data in China

Teng Li Panle Jia Barwick Yongheng Deng
Xinfei Huang Shanjun Li

CONTENTS

1. INTRODUCTION

Impact of the COVID-19 pandemic on Unemployment?

Effective and targeted policies to address the adverse consequences of the COVID-19 pandemic for the economy

RELY ON

Prompt and accurate measures of the labor market effects

1. INTRODUCTION

Prompt and accurate measures of the labor market effects

HOWEVER

Traditional Measures are based on Surveys

ROUTINELY

1. The substantial **time lag**
2. Limited availability for **small geographic areas**

SPECIFICALLY

In China

Measuring unemployment accurately is particularly challenging due to

1. A large fraction of the population **excluded** from the unemployment surveys
2. **Reporting and aggregation errors**

1. INTRODUCTION

Prompt and accurate measures of the labor market effects

High-frequency and **high-resolution** mobile phone usage data in **Guangdong**

Merits

1. **Representative**: mobile phone records for **71 million** users and location tracking information for one million **randomly selected** users
2. **Period**: from January 2018 to September 2020

The **most populous province** in China, with a GDP larger than all but the top 12 countries in the world

2. BACKGROUND AND DATA

Why Mobile Phone Data? → **It is particularly advantageous in China**

Because of the high cellphone penetration rate

1. According to the 2018 China Family Panel Studies, a nationally representative longitudinal survey of individuals' social and economic status, **89%** of correspondents sixteen years and older reported possessing a cellphone.

2. In addition, each household owns **2.5** cell phones on average (National Bureau of Statistics 2018).

2. BACKGROUND AND DATA

Why Mobile Phone Data? → **It is particularly advantageous in China**

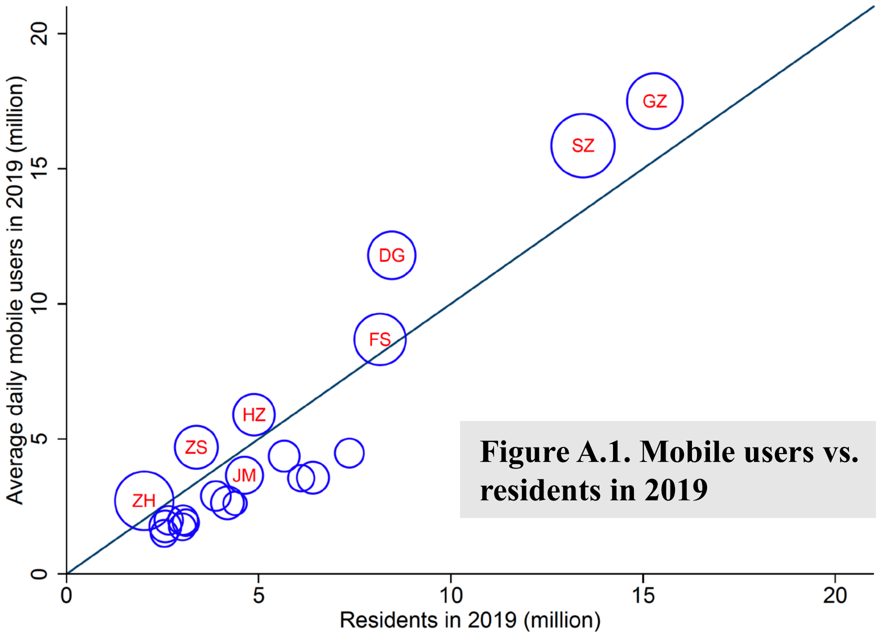


Figure A.1. Mobile users vs. residents in 2019

Because of the high cellphone penetration rate

Figure A.1 shows **a strong correlation** between the number of phone users and the number of residents by city.

Cities with a higher GDP per capita (represented by the size of the circles in Figure A.1) tend to have higher mobile phone ownership.

2. BACKGROUND AND DATA

Why in Guangdong? → **It is highly relevant**

The most populous province with the largest provincial GDP in China

1. Guangdong contributes to **11%** of China's GDP and around a quarter of China's foreign trade (China Statistical Yearbook 2020).

2. The number of daily confirmed COVID cases are under a few handfuls since the full reopening, similar to most other provinces in China.

Our measures on the pandemic's consequences could apply to other regions as well.

2. BACKGROUND AND DATA

Why in Guangdong? → **It is highly relevant**

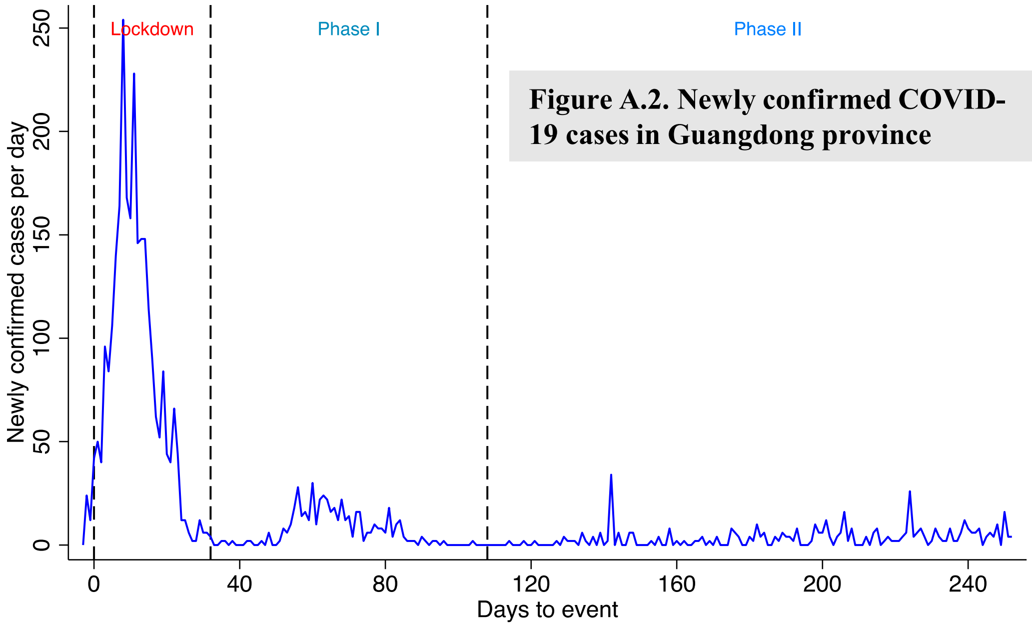


Figure A.2. Newly confirmed COVID-19 cases in Guangdong province

The number of daily confirmed COVID cases are under a few handfuls since the full reopening, similar to most other provinces in China.

Our measures on the pandemic's consequences could apply to other regions as well.

2. BACKGROUND AND DATA

Data Resources

Our data come from a dominant cellular service provider in China

1. We have access to **detailed phone usage** (encrypted IDs of the calling party and the receiving party, date of calls, and call duration in seconds) for all of **its 71 million users** in Guangdong Province from January 2018 to September 2020, accounting for **63%** of all mobile users in the province.

2. We observe some **user demographic information**, such as **age, gender, and the place where the phone number is registered**. In addition, we have access **to the location records every five-minute interval** for **one million randomly selected users** during the same period.

2. BACKGROUND AND DATA

Unemployment Measures

We leverage **two features** of the mobile phone data to construct unemployment measures

WORK COMMUTING	CALLS TO UNEMPLOYMENT HOTLINE
Reductions in the number of people working on-site before and after the lockdown and relative to 2019	The number of individuals calling the designated government hotline (12333) for unemployment benefits

2. BACKGROUND AND DATA

Unemployment Measures | **WORK COMMUTE**

We use increase in the number of non-commuters before and after the lockdown and relative to 2019 as our major measure of pandemic-induced unemployment.

1. We define **the work location** as the location where a user spends at least 5 hours a day between 9 am and 6 pm for at least fifteen workdays in a given month.

2. **Changes in commuting patterns** provides valuable barometer of changes in unemployment, especially when participation in the unemployment benefit programs is low (as is the case in China).

2. BACKGROUND AND DATA

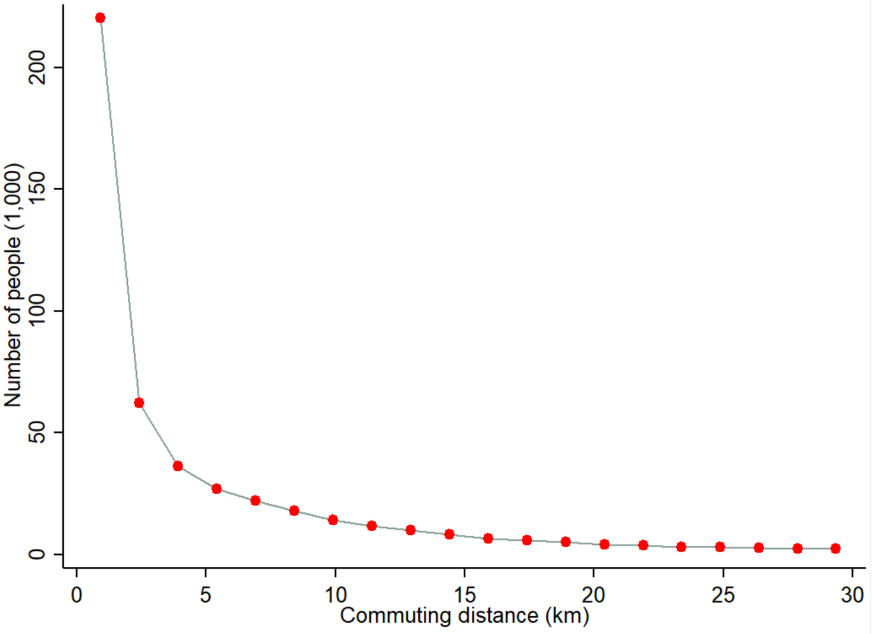


Figure 1. The number of people commuting by commuting distance

2. BACKGROUND AND DATA

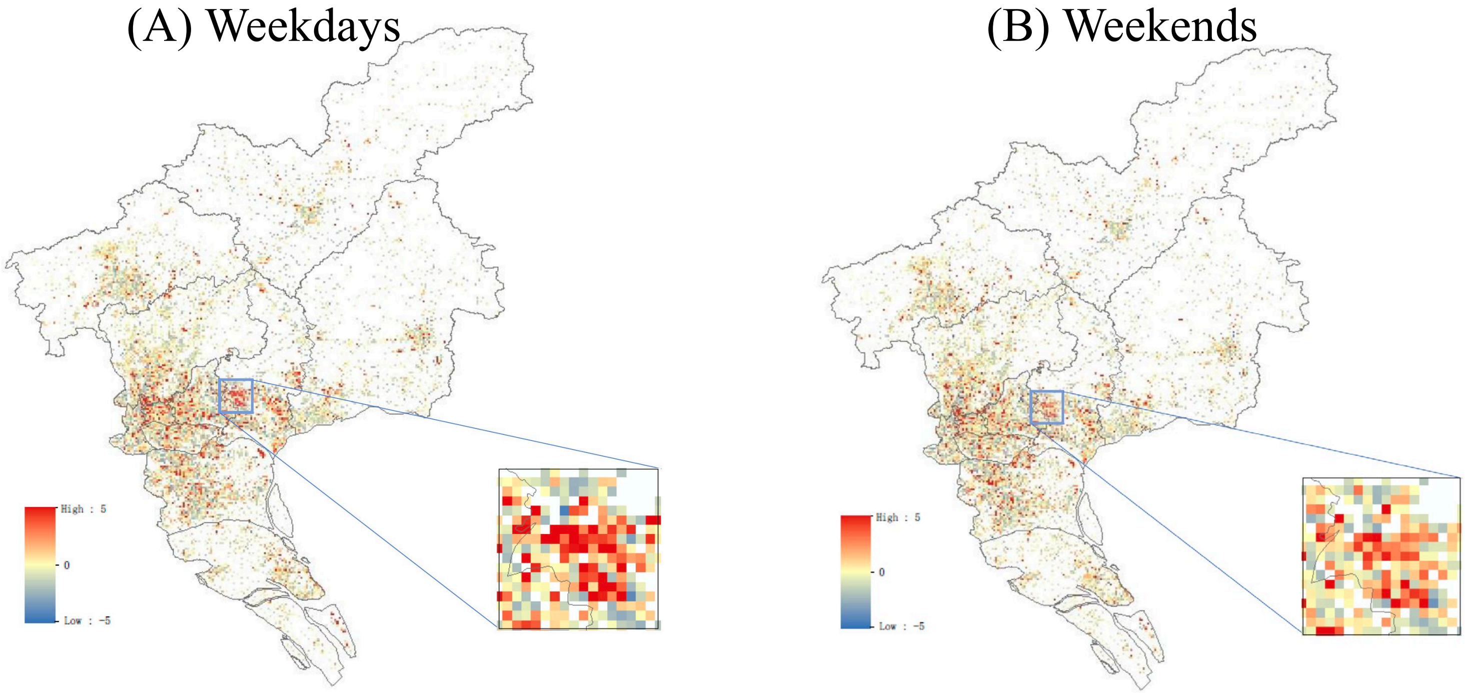


Figure A.3. Differences in (log) population during daytime and nighttime in Guangzhou

2. BACKGROUND AND DATA

Unemployment Measures

UNEMPLOYMENT BENEFIT CLAIMS

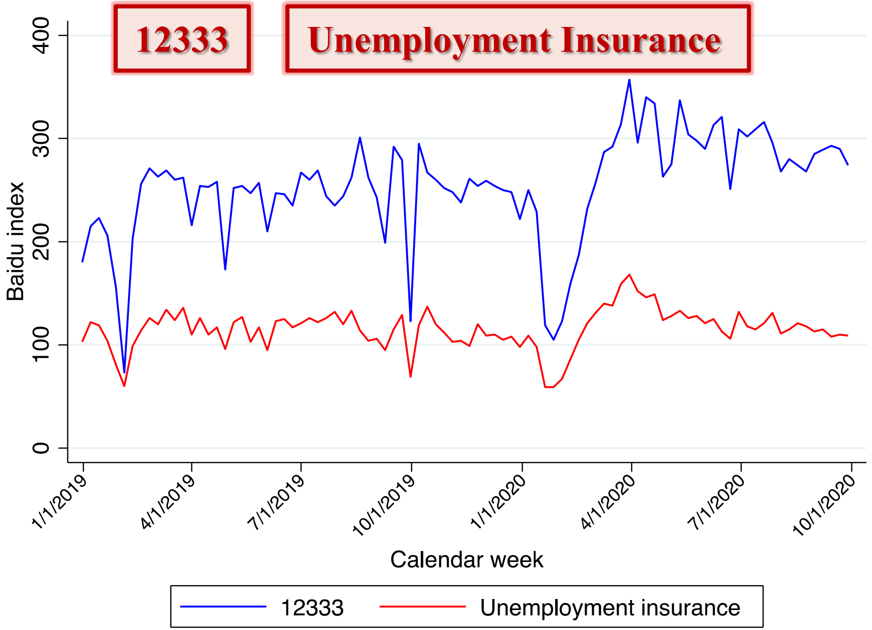


Figure 2. Correlation of Baidu index between keywords of "12333" and "unemployment insurance" in Guangdong province

The correlation of the Baidu Index of the two keywords is **0.83** during the sample period. The co-movement of the index for the two keywords offers additional support for using the 12333 hotline as a proxy for individuals claiming unemployment benefits.

2. BACKGROUND AND DATA

Unemployment Measures

UNEMPLOYMENT BENEFIT CLAIMS

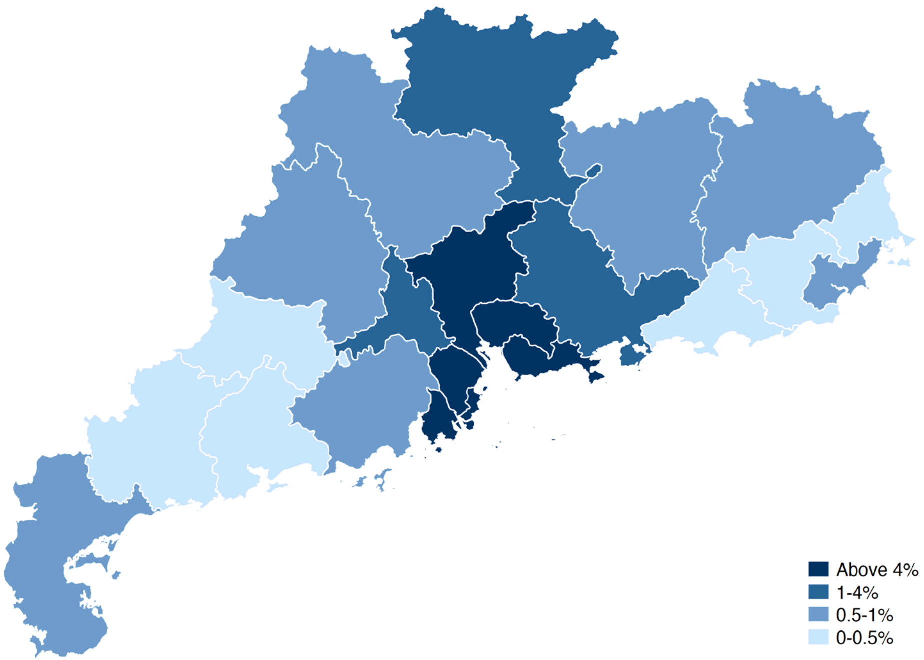


Figure A.4. Unemployment rate by city in 2019 based on the number of individuals making unemployment calls

The correlation between city-level unemployment calls and the official unemployment rate released by the NBS, which is only available annually for city-level statistics, **is reasonably high at 0.7 in 2019**.

2. BACKGROUND AND DATA

Sample Construction

1. Our main analysis **excludes users** under the age of 18, and results excluding users under the age of 25 (to eliminate those still in school) are almost identical.

2. We define **migrants** as individuals who registered their phone numbers outside Guangdong province.

3. We only count **the first time** when a user reaches out to the unemployment benefit hotline.

2. BACKGROUND AND DATA

Descriptive Analysis

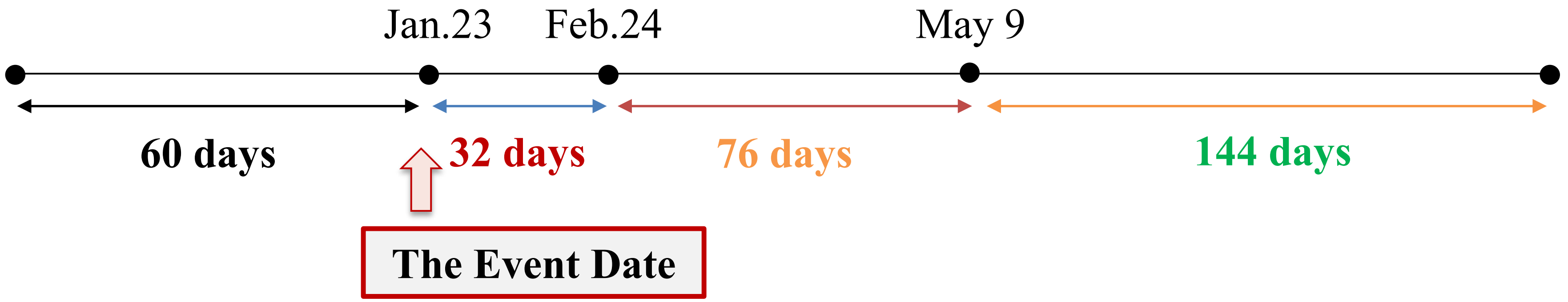
	Jan.23	Feb.24	May 9
60 days	**32 days**	**76 days**	**144 days**
Before	**Lockdown**	**Phase I Reopening**	**Phase II Reopening**

Phase I Reopening: People were allowed to **go back to work** and **visit outdoor public places**.

Phase II Reopening: **All businesses**, including shopping malls, supermarkets, and restaurants, were allowed to reopen fully. (The only exception was movie theaters)

2. BACKGROUND AND DATA

Our analyses use **a standard DID framework** and exploit differences in our two key measures of labor market outcomes between **2020 (the treatment group)** and **2019 (the control group).**



Our analyses compare **changes in labor market outcomes** before and after **the event date** in 2020 with changes in labor market measures before and after **the exact event date** in 2019.

3. EMPIRICAL FRAMEWORK

The standard difference-in-differences (DID) approach

i denotes the treatment group (year 2020) or the control group (year 2019)

t denotes the event-day ($t = 0$ stands for January 23 in 2020 and February 3 in 2019)

$$y_{cit} = \sum_{q=-5}^{24} \boldsymbol{\beta_q} \cdot d_i \cdot \mathbb{1}(t \in [q * 10 + 1, (q + 1) * 10]) + \alpha_c + \gamma_i + \eta_t + \xi_{it} + \varepsilon_{cit} \quad (1)$$

c denotes a neighborhood (a cell-tower area)

$\boldsymbol{\beta_q}$ are event-study coefficients, capturing differences between the treatment group and the control group

$\mathbb{1}(\cdot)$ is an indicator variable for each 10-day interval of the sample.

3. EMPIRICAL FRAMEWORK

The standard difference-in-differences (DID) approach

Neighborhood fixed effects

The holiday fixed effect

$$y_{cit} = \sum_{q=-5}^{24} \boldsymbol{\beta_q} \cdot \mathrm{d}_i \cdot \mathbb{1}(t \in [q * 10 + 1, (q+1) * 10]) + \boldsymbol{\alpha_c} + \boldsymbol{\gamma_i} + \boldsymbol{\eta_t} + \boldsymbol{\xi_{it}} + \boldsymbol{\varepsilon_{cit}} \quad (1)$$

Group fixed effects

312 event-day fixed effects

3. EMPIRICAL FRAMEWORK

We also extend the model (1) to account for various heterogeneities

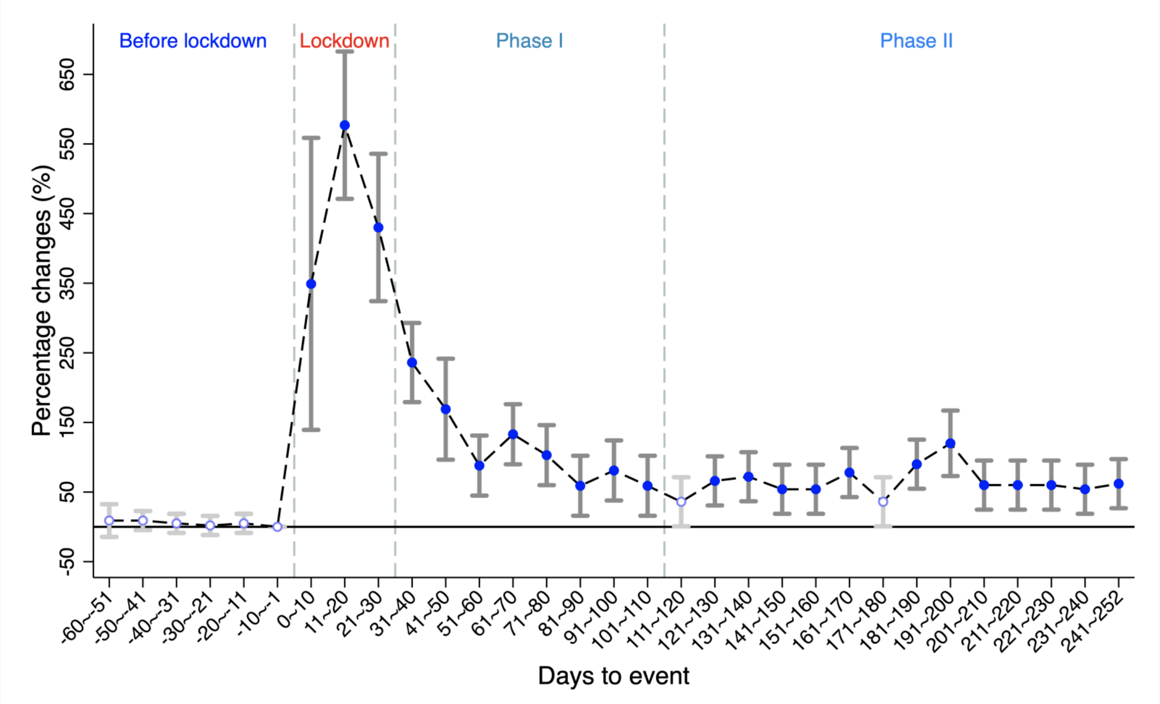
$$y_{cit} = d_i \cdot \mathbb{1}(t \in [0,252]) \cdot \boldsymbol{Z}'\tau + d_i \cdot \boldsymbol{Z}'\mu + \mathbb{1}(t \in [0,252]) \cdot \boldsymbol{Z}'\rho$$

$$+ \beta \cdot d_i \cdot \mathbb{1}(t \in [0,252]) + \alpha_c + \gamma_i + \eta_t + \xi_{it} + \varepsilon_{cit} \qquad (2)$$

Where $\boldsymbol{Z}$ is a vector of city attributes in 2019 and τ, μ, ρ are corresponding coefficients.

For example, $\boldsymbol{Z}$ could be a city's labor share in each of the 13 major industries, dummies for the 21 cities, or a city's export-over-GDP ratio.

4. MAIN RESULTS

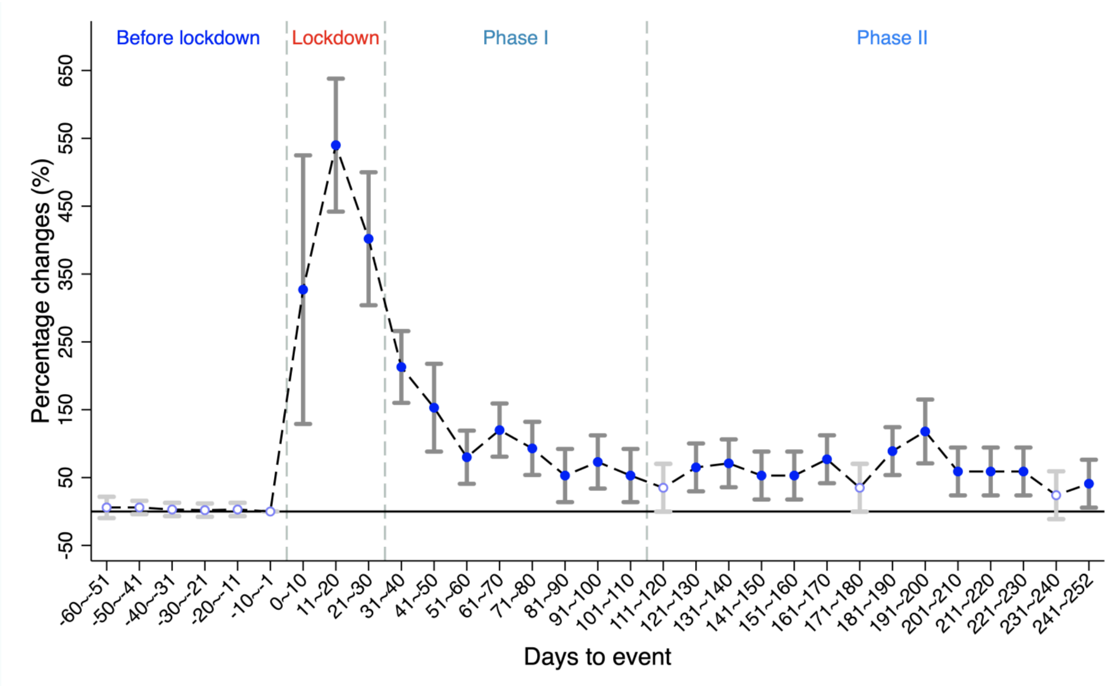


1st Measure: Commuting

Figure 3(a). Changes in the number of non-commuters between 2019 and 2020

The increase gradually came down to about 50% by the end of Phase I reopening and the change remained stable even four months after Phase II.

4. MAIN RESULTS



Unemployed or working from home?

Figure (b). Non-commuters who did not use emails/virtual meeting apps

While the increase is slightly smaller but the pattern stays the same.

4. MAIN RESULTS

Effect on non-commuters

β_q **Commuting**

Table 1: Effects on commuters and working hours on-site

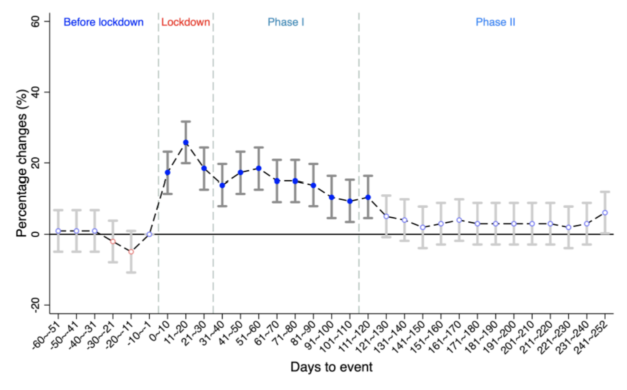
VARIABLES	(1) No. of non-commuters (in log) Two-week window	(2) Working hours (in log)
1-30 days before lockdown	0.07	0.01
	(0.05)	(0.01)
Lockdown period	4.51***	-0.21***
	(1.26)	(0.02)
Phase I re-opening	1.03***	-0.08***
	(0.36)	(0.01)
Phase II re-opening	0.59**	-0.02
	(0.30)	(0.02)
Observations	34,965	34,965
R-squared	0.92	0.95
Neighborhood FE	Yes	Yes
Event-fortnight FE	Yes	Yes
Treatment group FE	Yes	Yes

Column 1 examines the percentage change in **the number of non-commuters and working hours** as a result of the pandemic following equation (1), except that the ten-day intervals are grouped into four periods: before lockdown, during the lockdown, Phase I reopening, and Phase II full reopening.

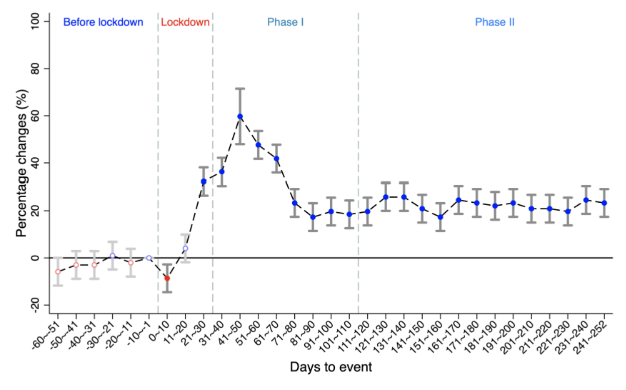
4. MAIN RESULTS

Heterogeneity Analysis β_q **Commuting**

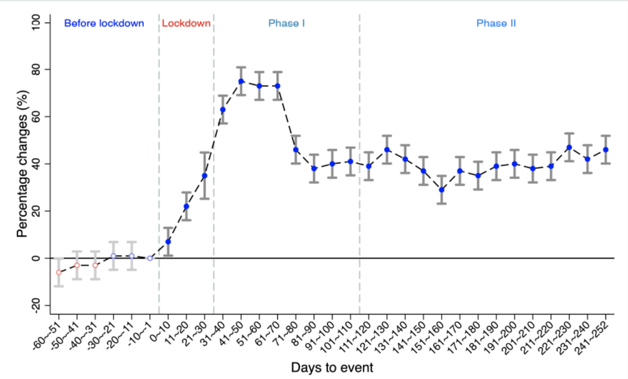
(a) Female v.s. male



(b) 40 and above v.s. under 40 years old



(c) Migrants v.s. non-migrants



1. **Females** are more affected by the pandemic.
2. **Older** workers fared worse than younger cohorts.
3. **Migrants** are most severely affected by the pandemic.

4. MAIN RESULTS

Heterogeneity Analysis

β_q **Commuting**

Cities	Population in 2019 (million)	GDP in 2019 (billion USD)
Shenzhen (SZ)	13.44	390.25
Guangzhou (GZ)	15.31	342.44
Foshan (FS)	8.16	155.81
Dongguan (DG)	8.46	137.43
Huizhou (HZ)	4.88	60.54
Zhuhai (ZH)	2.02	49.8
Maoming (MM)	6.41	47.13
Jiangmen (JM)	4.63	45.6
Zhongshan (ZS)	3.38	44.94
Zhanjiang (ZJ)	7.36	44.42
Shantou (ST)	5.66	39.04
Zhaoqing (ZQ)	4.19	32.59
Jieyang (JY)	6.11	30.46
Qingyuan (QY)	3.89	24.61
Shaoguan (SG)	3.03	19.11
Yangjiang (YJ)	2.57	18.73
Meizhou (MZ)	4.38	17.2
Chaozhou (CZ)	2.66	15.67
Shanwei (SW)	3.02	15.66
Heyuan (HY)	3.11	15.65
Yunfu (YF)	2.55	13.36

Table A.2. Cities in Guangdong Province

There is considerable variation **across cities** in Guangdong in terms of population and GDP (Appendix Table A.2).

4. MAIN RESULTS

Heterogeneity Analysis β_q **Commuting**

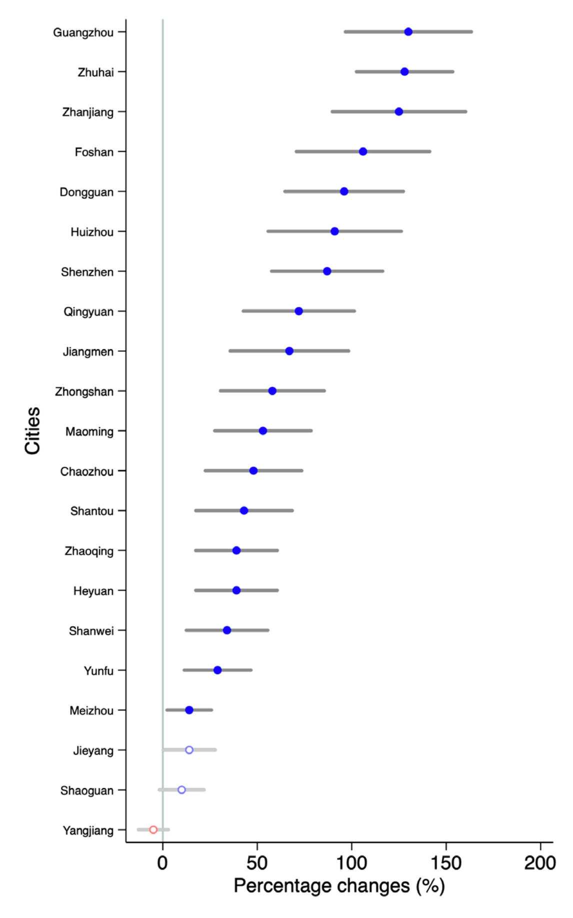


Figure 6(a). Heterogeneity across cities

The figure reports **coefficients** on the interactions of the pandemic treatment dummy (Jan 23 - Sept 30, 2020) and city dummy.

Heterogeneity across cities is sizeable

4. MAIN RESULTS

Heterogeneity Analysis

β_q **Commuting**

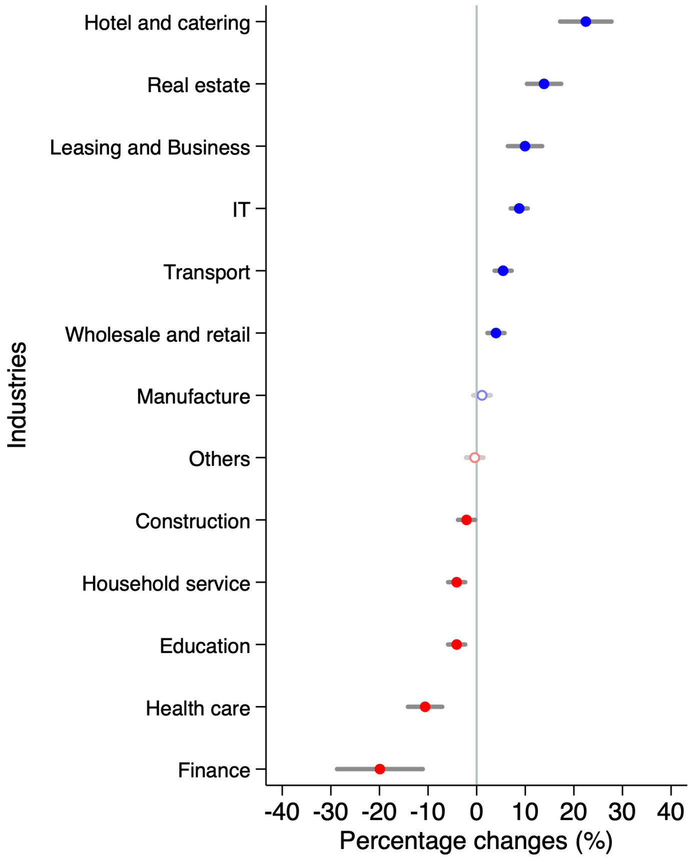


Figure 6(b). Heterogeneity across industries

The **hotel and catering**, **real estate**, and **transportation** sectors witnessed the largest increase in non-commuters.

The **finance**, **health care**, and **education** sectors experienced *reductions* in non-commuters after the lockdown in January.

4. MAIN RESULTS

Heterogeneity Analysis

β_q **Commuting**

Table 2. Heterogeneity by export-to-GDP ratio

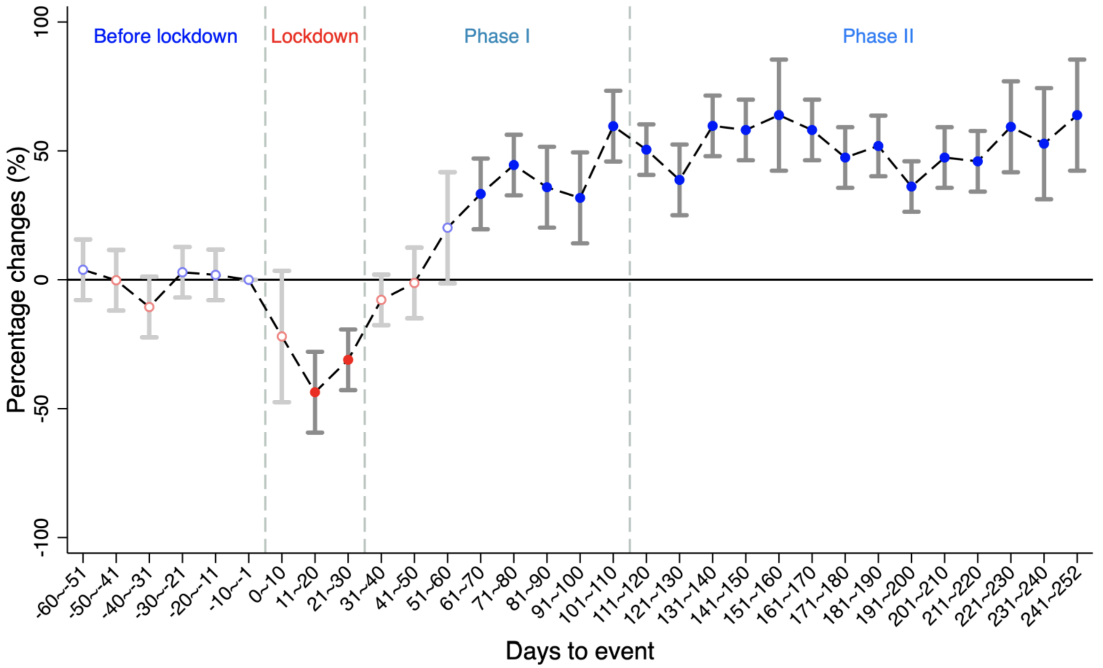
Table 2: Heterogeneity by export-to-GDP ratio using commuting patterns

VARIABLES	(1) No. of non-commuters (in log) Two-week window	(2) Working hours (in log)
Phase II re-opening * Export/GDP in 2019 (%)	0.0037***	-0.0002*
	(0.0007)	(0.0001)
Phase II re-opening (=1)	0.0329***	-0.0623***
	(0.0092)	(0.0065)
Observations	34,965	34,965
R-squared	0.93	0.81
Neighborhood FE	Yes	Yes
Event-fortnight FE	Yes	Yes
Treatment group FE	Yes	Yes

Table 2 interact the **pandemic treatment variable** with **a city's export-to-GDP share**.

A one percentage point increase in 2019's export-to-GDP ratio is associated with a **0.37%** increase in the number of non-commuters and a **0.02%** decrease in working hours.

4. MAIN RESULTS



2nd Measure:

The number of individuals calling the unemployment hotline 12333

Figure 4. Differences in unemployment calls between 2019 and 2020

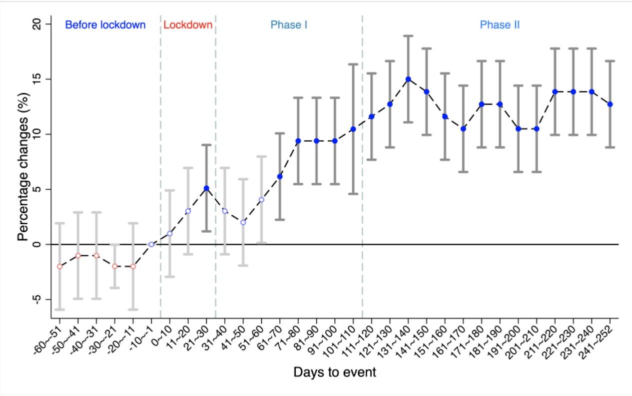
The increase in the number of calls to 12333 stabilized at about 50% by the end of the Phase I reopening and remained till the end of our data period.

4. MAIN RESULTS

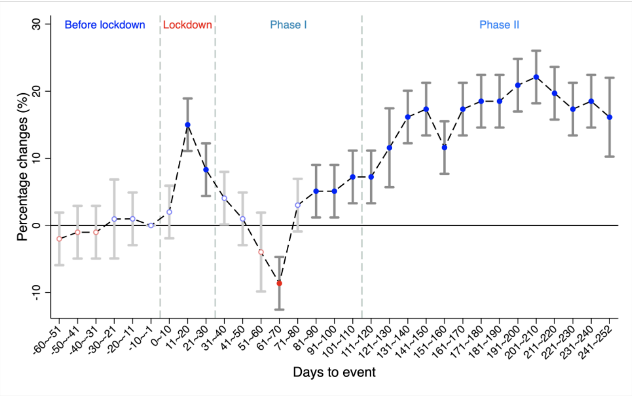
Heterogeneity Analysis

β_q **Unemployment hotline calls**

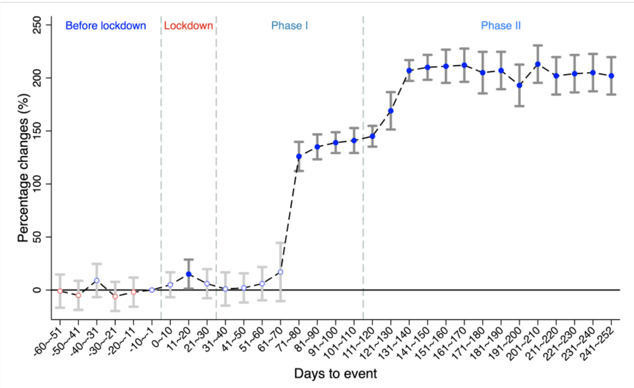
(a) Female v.s. male



(b) 40 and above v.s. under 40 years old



(c) Migrants v.s. non-migrants

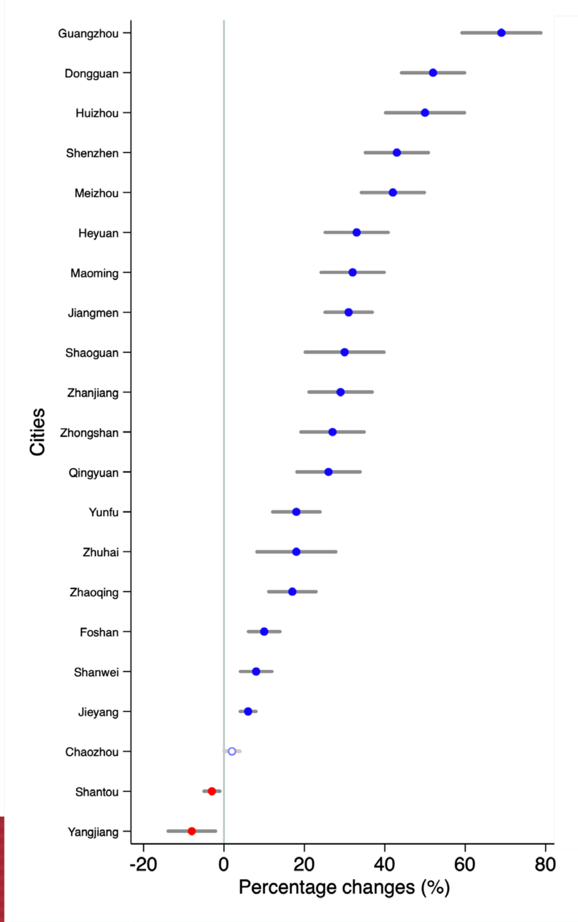


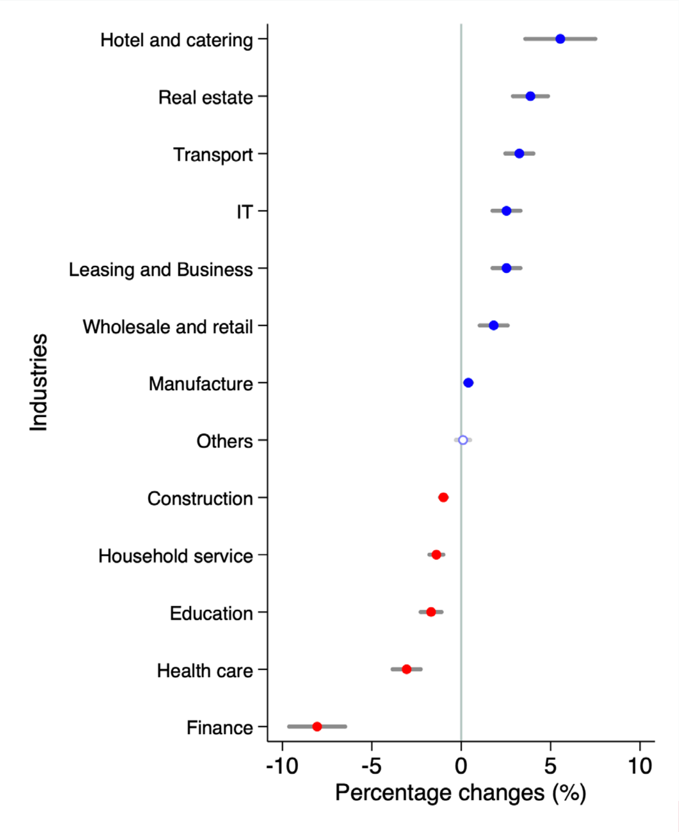
1. **Females** are more affected by the pandemic.
2. **Older** workers fared worse than younger cohorts.
3. **Migrants** are most severely affected by the pandemic.

4. MAIN RESULTS

Heterogeneity Analysis

β_q Unemployment hotline calls





Sizable heterogeneity by city.

The **hotel and catering**, **real estate**, and **transportation** sectors witnessed the largest increase in unemployment benefit claims.

The **finance**, **health care**, and **education** sectors experienced *reductions* in benefit claims after the lockdown in January.

4. MAIN RESULTS

Measure benefit claims

β_q **Unemployment hotline calls**

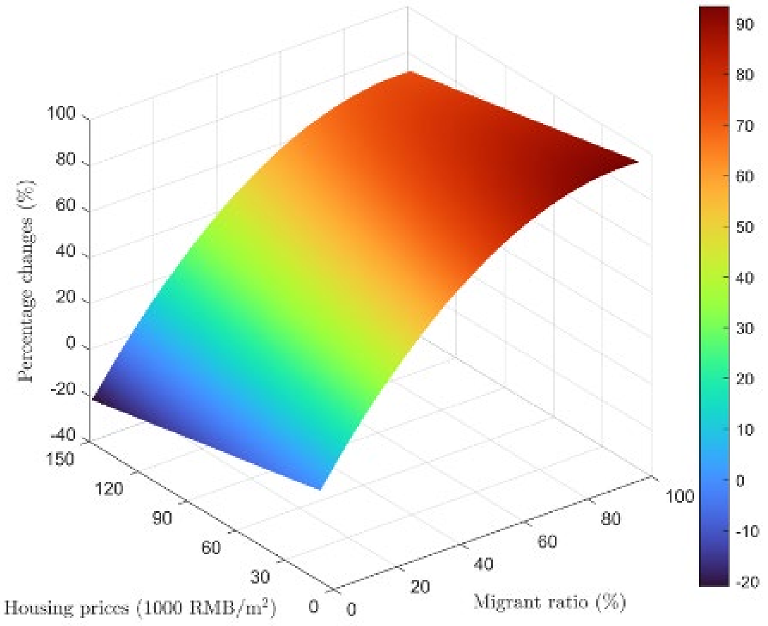


Figure 10. Changes in unemployment claims by income and migrant share in Guangzhou

- This graph plots the predicted **percentage changes** in the individuals calling 12333 between 2019 and 2020 against the **average housing price** and the **migrant share**.
- It shows that the **impact on unemployment** is **stronger** in neighborhoods with a **lower housing price** and a **higher share of migrants**.

* The housing prices (from Soufang.com) and migrant shares are based on 2018 data in each neighborhood (i.e., cell-tower-area).

4. MAIN RESULTS

Heterogeneity Analysis

β_q **Unemployment hotline calls**

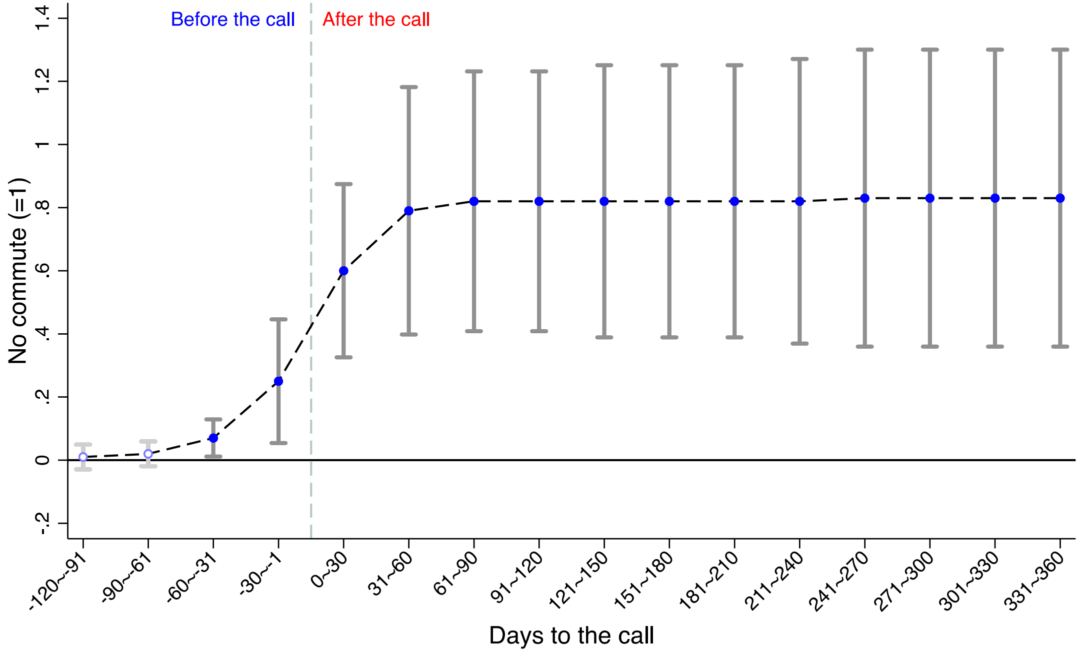
Table 4: Heterogeneity by export-to-GDP ratio using unemployment calls

VARIABLES	(1) No. of individuals making calls (in log)	(2) Duration per call (in log)
Pandemic period * Export/GDP in 2019 (%)	0.0020***	0.0022***
	(0.0004)	(0.0004)
Pandemic period (=1)	0.2137***	0.2548***
	(0.0221)	(0.0368)
Observations	489,514	489,514
R-squared	0.79	0.56
Neighborhood FE	Yes	Yes
Event-day FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

Table 4 shows cities with a larger export-to-GDP ratio experience a larger increase in the number of unemployment benefit claims during the pandemic. This is consistent with the commuting results.

4. MAIN RESULTS

Unemployment hotline callers stop commuting



More than 80% users have stopped commuting for at least a month during the two months after the call.

4. MAIN RESULTS

Unemployment hotline calls before and after the pandemic

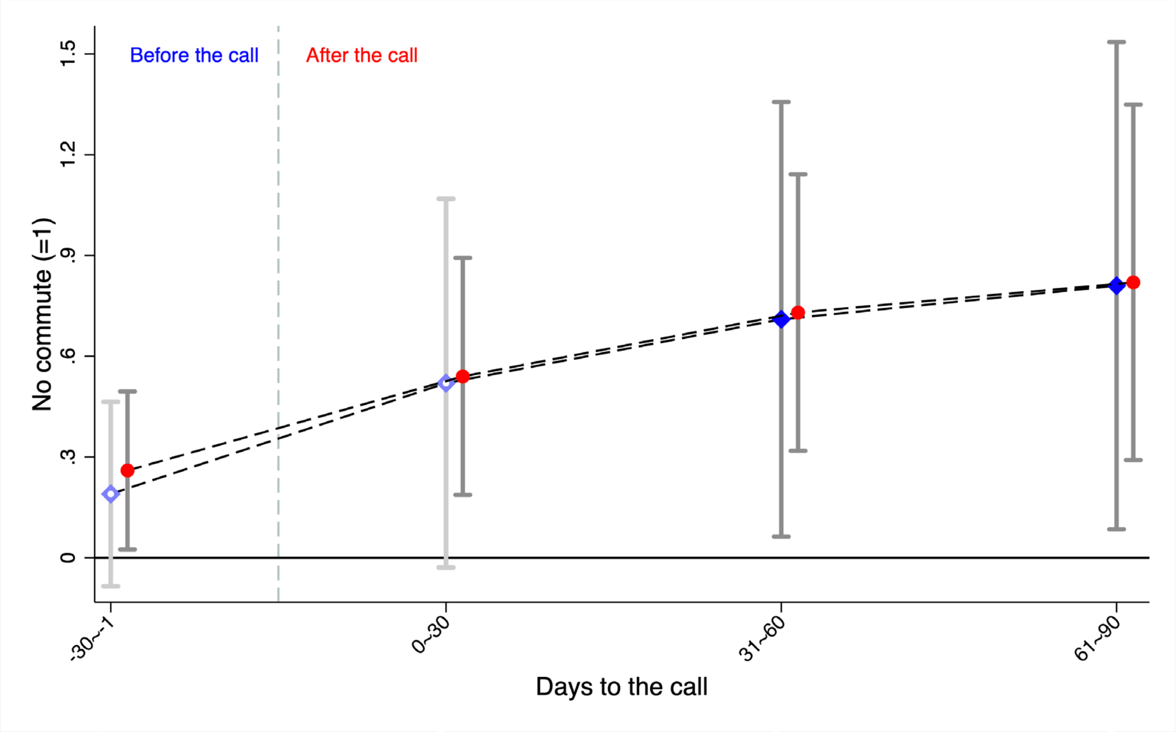


Figure A.5 compares the commuting patterns after calling 12333 in pre- and post-COVID period. Reassuringly, the commuting patterns after the call look very much similar in pre- (blue diamonds) and post-COVID (red dots) period.

4. MAIN RESULTS

Unemployment calls and stopping commuting

VARIABLES	(1) No. of individuals making calls (in log)	(2) Duration per call (in log)
Panel A:		
1-30 days before lockdown	0.02	0.03
	(0.04)	(0.04)
Lockdown period	-0.42***	-0.40***
	(0.07)	(0.07)
Phase I re-opening	0.23***	0.26***
	(0.03)	(0.04)
Phase II re-opening	0.40***	0.58***
	(0.03)	(0.05)
Panel B:		
Pandemic period	0.24***	0.37***
(Lockdown + Phases I + II)	(0.03)	(0.03)
Observations	489,514	489,514
R-squared	0.81	0.57
Neighborhood FE	Yes	Yes
Event-day FE	Yes	Yes
Day-of-week FE	Yes	Yes
Holidays FE	Yes	Yes
Treatment group FE	Yes	Yes

The unemployment rate can be computed from the estimated coefficient, such that

$$Unemp = \left(\exp\left[\beta - \frac{\mathrm{var}(\beta)}{2} \right] - 1 \right) \times 100$$

Therefore, the unemployment rate during the entire pandemic period (Lockdown + Phase I + Phase II) is

$$Unemp = \left(\exp\left[0.24 - \frac{0.03}{2} \right] - 1 \right) = 27\%$$

5. CONCLUSION

- The pandemic led to a **72% increase in unemployment** and a **57% increase in unemployment benefit claims** in Sept 2020 (nearly five months after the full reopening) relative to the pre-pandemic level in Sept 2019, based on the two strategies.
- The unemployment rate **during the entire pandemic period** (Lockdown + Phase I + Phase II) has **increased by 27%** relative to the same period in 2019.
- **Females, workers over 40,** and **migrant workers** experienced a more pronounced reduction in employment.
- The pandemic's impact is larger in cities with a high labor share of **hospitality, real estate,** or **transportation industries** but less severe in cities where jobs are concentrated in finance, health care, or education industries.
- Cities with a **higher export intensity** have been more affected, reflecting the **global nature** of the pandemic and the **interconnectedness of the world economy**.

Thanks!